



Archaeology in the Digital Era

Papers from the 40th Annual
Conference of Computer
Applications and Quantitative
Methods in Archaeology (CAA),
Southampton, 26-29 March 2012

Edited by Graeme Earl, Tim Sly,
Angeliki Chrysanthi, Patricia
Murrieta-Flores, Constantinos
Papadopoulos, Iza Romanowska
and David Wheatley

AMSTERDAM UNIVERSITY PRESS

CAA
Computer Applications &
Quantitative Methods in Archaeology

ARCHAEOLOGY IN THE DIGITAL ERA

Archaeology in the Digital Era

Papers from the 40th Annual Conference of
Computer Applications and Quantitative Methods in
Archaeology (CAA), Southampton, 26-29 March 2012

Edited by

Graeme Earl, Tim Sly, Angeliki Chrysanthi, Patricia Murrieta-Flores,
Constantinos Papadopoulos, Iza Romanowska and David Wheatley

AMSTERDAM UNIVERSITY PRESS

This book is published in the CAA series *Computer Applications and Quantitative Methods in Archaeology*.

Cover design: Magenta Ontwerpers, Bussum
Lay-out: Iza Romanowska

Amsterdam University Press English-language titles are distributed in the US and Canada by the University of Chicago Press.

ISBN 978 90 8964 663 7
e-ISBN 978 90 4851 959 0 (pdf)
e-ISBN 978 90 4851 960 6 (ePub)
NUR 684

© Computer Applications and Quantitative Methods in Archaeology (CAA) /
Amsterdam University Press, Amsterdam 2013

All rights reserved. Without limiting the rights under copyright reserved above, no part of this book may be reproduced, stored in or introduced into a retrieval system, or transmitted, in any form or by any means (electronic, mechanical, photocopying, recording or otherwise) without the written permission of both the copyright owner and the author of the book.

Every effort has been made to obtain permission to use all copyrighted illustrations reproduced in this book. Nonetheless, whosoever believes to have rights to this material is advised to contact the publisher.

Contents

11 Foreword

HUMAN COMPUTER INTERACTION, MULTIMEDIA, MUSEUMS

13 Disciplinary Issues: Challenging the Research and Practice of Computer Applications in Archaeology
Jeremy Huggett

25 Paperless Recording at the Sangro Valley Project
Christopher F. Motz and Sam C. Carrier

31 ‘Tangible Pasts’: User-centred Design of a Mixed Reality Application for Cultural Heritage
Angeliki Chrysanthi, Constantinos Papadopoulos, Tom Frankland and Graeme Earl

40 The Virtual Museum: a Quest for the Standard Definition
Laia Pujol and Anna Lorente

49 Can you Hack (the) Communication?
Hugh Corley

55 Identifying and Tracing Archaeological Material with RFID Tags
Ana María López, Ana María Salinas, Eduardo Pascual, Guillermo Ignacio Azuara, Gloria Fernández, Elena Gallego and Francisco Burillo

65 Matera Città Narrata Project: a Multimedia and Multi-Platform Guide for Mobile Systems
Eva Pietroni

74 Evaluating Virtual Museums: Archeovirtual Case Study
Sofia Pescarin, Alfonsina Pagano, Mattias Wallergård, Wim Hupperetz and Christie Ray

SIMULATING THE PAST

83 Extracting Scar and Ridge Features from 3D-scanned Lithic Artifacts
Eitan Richardson, Leore Grosman, Uzy Smilansky and Michael Werman

93 Retracing Prehistoric Population Events in Finland Using Simulation
Tarja Sundell, Juhana Kammonen, Martin Heger, Jukka U. Palo and Päivi Onkamo

105 SteppingIn - Modern Humans Moving into Europe - Implementation
Fulco Scherjon

118 Geospatial Virtual Heritage: A Gesture-Based 3D GIS to Engage the Public with Ancient Maya Archaeology
Heather Richards-Rissetto, Jim Robertsson, Jennifer von Schwerin, Giorgio Agugiaro, Fabio Remondino and Gabrio Girardi

- 131 Integration of 3D Laser Scanning and Virtual Reconstructions as Research and Educational Tools for Representing the Past. The Case Study of Roman Baths of Edeta**
Vito Porcelli, Fernando Cotino Villa, Josep Blasco i Senabre, Vicent Escrivá Torres and Julian Esteban Chapapriá
- 143 Reconstructing the Baths of Caracalla**
Taylor A. Oetelaar
- 147 Instruments and Methods for the Survey and Analysis of Amphitheatres**
Martina Ballarin, Francesco Guerra and Luigi Sperti
- 154 Why Hunter and Gatherers did not Die More Often? Simulating Prehistoric Decision Making**
Florencia Del Castillo and Juan A. Barceló
- 164 Reconstruction of Ruined Archaeological Structures Using Structural Analysis Methods**
James Miles, Aykut Erkal, Dina D'Ayala, Simon Keay and Graeme Earl

FIELD AND LAB RECORDING

- 177 Telling the Story of Ancient Coins by Means of Interactive RTI Images Visualization**
Gianpaolo Palma, Eliana Siotto, Marc Proesmans, Monica Baldassari, Clara Baracchini, Sabrina Batino and Roberto Scopigno
- 186 IBISA 3D: Image-Based Identification/Search for Archaeology Using a Three-dimensional Coin Model**
Sylvain Marchand
- 194 Using Image Analysis to Match a Coin to a Database**
Sebastian Zambanini and Martin Kampel
- 199 Pompeii Revived: Scanning Mission - Insula V 1**
Nicolò Dell'Unto, Matteo Dellepiane, Marco Callieri, Anne-Marie Leander, Stefan Lindgren and Carolina Larsson
- 208 A Metrology Tracking System Used as a 3D Drawing Tool for Faster Excavation Recording**
Maarten Smeets, Wouter Yperman and Geoff Avern

DATA MODELLING AND SHARING

- 216 From the Slope of Enlightenment to the Plateau of Productivity: Developing Linked Data at the ADS**
Michael Charno, Stuart Jeffrey, Ceri Binding, Doug Tudhope and Keith May
- 224 Reflections on the Rocky Road to E-Archaeology**
Geoff Carver and Matthias Lang
- 237 Least-cost Networks**
Irmela Herzog

- 249 Linking Roman Coins: Current Work at the American Numismatic Society**
Ethan Gruber, Gilles Bransbourg and Sebastian Heath and Andrew Meadows
- 259 GeoDia: or, Navigating Archaeological Time and Space in an American College Classroom**
Adam Rabinowitz

DATA ANALYSIS, MANAGEMENT, INTEGRATION AND VISUALISATION

- 269 Dating Historical Rock Art on Marble Surfaces by Means of a Mathematical Model for Natural Erosion Processes**
Paolo Emilio Bagnoli
- 279 Cultural Heritage Application Schema: a SDI Framework within the Protected Sites INSPIRE Spatial Data Theme**
Antonio Uriarte González, César Parcero Oubiña, Alfonso Fraguas Bravo, Pastor Fábrega Álvarez, Juan Manuel Vicent García, Esther Pérez Asensio, Carlos Fernández Freire and Isabel del Bosque González
- 291 Old Places, New ideas: New Routes into Canmore, the National Inventory of Scotland**
Susan Hamilton, Rebecca Jones and Peter McKeague
- 299 ADS easy: an Automated e-archiving System for Archaeology**
Ray Moore, Catherine Hardman, Julian Richards and Lei Xia
- 307 Archaeology in Broad Strokes: Collating Data for England from 1500 BC to AD 1086**
Chris Green
- 313 Beyond Inspire: Towards Delivering Richer Heritage Data in Scotland**
Peter McKeague and Mike Middleton
- 320 SEAD - The Strategic Environmental Archaeology Database Inter-linking Multiproxy Environmental Data with Archaeological Investigations and Ecology**
Philip Iain Buckland
- 332 Obscura itinera: a GIS-based Approach to Understand the pre-Roman and Roman Transhumance Pathways in Umbria and Sabina Regions (Central Italy)**
Paolo Camerieri and Tommaso Mattioli
- 340 Transparency, Testing and Standards for Archaeological Predictive Modelling**
William Wilcox

SPATIAL ANALYSIS

- 348 Zooming Patterns Among the Scales: a Statistics Technique to Detect Spatial Patterns Among Settlements**
Alessio Palmisano
- 357 The Long and Winding Road: Combining Least Cost Paths and Network Analysis Techniques for Settlement Location Analysis and Predictive Modelling**

Philip Verhagen, Tom Brughmans, Laure Nuninger and Frédérique Bertoncello

367 Can Infovis Tools Support the Analysis of Spatio-Temporal Diffusion Patterns in Historic Architecture?

Jean-Yves Blaise and Iwona Dudek

379 Introducing the Human Factor in Predictive Modelling: a Work in Progress

Philip Verhagen, Laure Nuninger, François-Pierre Tourneux, Frédérique Bertoncello and Karen Jeneson

389 Changing Settlement Patterns in the Mediterranean Context: a Case Study of Menorca (Balearic Islands) from Prehistory to the 19th Century AD

Monica De Cet, Rainer Duttmann, Simón Gornés, Joana Gual, Johannes Müller, Roberto Risch, Elena Sintes and Bianca Willié

400 A New Method of Spatial Analysis Based on the Extraction of Proximity Graphs

Diego Jiménez-Badillo

414 Bayesian Spatial Modelling of Radiocarbon Dated Archaeological Artefacts Using R-INLA

Juhana Kammonen, Tarja Sundell, Petro Pesonen, Markku Oinonen and Päivi Onkamo

420 From Space to Graphs to Understand Spatial Changes Using Medieval and Modern Fiscal Sources

Xavier Rodier, Mélanie Le Couédic, Florent Hautefeuille, Samuel Leturcq, Bertrand Jouve and Etienne Fieux

428 Exploring the Effects of Curvature and Refraction on GIS-based Visibility Studies

Mariza Kormann and Gary Lock

438 Process Formalization and Conceptual Modelling in the Study of Territorial Dynamics

Laure Saligny, Xavier Rodier, Estelle Gauthier, Nicolas Poirier, Murielle Georges Leroy, Frédérique Bertoncello and Olivier Weller

THEORETICAL APPROACHES AND CONTEXT OF ARCHAEOLOGICAL COMPUTING

449 Defining and Advocating Open Data in Archaeology

Stefano Costa, Anthony Beck, Andrew Bevan and Jessica Ogden

457 The Evolution of Territorial Occupation: Exploratory Spatial Data Analysis. Uncertainty and Heterogeneity of Data in Archaeology

Lucile Pillot and Laure Saligny

470 Visualising Time with Multiple Granularities: a Generic Framework

Iwona Dudek and Jean-Yves Blaise

Foreword

This volume consists of a selection of the peer-reviewed papers presented at the Computer Applications and Quantitative Methods in Archaeology 2012 conference hosted by the Archaeological Computing Research Group at the University of Southampton, UK between 26th and 30th March 2012. The conference included 53 sessions divided between the themes of simulating the past, spatial analysis, data modelling and sharing, data analysis, management, integration and visualisation, geospatial technologies, field and lab recording, theoretical approaches and the context of archaeological computing, and a general theme. In addition there were 12 workshops. A total of 380 papers and posters were presented, and two key note addresses. Alongside the lively conference atmosphere at the venue there was a thriving social media back channel. In addition to these proceedings there is therefore a broad ranging multimedia record of the event, accessible via the conference website.

The co-organisers of CAA2012 and myself would like to thank the CAA Steering Committee for their advice and assistance. We are also indebted to Professor Anne Curry (Dean of Faculty of Humanities) and Professor Jonathan Adams (Head of Archaeology) for their support and encouragement. Many individuals and organisations in Southampton and further afield, including the sponsors and exhibitors, contributed to making the conference such a success. Of course without the many delegates travelling from across the globe and offering such exciting contributions there could have been no conference, and we are very grateful to them for their lively contributions to all aspects of the event. Finally, we would like to offer our thanks to the superb team of volunteers that made CAA2012 possible. The Archaeological Computing Research Group at Southampton was very proud indeed to be able to host the 40th CAA conference and we know that this was demonstrated by the enthusiasm, dedication and professionalism of the postgraduate and undergraduate students that gave so much of their time to the event.

I very much hope that you enjoy these proceedings and all the many related outputs from CAA2012, and I look forward to seeing you at future CAA conferences.

Graeme Earl

Southampton, United Kingdom, November 2012

Disciplinary Issues: Challenging the Research and Practice of Computer Applications in Archaeology

Jeremy Huggett

University of Glasgow, United Kingdom

Abstract:

An annual international conference on computer applications in archaeology (CAA) has been meeting annually for almost forty years, so one might expect that there would be a reasonable idea of the nature and role of archaeological computing. However, some commentators see it as an emerging field while others suggest the need for a new archaeological speciality: Archaeological Information Science. Even the Wikipedia page on computational archaeology describes archaeoinformatics as an emerging discipline. Is this a sign of a lack of confidence in forty years-worth of enterprise and development or is it instead an indication of growing self-assurance in the subject? In recent years other fields, including GIScience and Information Systems, have sought to evaluate their intellectual core and identity; this paper suggests that it is time that archaeological computing does likewise.

Keywords:

Archaeoinformatics, Archaeological Information Science, Archaeological Computing, Digital Archaeology, Anxiety Discourse, Grand Challenges

1. Established or Emergent?

In 1967 George Cowgill suggested that the earliest use of electronic data processing in European archaeology was by Peter Ihm and Jean-Claude Gardin in 1958/1959 and in the USA by James Deetz in 1960 (Cowgill 1967, 17). Since then, activity in archaeological computing has grown substantially, especially since the first personal computer revolution in the 1980s, and the annual Computer Applications in Archaeology (CAA) conference has been meeting since 1973. It is surprising, then, that a series of articles presenting examples of archaeological computing in a recent issue of the Institute of Electrical and Electronics Engineers (IEEE) Computer Society magazine *Computer* were described as coming from “the emerging field of computational archaeology - the interface where archaeology meets computer science” (Bimber and Chang 2011, 31). Archaeological computing practitioners might be forgiven for thinking that work on this interface had been going on for over 50 years. Why might this apparent disconnect have arisen?

A simple reaction would be to suggest that the perception of the field as ‘emerging’ is simply mistaken. However, Bimber and Chang

Corresponding author: jeremy.huggett@glasgow.ac.uk

are not alone their characterisation. For example, Marcos Llobera has argued for the creation of a new focus in archaeology, Archaeological Information Science (AISc), “... concerned with the generation, representation and manipulation of archaeological information within the context of information systems.” (Llobera 2011, 218). The University of Washington’s Digital Archaeology Research Lab where he is based describes itself as promoting the development of Archaeological Information Science: ‘an emerging field that results from the intersection of Archaeology with Computer and Information Science, Engineering and Mathematics’ (DigAR Lab 2011). Popular presentations of archaeological computing on the web describe it as a novel field (for example, Bryant (nd) calls computational archaeology an “emerging facet” of archaeology) while Wikipedia describes Archaeological Information Science as “an emerging discipline that attempts to uncover, quantitatively represent and explore specific properties and patterns of archaeological information” (Wikipedia contributors, nd) and has no page for archaeological computing.

Emergence therefore seems to be associated with computational archaeology and Archaeological Information Science, not archaeological computing *per se*. Whether these are significantly different

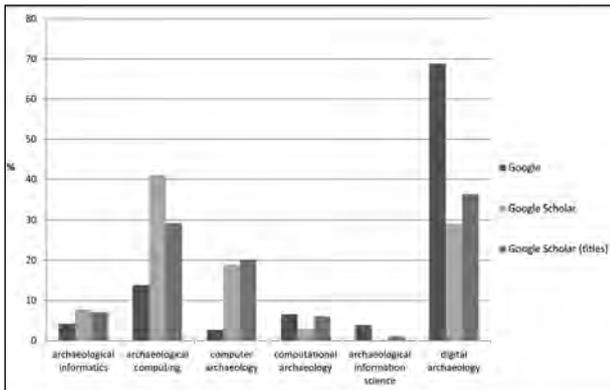


Figure 1. Counts of terms associated with archaeological computing returned by Google and Google Scholar expressed as a percentage of total hits.

from archaeological computing is debateable - Archaeological Information Science can be seen as concerned with the fundamental computational structures underpinning archaeological information; computational archaeology has strong connotations of mathematical and computational modelling and analysis, but neither definition is a stranger to the traditional field of archaeological computing.

2. Patterns of Use

A basic textual analysis demonstrates the shifting sands of the terms associated with archaeological computing. For example, searching Google for the phrases *archaeological computing*, *archaeological informatics*, *computer archaeology*, *computational archaeology*, *archaeological information science*, and *digital archaeology* demonstrates that digital archaeology is by far the most common term, with archaeological computing a long way behind in second place (Fig. 1). If the same search is carried out in Google Scholar, archaeological computing is the largest, followed by digital archaeology and computer archaeology (Fig. 1). If the Google Scholar search is restricted to titles only, digital archaeology comes to the fore, followed by archaeological computing and computer archaeology (Fig. 1). Of course, these searches take no account of the context in which the terms appear, and not all references will be strictly archaeological - terms such as digital archaeology are used in other disciplines and fields. By way of comparison, the same searches were conducted on Web of Knowledge and generated only single figure results for each term.

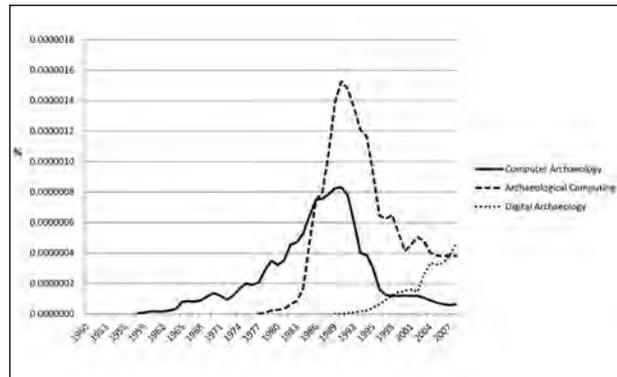


Figure 2. Trending terms associated with archaeological computing in Google's English corpus 1950-2008, incorporating upper and lower case variants. The graph is created from data extracted from the Google Ngram Viewer (<http://books.google.com/ngrams>) using the Engauge Digitizer software written by Mark Mitchell (<http://digitizer.sourceforge.net/>).

This underlines the predominance of terms other than Archaeological Information Science and computational archaeology. However, it does not reveal whether the pattern of use has changed over time.

The same terms can be entered into the Google nGram viewer: this searches for phrases within a sample of over 5.2 million books scanned by Google up to 2009 and normalises the results by the number of books published each year (Michel et al. 2011). Limitations of the tool include a requirement that the ngram term must occur in at least 40 books, and searches are case-sensitive, unlike the examples discussed above. One outcome is that terms such as computational archaeology, archaeological informatics, and Archaeological Information Science generate null results. Of those terms that do appear (Fig. 2), it is apparent that *computer archaeology* has the longest pedigree, in use since the late 1950s, but having peaked in the late 1980s the term has been in decline and is now the least common of the three in use. The usage of *archaeological computing* picks up in the late 1970s, quickly overtaking *computer archaeology* by the mid 1980s before declining after 1990. In contrast, *digital archaeology* appears around 1990 and has been on a steady rise in use since then, with a noticeable jump in popularity since 2000. It would be interesting to speculate what caused the peaks in occurrence of *archaeological computing* and

computer archaeology in the late 1980s and their subsequent decline alongside the rise in the use of *digital archaeology* from 1990. A similar analysis could be undertaken for other languages, but the French, Spanish and German corpora used by the Google tool are much smaller and currently generate null results for equivalent terms.

In summary, it can be shown that terms such as archaeological computing and computer archaeology have a long pedigree, but their recent decline in use has been matched by an increase in the occurrence of digital archaeology, while the use of computational archaeology and Archaeological Information Science is too recent to figure in the results. The pattern might suggest that we are at an inflection point, with traditional labels declining and new ones arriving, hence the references to 'emerging' fields.

It may be that the significance of the emerging terms is that they have a greater political value. For example, Kirschenbaum has characterised the label 'digital humanities' as a term of tactical convenience: it is employed to get things done, whether it is to capture funding, acquire a post, establish a curriculum, or launch a centre (Kirschenbaum 2012, 415). In much the same way, digital archaeology, Archaeological Information Science and computational archaeology may have a tactical purpose or value. Digital archaeology, the closer analogue of digital humanities, appears in the titles of books (for instance, Evans and Daly 2006), archaeological organisations (for example, Digitale Archäologie, based in Freiburg, the Center for Digital Archaeology (CoDA) at the University of California, Berkeley, the Laboratorio di Archaeologia Digitale at the University of Foggia, the Digital Archaeology Research Lab at the University of Washington, Seattle, and the Digital Archaeological Record (tDAR)), exhibitions which chart the development of the web (at Internet Week Europe in 2010 and Internet Week New York 2011), and is used in a wide variety of contexts as shorthand for activities such as forensic data recovery (for example, Ross and Gow 1999) and the investigation of digital media (for example, Wardrip-Fruin 2011). While digital archaeology as a term is increasingly widely employed both within and without the subject, computational archaeology and Archaeological Information Science are in less widespread use

as yet but do appear as titles of organisations or as part of their mission statements (for example, the Computational Archaeology Laboratory at McGill University, the DigAR Lab at Washington University).

3. What's in a Name?

In Shakespeare's terms, none of this may be significant - what we choose to call the archaeological computing rose does not change its inherent qualities. On the other hand, the changing terms associated with a subject - the introduction of new labels and the decline of old ones - may reveal more fundamental changes within that subject. For example, the increased usage of 'digital archaeology' may parallel a discursive shift from 'humanities computing' to 'digital humanities' observed by Svensson (2009) and others, who point to the way in which naming is related to changes in institutional, disciplinary, epistemological, and social organisation (Svensson 2009, 2). Certainly Llobera's case for the development of Archaeological Information Science is based on a dissatisfaction with the current impact and use of computer applications within archaeology (Llobera 2011, 216).

The tensions evident in the process of naming and renaming can be seen as forming part of an 'anxiety discourse' surrounding archaeological computing, one which questions the identity, nature and academic legitimacy of the field, expressing concerns about the theoretical core, the rigour and relevance of the methodologies used, the usefulness and value of outputs generated, and the level of significance and recognition of contributions made. This in itself is not unusual - for example, crises of confidence have arisen in areas such as Information Systems (for instance, Hassan 2011; Sawyer and Winter 2011; King and Lyytinen 2004), Digital Humanities (for example, Berry 2011; Davidson 2008; Hall 2011), and Computing Science (Tedre 2011), and contributed to the creation of Geographical Information Science (Raper 2009); indeed, they can be seen as part of the process of discipline-building.

A characteristic of fields which undergo this anxiety discourse is that they frequently meet their disciplinary challenges by drawing down concepts and methodologies from external subjects, and

have an intellectual centre primarily focused on praxis with theory being derived from outside (for example, Lyytinen and King 2004, 222). Since Dunn has characterised archaeology as “a disciplinary mashup, needing support from a range of technological infrastructures, at all levels of scale and complexity” (Dunn 2011, 98), while Daly and Evans (2006, 3) defined digital archaeology as “not so much a specialism, nor a theoretical school, but an approach - a way of better utilizing computers based on an understanding of the strengths and limits of computers and information technology as a whole”, it would be surprising if archaeological computing were not to experience some degree of uncertainty regarding its focus, its boundaries, and its relevance to the broader subject. Nor would it be alone in this regard: an example of another field within archaeology which has recently experienced ‘anxiety discourse’ is the archaeology of modernity (for example, see Gonzalez-Ruibal 2008; Harrison and Schofield 2010).

4. An Anxiety Discourse

An anxiety discourse for a subject which attracts hundreds of participants to its annual conference, a truly international event, may seem a strange undertaking, but wrapping ourselves in the warm blanket of CAA may disguise underlying issues or concerns. By way of demonstrating (or encouraging) a sense of anxiety concerning archaeological computing or digital archaeology, a number of observations can be made.

As far as publication is concerned - the lifeblood of a discipline - there is only one journal of archaeological computing (*Archeologia e Calcolatori*) although arguably one of the strengths of the subject is the way that archaeological computing papers appear in mainstream archaeology journals and, to a lesser extent, in journals outside the field, which implies a participation in archaeology more generally. There are no book series, and relatively few books in the field. The main publication remains the annual Computer Applications and Quantitative Methods in Archaeology proceedings, but concerns regarding the profile and relevance of the proceedings have been a regular topic of debate at the annual general meetings although published proceedings in recent years have been more selective than in the past.

More broadly, how would we measure the ‘impact’ of archaeological computing within archaeology more generally, and within those external disciplines which we draw upon such as Geographical Information Science? A citation analysis within digital humanities journals concluded that the humanities community more generally tended not to read or at least cite work published in humanities computing, suggesting that there was a failure to produce digital humanities scholarship of outstanding importance and significance within the broader field (Juola 2008, 73-75). Gauging the impact of the CAA proceedings using citation analysis is difficult because of the changing names of the proceedings as each conference seeks to stamp its own mark; however, the suspicion must be that the significance of the proceedings is largely internal to the field with little impact beyond.

Archaeological scholarship more generally is already digital in many respects, but to what extent has the archaeological computing field shaped this? What proportion of the outcomes of archaeological computing are used by the wider community? We can point to the transformative effect of online access to archaeological data, perhaps, or the development of excavation software, for example, but is a prerequisite of these the existence of a digital archaeology field? Indeed, most of the tools we use are borrowed from elsewhere (Schollar 1999, 8; Lull 1999, 381). Does this compound a perception that archaeological computing tends to be practice-based and can consequently appear to be under-theorised which, amongst other things, leads to an undervaluing of archaeological computing in research assessment terms? To what extent are the leading scholars in archaeological computing also leading in the archaeological community and beyond? Individual scholars may make contributions to borrowed fields, but this tends to be on an individual basis, limiting the impact of archaeological computing more generally. As a subject area, digital archaeology receives rather than gives, which means that it is rarely perceived as a donor or reference field. Where are the ‘big ideas’ that are generating large-scale international projects which require petaflop performance and exabyte storage, crucially attract large-scale finance, and which have the potential to transform the practice of archaeology and contribute to theories and methods in other disciplines?

In proposing a new field of Archaeological Information Science, Llobera's criticisms of archaeological computing cut even deeper than this. For example, he argues that the impact of computer applications in archaeology is more quantitative than qualitative:

"We are able to record information much more quickly in the field but to what degree is this "new information"? How much has it changed the way we conduct our analysis? We have the capacity to process and visualize information in novel ways but are we actually doing this? More importantly, are we even thinking about new possibilities? How do these new developments relate, if at all, with theoretical orientations currently found in archaeology? Has the introduction of information systems precipitated new ways of doing archaeology?" (Llobera 2011, 217).

Worse still, he suggests that through ignorance and lack of training we fail to deal with the limitations inherent in the software appropriately:

"In some cases, archaeologists brush off any concerns about these deficiencies by claiming that they are just archaeologists and not computer scientists. They are simply applying what is on offer and that is that. All that is needed then is to mention the limitation to make it right and to proceed as usual. Actively engaging in understanding the underpinnings of applications, let alone developing them, is generally not considered to be archaeology even among many of those trained in IT within the discipline." (Llobera 2011, 217).

Consequently, he concludes that archaeological computing practitioners are little more than technicians (Llobera 2011, 218), and not very good ones at that.

Of course, anxiety discourses can rapidly become excessively negative and inward-looking, and that is not the objective here. Nor is it the intention to suggest that there is not a large amount of valuable, high quality research taking place in archaeological computing. An anxiety discourse can be seen as being associated with questions of disciplinarity, in the sense that it seeks validation for the methods of a particular intellectual community and its relationships with other disciplines (for

example, Post 2009). At the same time, it can be seen as simply presenting a series of challenges to a subject, field or discipline. To be valuable, however, an anxiety discourse, having drawn attention to a range of issues concerning role, methods and relationships, needs to develop into a positive discourse which actively addresses these concerns.

5. Grand Challenges?

One of the dangers for a subject area which derives much of its methods from other disciplines, which operates on the boundaries of different disciplines, and which is seen as promoting technique and technology over theory, is that it is always seen as peripheral, ever playing a supporting role, and lacking a coherent central core which provides a clear-cut identity. The outcome of an anxiety discourse coupled with a positive response may be the strengthening of a formal discipline, or the creation of an 'indiscipline' which routinely questions and disrupts practice (Post 2009, 761), or a 'shadow discipline' (Chandler 2009, 757) arising out of changing organisation and interests within archaeology and allied disciplines. Indeed, 'disciplinification' may be seen as too rigid and restrictive for a field which thrives on shifting practices and nimble knowledge-making (Biagioli 2009, 819). Regardless, the fundamental question arising out of this discourse essentially concerns whether archaeological computing is capable of getting us to rethink archaeology, or whether we are content to see it absorbed as a methodology or series of techniques? Should archaeological computing continue to rely on core techniques and methodologies derived from other disciplines, or should there be a greater emphasis on developing new tools and approaches?

Such questions move into the realms of 'grand challenges', which have been essentially lacking in archaeological computing over the years. However, the opportunity for a grand challenges within a subject only arises occasionally, once a field reaches a sufficient level of maturity to be able to predict and plan the direction of its future progress (Hoare 2003, 63). The value of grand challenges lie in their intellectual aspiration and technical ambition, and their potential to transform not only the immediate subject area, but multiple academic fields and communities beyond. For example, Winter and

Butler suggest that “Engaging grand challenges helps mobilize requisite resources, provide legitimacy, and focuses economic, social, and research activities” (2011, 99). Grand challenges:

“seek to drastically alter the boundaries of existing knowledge, established disciplines, and available capabilities ... [they] require cooperation and interaction between groups with differing perspectives over years and decades. New norms, structures, and practices must be developed to provide the support and incentives necessary to sustain these long-term, large-scale collaborative efforts. Addressing grand challenges also requires mobilization of substantial resources and significant participation from members of many relevant academic, practitioner, and policy-oriented communities.” (Winter and Butler 2011, 100).

They use the example of the project to map the human genome - a 20-year effort by researchers in six countries and 20 research centres, which developed from a peripheral project to become critical infrastructure (Winter and Butler 2011, 100). Where is the digital archaeology equivalent? A project in its early stages is the National Science Foundation (NSF) funded programme on Planning Archaeological Infrastructure for Integrative Science (NSF 2011; see also Kintigh 2012) which is using the identification of ‘grand challenges’ in order to develop proposals for major future NSF investment in cyberinfrastructures for archaeology. However, for the most part archaeological computing is implicit in, but largely unengaged with, a number of national and international sets of computing ‘grand challenges’ and consequently many archaeologists will be blissfully unaware of them.

For example, an initiative in the UK was launched in 2002 by the UK Computing Research Committee (UKCRC), supported by the National e-Science Centre and the Engineering and Physical Sciences Research Council (EPSRC). It sought to investigate possibilities for the advancement of computing research and identify ambitious long-term research initiatives. By 2008, a total of eight grand challenges had been identified (Kavanagh and Hall 2008, 1), with additional proposals under consideration in 2010 (Allan 2010, 2) although there seems to have been little further development. Of relevance to archaeological computing is grand

challenge *GC9: Bringing the Past to Life for the Citizen* (Arnold 2008; Allan 2010, 15-16). Key areas which were identified for research (Allan 2010, 16) were:

- an integrated data infrastructure for cultural heritage, requiring the definition of data formats to allow interoperability of tools and data
- digitisation and preservation of collections
- intelligent interactive tools for non-expert use to enable cultural heritage professionals to work in their domain of expertise rather than struggle with general purpose tools
- modelling and visualisation systems which are capable of differentiating between interpretation and evidence supported by fact, to avoid misleading prettified presentations
- management of very large data sets
- algorithms, data structures and systems for visualising very large animated and detailed multimedia data sets
- natural language technologies for interpreting historical accounts

These issues are essentially driven by computing science although it is noted that cultural heritage professionals may perceive the demands of computing science as being crass and insensitive (Allan 2010, 15).

In parallel with the UK Computing Research Committee’s grand challenges initiative, a European Network of Excellence was funded under the European Union’s Sixth Framework Programme. *EPOCH - Excellence in Processing Open Cultural Heritage* - involved 95 university-based and other research centres, commercial technology organisations, and public agencies and individual institutions responsible for cultural heritage (Arnold and Geser 2008, 21). The objective was to define a common research agenda for the application of ICT to cultural heritage, with the focus being on support for the work processes of cultural heritage organisations (Arnold and Geser 2008, 22). In a lengthy report, a total of 65 recommendations are

made alongside a series of technological research topics and priorities (Arnold and Geser 2008, 63-106). These include:

- data and metadata capture, including intelligent data capture tools and artefact/collection digitization,
- documentation of 3D digital objects, including the definition of sustainable file formats, markup, and processing history,
- user-created cultural heritage content, ranging from personal records to tagging of collections,
- intelligent tools built using semantic domain-specific knowledge (although some resistance amongst cultural heritage professionals is recognised (Arnold and Geser 2008, 77-78)),
- semantic and multi-lingual processing for content analysis, indexing, classifying and interrogating collections, and knowledge discovery,
- visualisation and presentation, covering reconstruction and recreation of objects, structures and environments, including their subsequent dissemination,
- systems for cultural heritage monitoring, risk assessment and damage prevention,
- distributed and mobile systems for site monitoring, excavation and fieldwork, including the integration of GIS,
- mobile location-based, context-aware applications for visitor information services,
- augmented reality applications combining real and virtual scenes within a museum or site environment.

As with the UKCRC initiative, it is clear that archaeological computing practitioners are already working in a number of these areas, and a great deal of effective work has been carried out, even if not under the auspices of these 'grand challenges' or research agendas. For example, the Archaeology Data Service in the UK, and Open Context and the Digital Archaeological Record (tDAR) in the USA have been

working in several of these areas, as have English Heritage and the Royal Commission on the Ancient and Historical Monuments of Scotland (RCAHMS), while a range of university-based research projects have been addressing other issues, as have a number of the larger archaeological units. However, both the UKCRC Grand Challenges and the EPOCH Research Agenda are largely developed outside archaeology, with few archaeologists involved in their creation (for example, see the acknowledgements in Arnold and Geser 2008, 24-25). Consequently, despite a certain sensitivity evident in both reports, they suffer from a 'not invented here' syndrome and address a range of issues which, as they are not derived from an archaeological perspective, may not meet the needs, requirements, or interests of archaeologists, whether computing experts or not. What these challenges do reveal is the diversity of archaeological computing - that fact that computing archaeologists can express an interest in or claim to be active in these different areas underlines what can be both a strength and a weakness of the field. Too much diversity can be as bad as too little: it may result in a diffuse lack of focus or it may reduce the potential for innovation and invention.

This is not to suggest that these grand challenges or research agendas have no value - they evidently deal with areas which are highly relevant to archaeological computing and in which there is already a good deal of activity. The argument instead is that digital archaeology should be defining its own grand challenges or research agendas rather than relying on those defined by others. In this way, the field can address concerns and issues identified within it, and define and determine its relationship and contributions to the broader archaeological and other boundary disciplines.

6. Evaluating Challenges

The first stage in identifying a grand challenge is determining how it might best be characterised. By definition, a grand challenge will commit resources, in terms of people, time, finance, etc. for several years, so a means of recognising a valuable grand challenge and distinguishing it from the futile or the facile is necessary. A challenge needs to be sufficiently complex and at the same time sufficiently simple that it can be defined and there is a reasonable prospect of achieving it. It will also need to have an impact

on archaeological computing, on archaeology more generally, and beyond archaeology.

A simple approach would be to adopt the SMART (Specific, Measurable, Achievable, Realistic and Time-bound) criteria widely used for defining objectives, or one of its variants (for instance, SMARTER - including Evaluation and Review, or C-SMART - making the criteria Challenging). Useful though these are, they are not a comprehensive list and are insufficiently focused for a 'grand challenge' for archaeological computing. Fortunately, establishing these criteria from scratch is not necessary.

For example, the UK Computing Research Committee's grand challenges initiative defined a series of criteria against which challenges could be measured. According to their website (UKRC 2009 - see also Hoare 2003, 63-64) a grand challenge should:

- arise from scientific curiosity about the fundamental aspects of the nature or the limits of the discipline,
- give scope for ambition to create something that is truly novel and innovative,
- provide a means of indicating the extent to which and when the challenge has (or has not) been met,
- have enthusiastic support from the research community, even those who do not participate and do not benefit from it,
- have international scope and involvement,
- be comprehensible, so that it can attract the attention of those in other disciplines and capture the imagination of the general public,
- be a long-standing issue,
- promise to go beyond what is initially possible, and ultimately requires the development of understanding, techniques and tools which are currently unknown,
- involve planned co-operation between different

researchers and research teams,

- encourage competition among researchers and teams, with clear criteria on who is winning, or who has won,
- be capable of decomposition into identified intermediate research goals, whose achievement is beneficial even if the project as a whole ultimately fails,
- lead to radical paradigm shift,
- be unlikely to be met from evolutionary professional/commercial development.

No individual challenge was expected to meet all the criteria.

Similarly, the EPOCH Research Agenda defined a series of criteria which were considered in the selection of research priorities (Arnold and Geser 2008, 22-23):

- potential for new or enhanced technological capability rather than customising what is already available; technology transfer from other disciplines is considered to be low priority,
- relevance for cultural heritage organisations and their constituencies,
- potential wider benefits for areas such as tourism, cultural and creative industries, and regional regeneration,
- interdisciplinarity - collaboration with domain experts and non-professional user-groups.

In addition, a series of archaeological criteria have already been defined but have not received wider consideration. An archaeological seminar, sponsored by the then Arts and Humanities Data Service was held in June 2006 as part of an AHRC-funded e-Science Scoping Study which was intended to evaluate how e-Science might be exploited within the arts and humanities community (for example, Blanke et al. 2009; Blanke and Dunn 2006). Directly related archaeological outcomes included the Silchester-based Virtual Research Environment project (VERA) (Rains 2011), the ADS Archaeotools

project (Richards et al. 2011), and the Medieval Warfare on the Grid project at Birmingham University's IBM-VISTA Visualisation lab (Craenen et al. 2010). The archaeological e-Science seminar produced a series of generic principles on which archaeological e-Science projects should be judged (Kilbride 2006, 8-10). These can be summarised in terms of the following criteria:

- focus on the needs of archaeology, rather than being technology-driven, and be assessed on the basis of archaeological values, rather than the standards of the discipline which may have initiated or invented a specific application,
- be considered to be fundamental research, pushing the boundaries in the use of ICT in research,
- involve and be embedded within the whole sector, not just the academic community, working with existing standards and professional groupings and including work on training and skills,
- be international, scalable, sustainable, implementable,
- go beyond resource discovery, finding aids, extensions to existing metadata or ontological tools (in this respect, archaeology was distinct from the humanities more generally, where location and access were seen as being fundamental to the 'grand challenge' (Blanke and Dunn 2006)),
- distinguish between analysis and interpretation,
- measure success not simply in terms of successful outcomes: failure is not necessarily bad as long as lessons are learned.

On their own, each set of criteria has its weaknesses. Although comprehensive, the UK Computing Research Committee's criteria are by definition focused on computing science; the EPOCH criteria are notable for not including basic SMART-like elements; while the e-Science seminar group criteria were focused on how archaeology might contribute to the e-Science agenda more broadly and as a result perhaps do not go sufficiently far in terms of archaeological computing itself. However,

there is some commonality between the three sets of criteria, suggesting that a combination of the three might have some success in capturing the range and richness of the criteria needed to evaluate potential grand challenges.

As Hoare emphasises (2003, 65), a putative grand challenge is not expected to satisfy all the criteria. The criteria are demanding since they need to distinguish between a typical research project and a grand challenge - what may be adequate and sufficient for a research project will not measure up to a grand challenge. For example, a research project may well involve the transfer of concepts and technology from a donor discipline into an archaeological context, and this is quite commonplace; however, it does not constitute a grand challenge unless it is offset by a number of the other criteria. Furthermore, the extent of disciplinarity will be determined by the range of criteria which apply; so, for example, the requirement for developing theories, concepts and techniques within the field rather than borrowing them elsewhere has clear implications for the character of the subject. Clearly, the identification of challenges is not straightforward and is open to debate as different challenges and criteria are played off against each other along with beliefs and perceptions of the subject itself.

7. Rising to the Challenge

It would be hubristic to end with a series of proposals for grand challenges - however, some indication of what might constitute grand challenges might be useful by way of providing examples. The following suggestions are therefore no more than that - they are not recommendations or claims for pre-eminence in any subsequent discussion of grand challenges. Potential areas which might give rise to grand challenges include:

Data mining: current semantic methods of data processing are still relatively novel and employ restricted ontologies. A challenge might therefore be to create methods which overcome the current limitations of data documentation and which can be applied to information extraction from very large collections of diffuse and dispersed digital data, maximising quality of returns and minimising false hits.

Data capture: site recording using computers largely continues to record traditional forms of data in traditional ways and consequently the nature, quality and detail of those data are rarely greater than in the past. Developing tools and techniques which would allow high-resolution rapid capture of site data without recourse to expensive instrumentation, with simple incorporation of multiple categories of data, such as text, artefacts, samples etc. could radically transform the process of excavation.

Image processing: the large collections of satellite and aerial imagery, especially with the quality of international imagery increasingly available through Google Earth, are nevertheless notoriously slow and complex to process and record. A means of automatically identifying, classifying and extracting human activity from the imagery could have significant benefits; it could also entail the development of a crowd-sourcing tool to check and confirm the results, with the data feeding back into the extraction process.

Whether these constitute grand challenges remains open to question. However, they have transformative potential, are interdisciplinary in outlook and undertaking, are difficult to achieve and likely to represent a considerable degree of effort, are not specific to the needs of one region or country, require significant computing capacity, are amenable to different approaches and hence capable of being addressed by different research teams, and so on.

However, there are two further Grand Challenges which underlie this discussion. First is simply the challenge that archaeological computing as a field should confront the task of constructing and pursuing grand challenges in the first place. This undertaking requires community involvement, participation, and commitment if it is to have any meaning or value, and should not be simply devolved to an 'expert panel' but discussed and debated widely, over a period of time, before coming to a collective decision. Secondly, there is the challenge to the significance of the field of archaeological computing itself. If the field cannot come up with a set of agreed grand challenges, or cannot see the value in doing this, either archaeological computing is not yet sufficiently mature as a subject, or it will never have

a clear identity and rise above a perception that it provides little more than technical support for 'real' archaeologists.

The prospect of developing grand challenges arose out of a concern for the identity, purpose and future of archaeological computing, initiated by evidence of shifting terminology and consequent meaning. To paraphrase Foster (2010), instead of asking "what is 'digital archaeology'?" (or 'archaeological computing', or 'archaeoinformatics' etc.) we might ask with greater profit "what do you want from digital archaeology?" The challenge in the end is an intellectual one, and the answer depends on our beliefs, perceptions, ambitions and desires for the subject, and that is perhaps the greatest challenge of all.

References

- Allan, R.J. 2010. „Computing Grand Challenges.“ *Science & Technology Facilities Council Technical Report DL-TR-2010-008*. <http://epubs.cclrc.ac.uk/bitstream/5681/DLTR-2010-008.pdf>.
- Arnold, D. 2008. "GC9: Bringing the Past to Life for the Citizen." In *Grand Challenges in Computing Research Conference 2008*, edited by John Kavanagh and Wendy Hall, 17-18. London: UK Computer Research Committee/ British Computer Society. <http://www.ukerc.org.uk/grand-challenge/index.cfm>.
- Arnold, D., and G. Geser. 2008. *EPOCH Research Agenda for the Applications of ICT to Cultural Heritage*. Budapest: Archaeolingua. http://public-repository.epoch-net.org/publications/RES_AGENDA/final_res.pdf.
- Berry, D.M. 2011. "The computational turn. Thinking about the Digital Humanities." *Culture Machine* 12. <http://www.culturemachine.net/index.php/cm/article/view/440/470>.
- Biagioli, M. 2009. "Postdisciplinary liaisons. Science Studies and the Humanities." *Critical Inquiry* 35 (4): 816-833.
- Blanke, T., M. Hedges, and S. Dunn. 2009. "Arts and humanities e-science. Current practices and future challenges." *Future Generation Computer Systems* 25 (4): 474-480.

- Blanke, T., and S. Dunn. 2006. "The Arts and Humanities e-Science Initiative in the UK." Paper presented at the *Second IEEE International Conference on e-Science and Grid Computing, e-Science '06*. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4031109&isnumber=4030973>.
- Bryant, C.W. nd. "What can we learn from computational archaeology?" *HowStuffWorks* <http://science.howstuffworks.com/environmental/earth/geology/computational-archaeology.htm>.
- Chandler, J. 2009. "Introduction. Doctrines, Disciplines, Discourses, Departments." *Critical Inquiry* 35 (4): 729-746.
- Cowgill, G.L. 1967. "Computer Applications in Archaeology." *Computers and the Humanities* 2 (1): 17-23.
- Craenen, B., G. Theodoropoulos, V. Suryanarayanan, V. Gaffney, P. Murgatroyd, and J. Haldon. 2010. "Medieval military logistics: a case for distributed agent-based simulation." In *Proceedings of the 3rd International ICST Conference on Simulation Tools and Techniques (SIMUTools '10)*. Brussels: ICST. <http://eudl.eu/doi/10.4108/ICST.SIMUTOOLS2010.8737>.
- Daly, P., and T.L. Evans. 2006. "Introduction. Archaeological theory and digital pasts." In *Digital Archaeology: bridging method and theory*, edited by T. L. Evans and P. Daly, 3-9. London: Routledge.
- Davidson, C.N. 2008. "Humanities 2.0. Promise, perils, predictions." *Publications of the Modern Language Association of America (PMLA)* 123 (3): 707-717.
- DigAR Lab. 2011. "Welcome to the DigAR Lab!" Accessed 1 March 2011. <http://depts.washington.edu/digarlabor/>.
- Dunn, D. 2011. "Poor relatives or favourite uncles? Cyberinfrastructure and Web 2.0: a critical comparison for archaeological research." In *Archaeology 2.0: New Approaches to Communication and Collaboration*, edited by E. C. Kansa, S. Whitcher Kansa and E. Watrall, 95-118. Los Angeles: University of California. <http://escholarship.org/uc/item/1r6137tb>.
- Evans, T.L., and P. Daly (eds). 2006. *Digital Archaeology: bridging method and theory*. London: Routledge.
- Foster, C. 2010. "I'm Chris. Where am I wrong?" *HASTAC (Humanities, Arts, Science and Technology Advanced Collaboratory)*. Accessed 9 August 2010. <http://hastac.org/blogs/cforster/im-chris-where-am-i-wrong>.
- González-Ruibal, A.. 2008. "Time to destroy. An archaeology of supermodernity." *Current Anthropology* 49 (2): 247-79
- Hall, G. 2011. "The Digital Humanities. Beyond computing: A Postscript." *Culture Machine* 12. <http://www.culturemachine.net/index.php/cm/article/view/441/471>.
- Harrison, R., and J Schofield. 2010. *After Modernity. Archaeological approaches to the contemporary period*. Oxford: Oxford University Press.
- Hassan, N.R. 2011. "Is information systems a discipline? Foucauldian and Toulminian insights." *European Journal of Information Systems* 20: 456-76.
- Hoare, T. 2003. "The verifying compiler. A grand challenge for computing research." *Journal of the ACM* 50 (1): 63-69.
- Juola, P. 2008. "Killer applications in Digital Humanities." *Literary and Linguistic Computing* 23 (1): 73-83.
- Kavanagh, J., and W. Hall (eds). 2008. *Grand Challenges in Computing Research Conference 2008*. UK Computer Research Committee/British Computer Society. <http://www.ukcrc.org.uk/grand-challenge/index.cfm>.
- Kilbride, W. 2006. "Grand challenges: Grand opportunities? Archaeology, the historic environment sector and the e-Science programme." AHDS e-Science Scoping Study Expert Seminar Report. <http://www.ahds.ac.uk/e-science/e-science-scoping-study.htm>.
- King, J.L., and K. Lyytinen. 2004. "Reach and grasp." *MIS Quarterly* 28 (4): 539-551.
- Kintigh, K. 2012. "Planning Archaeological Infrastructure for Integrative Science." National Science Foundation Division of Behavioral and Cognitive Sciences award number 1202413. <http://sustainability.asu.edu/research/project/680/planning-archaeological-infrastructure-for-integrative-science>.
- Kirschenbaum, M.G. 2012. "Digital Humanities as/is

- a tactical term.” In *Debates in the Digital Humanities*, edited by M.K. Gold, 415-428. Minneapolis: University of Minnesota Press.
- Llobera, M. 2011. “Archaeological Visualisation. Towards an Archaeological Information Science (AISC).” *Journal of Archaeological Method and Theory* 18: 193-223.
- Lull, V. 1999. “The new technologies and designer archaeology.” In *New Techniques for Old Times: Computer Applications and Quantitative Methods in Archaeology 1998*, edited by J. Barceló, I. Briz and A. Vila, 379-383. Oxford: Archaeopress.
- Lyytinen, K., and J.L. King. 2004. “Nothing at the Center? Academic legitimacy in the Information Systems field.” *Journal of the Association for Information Systems* 5 (6): 220-245.
- Michel, J.-B., Y.K. Shen, A. Presser Aiden, A. Veres, M.K. Gray, The Google Books Team, J.P. Pickett, D. Hoiberg, D. Clancy, P. Norvig, J. Orwant, S. Pinker, M.A. Nowak, and E. Lieberman Aiden. 2011. “Quantitative analysis of culture using millions of digitised books.” *Science* 331: 176-182.
- National Science Foundation. 2011. “Planning Archaeological Infrastructure for Integrative Science.” NSF Award Number 1202413 November 2011. <http://www.nsf.gov/awardsearch/showAward.do?AwardNumber=1202413>.
- Post, R. 2009. “Debating disciplinary.” *Critical Inquiry* 35 (4): 749-770.
- Rains, M. 2011. “Creating a Virtual Research Environment for Archaeology.” In *Archaeology 2.0: New Approaches to Communication and Collaboration*, edited by Eric C. Kansa, Sarah Whitcher Kansa and Ethan Watrall, 159-169. Los Angeles: University of California. <http://escholarship.org/uc/item/1r6137tb>.
- Raper, J. 2009. “Geographic Information Science.” *Annual Review of Information Science and Technology* 43: 1-117.
- Richards, J., S. Jeffrey, S. Waller, F. Ciravegna, S. Chapman, and Z. Zhang. 2011. “The Archaeology Data Service and the Archaeotools Project: Faceted Classification and Natural Language Processing In *Archaeology 2.0: New Approaches to Communication and Collaboration*, edited by E. C. Kansa, S. Whitcher Kansa and E. Watrall, 31-56. Los Angeles: University of California. <http://escholarship.org/uc/item/1r6137tb>.
- Ross, S., and A. Gow. 1999. *Digital Archaeology: Rescuing Neglected and Damaged Data Resources*, JISC/NPO Study within the eLib Programme on the Preservation of Electronic Material. London: Library and Information Technology Centre. <http://eprints.erpanet.org/47/01/rosgowrt.pdf>.
- Sawyer, S., and S.J. Winter. 2011. “Special issue on futures for research on information systems. Prometheus unbound?” *Journal of Information Technology* 26: 95-98.
- Schollar, I. 1999. “25 years of computer applications in archaeology.” In *Archaeology in the Age of the Internet CAA97*, edited by L. Dingwall, S. Exon, V. Gaffney, S. Laflin, and M. van Leusen, 5-10. Oxford: Archaeopress.
- Svensson, P. 2009. “Humanities computing as digital humanities.” *Digital Humanities Quarterly* 3 (3). <http://www.digitalhumanities.org/dhq/vol/3/3/000065/000065.html>.
- Tedre, M. 2011. “Computing as a science. A survey of computing viewpoints.” *Minds and Machines* 21: 361-87.
- UKRC. 2009. “Criteria of maturity for a Grand Challenge.” UK Computing Research Committee. <http://www.ukerc.org.uk/grand-challenge/criteria.cfm>.
- Wardrip-Fruin, N. 2011. “Digital media archaeology. Interpreting computational processes.” In *Media Archaeology: approaches, applications, and implications*, edited by E. Huhtamo and J. Parikka, 302-322. Berkeley: University of California Press.
- Wikipedia contributors. nd. “Computational Archaeology.” *Wikipedia, The Free Encyclopedia*. http://en.wikipedia.org/wiki/Computational_archaeology.
- Winter, S.J., and B.S. Butler. 2011. “Creating bigger problems. Grand challenges as boundary objects and the legitimacy of the information systems field.” *Journal of Information Technology* 26: 99-108.

Paperless Recording at the Sangro Valley Project

Christopher F. Motz

University of Cincinnati, USA

Sam C. Carrier

Oberlin College, USA

Abstract:

This paper presents the results of the Sangro Valley Project's deployment of a paperless recording system in a mixed environment of excavation and survey. It also discusses some advances made in archaeological photography. Finally, it presents preliminary results from on-going experiments with automatically generating Harris Matrices from a FileMaker Pro database and with using iPads and iPhones as GPS units for survey.

Keywords:

Recording, iPad, Paperless, Photography, Database

1. Introduction

The Sangro Valley Project (hereafter abbreviated SVP) was founded in 1994 and is now managed by Oberlin College in collaboration with the Soprintendenza per i Beni Archeologici dell'Abruzzo and the University of Oxford. The project operates a summer field school in Italy for Oberlin and other students; it employs a multi-disciplinary team of specialists from Canada, Italy, the United Kingdom, and the United States. The project's goal is to characterize and investigate the nature, pattern and dynamics of human habitation and land use in the *longue durée* within the context of a Mediterranean river valley system—the Sangro River valley of the Abruzzo region of Italy, the territory of the ancient Samnites (<http://www.sangro.org>).

Over its first sixteen years the SVP employed various formats to record, store, manage, and analyse its data, with recording methods based on the Museum of London Archaeological Service's site manual (Museum of London 1994). The opening of a new site in 2011 provided an opportunity to rethink the project's data systems. The Pompeii Archaeological Research Project: Porta Stabia, directed by Professor Steven Ellis of the University of Cincinnati, pioneered the use of Apple's iPad in 2010 for paperless recording of basic excavation data as well as for drawing and other functions (Wallrodt and Ellis 2011; Porta Stabia 2011). Building upon their success, the SVP developed an integrated paperless recording system in FileMaker on both laptop

Corresponding author: motzcf@mail.uc.edu

computers and iPads. The paperless system pushes digitization of data into the field—this replaces traditional recording using paper forms, followed by subsequent transcription into computer systems, with direct data entry into the digital database (an unlocked public version of this database has been made available at www.paperlessarchaeology.com to assist others in developing similar systems).

Such a system has many benefits. In addition to allowing for computerized analysis, it makes field data easier to access and eliminates errors introduced when interpreting handwriting. When dealing with paper records, the number of different forms can be overwhelming and confusing. This, coupled with the sheer volume of records, can make finding and organizing data during the season quite difficult.

2. Project Infrastructure

As a regional project, the SVP does not excavate at a single site. Instead, excavators move from site to site; the duration of study at each site depends on the amount of time required for a proper investigation. The SVP also employs survey and various other methods of data collection. Therefore the project's infrastructure needs to be mobile and flexible, and researchers cannot count on having access to anything other than what they bring into the field. Among other things, this means that all field equipment must run on batteries and must stay useable for at least 6 hours before needing to be recharged. Cellular 3G coverage in the area is

quite unreliable, so internet access is assumed not to exist in the field. The SVP operates a dig house with temporary lab facilities located in the town of Tornareccio. The project's computer lab, located in this building, operates on a local network with both wired and wireless access.

3. The SVP Database

The SVP has used paper, Microsoft Access, Excel, fillable PDFs, and numerous other data formats throughout its history, but had not employed a fully integrated database until the 2011 season. This presented both challenges and opportunities when developing the paperless system. On one hand, recording practices could be completely redesigned with a paperless workflow in mind, and the database was designed to be used on iPads from the beginning. On the other hand, the lack of a pre-existing database meant that an entirely new system needed to be developed very quickly, with relatively little time for testing. Fortunately, the database worked mostly as planned and the project was able to work through minor bugs without much disruption to the excavation and survey teams.

3.1 Overview

Data about each context, small find, environmental sample, and field survey unit (among other record types) were captured in the field using FileMaker Go on iPads. Specialists in the labs then entered more detailed information about small finds, pottery, tile, and a handful of other items into the FileMaker Pro database hosted on the SVP's local Mac Mini server. In addition, an iPad with FileMaker Go was given to each of two survey projects: one that was mapping and gathering data on agricultural terraces (see Countryman, Carrier and Kane 2011) and another that was field walking in the area around the excavation site.

Each iPad operated independently, since there was no local network in the field; data could not be exchanged or synchronized between devices until they were brought back to the computer lab during lunch and at the end of the day. This necessitated the use of only one iPad per trench for entering information into the database. Each trench was also assigned one iPad for a field notebook, drawing, and photography. For the same reason, each survey



Figure 1. iPad database home screen, showing colour-coded record types.

team was assigned only one iPad for database entry. To communicate between the iPads and the main database hosted on the Mac Mini server, the SVP used the syncing method and scripts designed by John Wallrodt from the University of Cincinnati, with great success (Wallrodt 2011a; 2011b).

3.2 User interface

The decision was made early in the development process to put significant effort into the database's user interface, as the developers had become quite aware of the shortcomings of the cluttered paper forms used in previous seasons. It was determined that the user experience would be far from merely cosmetic, and would have a major impact on the efficiency of excavation and survey, the quality and consistency of the data entered, and the degree of acceptance of the new system from the research team. Therefore, the database was designed to present all of the needed information in a way that was both understandable and usable.

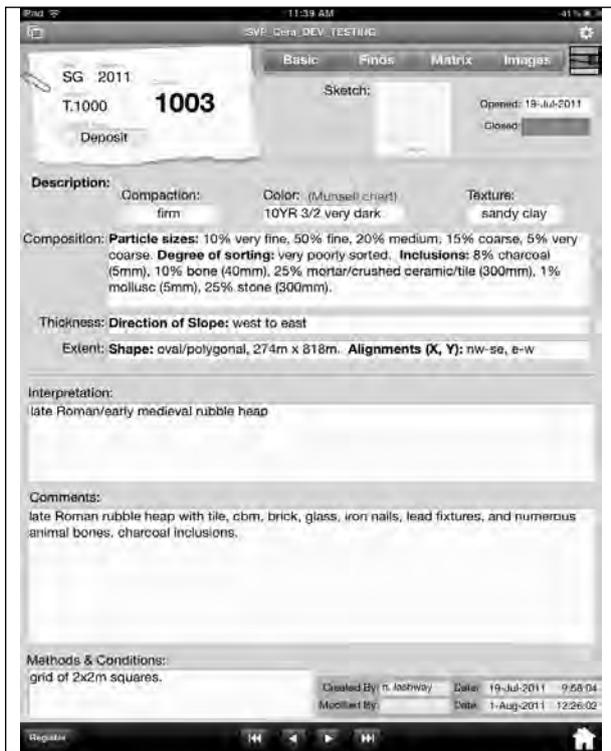
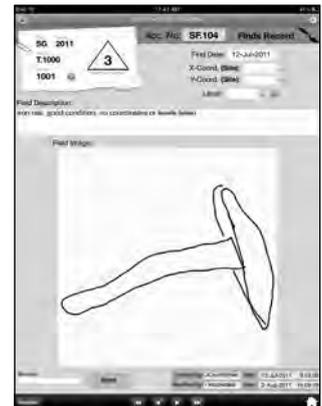


Figure 2. Main context record form with tabbed navigation.

To be understandable, the database needed to have an intuitive layout and navigation. In addition to the careful arrangement of fields on the screen, this was achieved first through colour coding. Different record types were identified in this way (contexts, survey forms, finds), and these colours were kept consistent across the entire database (Fig. 1). In addition, flags and other identifiers were colour coded. The second principle was minimizing scrolling, which was driven by a desire to avoid situations where fields did not get filled out because excavators did not realize there were more items off-screen. This meant that if all of the fields did not fit onto one screen, tabs or buttons were used to direct a user to a different screen instead of extending the form beyond the display (Fig. 2). Forms were greatly simplified from their paper predecessors, and included only what each user needed to see: rather than displaying all fields as on the earlier forms, the appropriate data entry fields were displayed according to the context type (deposit, masonry, or cut); excavators and finds specialists needed to see different amounts of information on small finds, so only the essentials were provided to the excavators

Figure 3a. Right. Small Find form (iPad layout for excavators). 3b. Below. Small Find form (computer layout for specialists).



while the Specialists could enter much more information (Figs 3a and 3b).

To be usable, the database also needed to be easily readable in sunlight, and each button or other actionable object needed to work the first time a user tried to interact with it. To deal with the sunlight, text was rendered in a large font size with special emphasis placed on the most important text; high-contrast colour palettes were also employed. Buttons, checkboxes, and fields were large, increasing the likelihood that they would be accurately tapped on the user's first try; this was especially important when dealing with rows or columns of checkboxes. Automation also improved the database's usability and efficiency; by automatically filling fields based on previously entered or calculable information, the project reduced the number of items excavators had to enter and increased the accuracy of many routine fields.

4. Photography

Image tagging had long been one of the least efficient areas of the SVP's workflow. Previously,

photos of the site taken during excavation were captured with digital cameras. At the end of the day, the excavators uploaded the digital images to a computer, labelled and captioned the files, and stored them on the project's local server. In the best of circumstances this meant that an image taken early in the morning might be tagged 9-10 hours later, likely after numerous other similar photos had been taken.

For the 2011 season the captioning process was moved out into the field. One of the key pieces of technology that enabled this was the Eye-Fi Connect X2, an SD camera memory card with built-in Wi-Fi. This technology enables direct communication between the cameras and iPads, allowing photographs to be immediately labelled and captioned in the field, and enabling a significant improvement in the accuracy of photo captions. The Eye-Fi card allowed excavators and surveyors to continue to use dedicated digital cameras, which currently produce images of a much higher quality than the cameras built into many tablets and mobile devices. Using Eye-Fi's Direct Mode, each card and camera was paired with one iPad. After a photo was taken, the card automatically broadcasted a Wi-Fi network to which the iPad connected. The card then transferred images to the iPad, putting the photos directly into the Photos app. The Eye-Fi app on the iPad performed the actual transfer, but it can run in the background. The entire process took from 30 seconds to 2 minutes, depending on the number of photos being transferred. Once the images were transferred, researchers began the process of adding captions and labels to the images' metadata using the Photosmith app on the iPad (Fig. 4). The SVP used only two metadata fields: *title* and *caption*. The *title* field was used for a descriptive subject code, while *caption* was for standard plain-text descriptions.

After the images had been tagged and the iPads had returned from the field, the images were then uploaded to the server. This was done by syncing Photosmith with Lightroom using Wi-Fi. Once the images were transferred to Lightroom, any necessary editing was performed and duplicates were removed. Next, a FileMaker Pro script was run to import the new photos into the database. The scripts created a thumbnail and ran some other actions on each file, then imported them into FileMaker Pro using the Troi File plugin. One of the



Figure 4. Screenshot of the Photosmith app, showing the fields used for labelling images.

key features of this plugin is its ability to import metadata along with the image itself. This allows the database to not only insert the metadata into fields within FileMaker so they can be read by users, but also create and manipulate other data based on the information contained in that metadata. For example, the SVP's scripts took the subject code that was entered earlier and parsed it in order to automatically link the photo with its subject. With the file renaming and relocating abilities of the Troi plugin, an image file could then potentially be renamed automatically according to its subject and moved to a particular folder on the server for that subject, a step which the SVP hopes to add to the database in time for the 2012 season. By pushing the tagging process out into the field and using structured tags, the image processing and importing workflow can potentially become much easier, with little or no input required by the user beyond the initial tagging in the field.

5. Other Mobile Apps

In addition to FileMaker and Photosmith, the SVP used a handful of other iPad apps to assist with field recording. Field notebooks were written with Apple's Pages, which allowed excavators to integrate both drawings and photos into their accounts. Several compass, calculator, and ruler utilities were tested in place of their more traditional counterparts, and Plaincode's Clinometer app proved particularly useful to the terrace survey team in measuring the approximate angles of slopes.

6. Alternative GPS Units

In the summer of 2011 two experiments were conducted with the goal of exploring the potential of the SVP's hardware and software. The first was a series of tests of the GPS chips that are built into the iPhone and the iPad Wi-Fi + 3G models (the Wi-Fi-only iPad models do not contain a GPS chip). The major shortcoming of GPS apps for iOS is that they do not provide as much detailed information about accuracy as a dedicated GPS unit. For example, the SVP's handheld Trimble GPS units report the number and configuration of satellites in the sky, while the iOS apps only report accuracy in terms of +/- X meters. Because all non-jailbreak GPS and GIS apps receive their location data through iOS's Core Location framework, all apps are essentially equal in the accuracy of their recorded coordinates, and no app will have access to detailed information on satellite configurations (Apple Inc. 2011). Apps differ greatly, however, in how they display and organize this data.

The experiments that the SVP conducted have shown that the iPhone and iPad can consistently achieve 5-20 meter accuracy under a variety of conditions in Italy's Apennine mountains. Professional GPS units can provide greater accuracy, but they are expensive, often require significant training to use properly, and do not always function as well or as consistently as one might like. The benefits of inexpensively equipping a large number of team members with iPhones or iPads may offset any decrease in GPS accuracy. While each project must take its own needs into consideration, the researchers determined that in some cases these devices would be sufficient for basic survey work.

7. Generating Harris Matrices

The second experiment was an attempt to automatically generate Harris Matrices from a FileMaker Pro database. The researchers wanted to create a system that would take the *physical* relationships recorded in the field and would produce a clean and accurate *stratigraphic* matrix. It was determined that creating such a system entirely within FileMaker Pro would be extremely difficult or impossible given the non-dynamic nature of FileMaker layouts; therefore, integration with additional software was needed. Stefano Costa's

method for using Graphviz's "tred" command to create clean Harris Matrices from manually entered physical relationship data provided the core of the workflow (Costa 2007). The first step of the system was a FileMaker script that searched for all relationships within a given trench and created a *.dot* file containing all of the physical relationship information. This file was then automatically passed through one of the Graphviz components, a UNIX command called "tred," an abbreviation of "transitive reduction." This command removes all of the extra connections that are introduced when graphing the physical relationships and reduces the graph to only the stratigraphic relationships. The key advantage to incorporating such a system into a FileMaker database was the ability to press a button, wait a few seconds while the script ran, then be presented with a fully editable Harris Matrix in OmniGraffle. This matrix can then be exported as an image file and easily added to the trench record in the database.

8. Cost

One of the drawbacks to much of the advanced technology used by archaeologists is that it can be very expensive. One of the goals of this project was to keep costs low, in the hope that whatever system the SVP developed would be affordable for other archaeologists to implement. To replicate the SVP's setup, the total cost would be around \$5,000. The cost of FileMaker Pro can drop considerably or be eliminated entirely through volume pricing and site licenses held by many educational institutions. These prices also assume the purchase of previous-generation iPads at a discount, which the researchers found worked just as well as the newest model.

9. Conclusions

The technology and systems described above quickly proved to have many advantages over traditional recording methods—much quicker exchange of information between the field personnel and specialists; immediate labelling and captioning of photos taken in the field; a significant decrease in human error through automation; improved consistency of terminology, by using pull-down menus and other structured fields; increased efficiency by eliminating the need to scan and digitize paper records; an increase in the accessibility of

information to all staff members; and improved security of field data due to twice-daily syncing and backup.

The paperless system proved to be a resounding success. It was used for excavation, two survey projects, and recording by specialists. While there were some growing pains, the benefits far outweighed the costs. For any large archaeological project, data organization is critical. The flexibility of both the hardware and software allowed the SVP to finally integrate several types of research into a single, cohesive database. This approach has enormous potential to revolutionize the way archaeological data is collected, managed, analysed, and disseminated.

10. Future Work

Future efforts will focus on the next stages of implementing paperless recording systems, namely the simplification and refinement of a successfully tested system. The system used in 2011 was highly effective, but required constant attention from the database developer. Further simplification and streamlining of field recording procedures will lessen the burden on excavators, and a reduction in the maintenance and development workload will allow the system to be managed by a non-specialist. The desired outcome is a recording system that can be employed by archaeological projects that do not have constant access to a developer. In addition the user interface will be completely redesigned with the goal of providing a more intuitive, consistent, and simplified experience to the researchers.

Acknowledgements

This project was made possible by a fellowship grant from the Concordia Foundation. Many thanks to Sam Carrier, Alexis Christensen, James Countryman, Beatrice Fidelibus, Susan Kane, Nick Lashway, and Chris Noon for their assistance in developing and testing the database; to all the staff and field school students for their assistance, support, and helpful feedback; to John Wallrodt, Steven Ellis, and the team at PARP:PS for their

valuable technical help and advice; to the Sangro Valley Project and Oberlin College; and to Sam Carrier and Susan Kane for supporting this project.

References

Apple Inc. 2011. "Location Awareness Programming Guide: Making Your Application Location-Aware." Last modified October 12. <http://developer.apple.com/library/ios/#documentation/UserExperience/Conceptual/LocationAwarenessPG/Introduction/Introduction.html>.

Costa, S. 2007. "Harris Matrix with Graphviz." *IOSA.it*, December 18. <http://www.iosa.it/content/harris-matrix-graphviz>.

Countryman, J.R., S.C. Carrier, and S.E. Kane. 2011. "Predictive Modelling: a Case Study of Agricultural Terraces at Monte Pallano (Abruzzo, Italy)." In *Revive the Past: Proceeding of the 39th Conference on Computer Applications and Quantitative Methods in Archaeology, Beijing, 12-16 April 2011*, edited by M. Zhou, I. Romanowska, Z. Wu, P. Xu and P. Verhagen, 307-317. Amsterdam: Pallas Publishing.

Museum of London. 1994. *Archaeological Site Manual*. London: Museum of London.

Porta Stabia. 2011. "iPads at Pompeii." Last modified April 12. <http://classics.uc.edu/pompeii/index.php/news/1-latest/142-ipads2010.html>.

Wallrodt, J. 2011. "Syncing Sample." *Paperless Archaeology*, May 27. <http://www.paperlessarchaeology.com/2011/05/27/syncing-sample>.

Wallrodt, J. 2011. "That's why it's called a beta." *Paperless Archaeology* August 20. <http://www.paperlessarchaeology.com/2011/08/20/thats-why-its-called-a-beta>.

Wallrodt, J., and S.J.R. Ellis. 2011. "The Paper-less Project: The use of iPads in the excavations at Pompeii." Paper presented at the *39th Conference on Computer Applications and Quantitative Methods in Archaeology, Beijing, China, April 12-17*.

'Tangible Pasts': User-centred Design of a Mixed Reality Application for Cultural Heritage

Angeliki Chrysanthi, Constantinos Papadopoulos, Tom Frankland and Graeme Earl
University of Southampton, United Kingdom

Abstract:

Tangible Pasts is a Mixed Reality (MR) prototype that combines a tangible interface with virtual content. The physical object, a book, enables users to manipulate the virtual models in real time by their actions performed in the physical world. This paper presents an overview of designing the prototype, and the evaluation of this application by participants who tested its usability and interpretive value. It also presents some preliminary results of the second round of design and evaluation. The attention is stressed on the importance of performing constant user evaluations in parallel to the design of the prototype in order to produce an application that is usable and of benefit to archaeological research and interpretation of cultural heritage knowledge.

Keywords:

Tangible Pasts, Mixed Reality, Iterative Design, User Evaluation

1. Introduction

This paper presents the preliminary results of a tangible prototype interface for cultural heritage dissemination in a Mixed Reality (MR) environment. This research project was developed in parallel to our research work, which is related to several aspects of computer visualisation, interpretation and presentation of archaeological material, with the purpose to explore available MR platforms from a human and user-centred perspective. The design of this application, which utilises available technologies, was structured around three main research questions: a) How can we provide an intuitive user-friendly application for cultural heritage, which blends virtual imagery with the actual world, where users operate and interact with the information? b) How effectively can archaeological information and visual interpretations of the past be disseminated through such technologies? c) How can user evaluations in early stages inform archaeological interpretive design? Based on the principles of iterative design the prototype was designed, evaluated, analysed, refined and re-evaluated. This cyclic process will continue until any issues identified by participants in this research will be kept at a minimum. This project will also employ new technologies as they become available to the public, attempting to smoothly combine the natural process of reading with MR content.

2. Background and Objectives of Tangible Pasts

A significant number of collaborative projects have already started to exploit the real potential of MR technologies and develop new methodologies in interpretive archaeology, cultural heritage sites and museums (Noh et al. 2009; Champion 2011). A considerable amount of work has also been undertaken on user evaluations for assessing the interpretive value and the technological development of applications, especially in museum contexts. Some of the most significant lessons learnt in such visitor evaluations is that the interactivity of systems operating in MR environments scores highly (Danks and Rodriguez-Echavarria 2007) and that novel interfaces can only become effective learning tools if the interface is easily understood by the users (Economou and Pujol-Tost 2007). However, there is little work on user evaluations for informing the actual interpretive design from the early development stages, despite the fact that such evaluations have admittedly played a crucial role for future revisions both on the interactivity of systems and their interpretive content (Danks and Rodriguez-Echavarria 2007).

In attempting to address this issue, we draw on the design methodologies from disciplines such as Human-Computer Interaction (HCI) and Computer Supported Cooperative Work (CSCW), where iterative design is adopted throughout the phases

Corresponding author: a.chrysanthi@gmail.com

of prototyping, testing, analysing, and refining a product or interface (Rogers et al. 2006). In those fields, iterative design is mainly utilised for testing design ideas and interactions with a system prior to developing its hardware and software components; a notion which was pointed out in the early surveys of Augmented Reality (AR) research due to the inherent technical complexities of implementing such applications (Azuma et al. 2001). Paper and more recently video prototyping (de Sá et al. 2011) have been used to inform the design of AR applications. In this project, the research objectives were mainly concerned with the design of heritage content and users' responses to the modalities provided by the system for interacting with CH information rather than the system itself. For this reason, we used an AR commercial platform and prototyped the tangible interface to test some initial ideas. Drawing from previous work on MR and the intuitiveness of Tangible User Interfaces (TUI), such as the 'Magic Book' described in Billingham et al. (2001), this study explores an alternative and interactive mode of engaging the public with archaeological information. It focuses on a multimodal approach for enhancing Cultural Heritage (CH) experiences by combining the interpretive nature of physical books with virtual CH content (Wright and McCarthy 2010).

Additionally, this work set out to investigate the benefits of exploring archaeological information in a MR environment and, to define and assess this environment in the virtuality continuum (Milgram and Kishino 1994). In this continuum, according to Milgram and Kishino, real environments are shown at one end and the virtual ones at the opposite. Augmented Reality and Augmented Virtuality (AV) are placed within the general area of MR, which is defined as the environment in which 'real world and virtual world objects are presented together within a single display, that is, anywhere between the extrema of the virtuality continuum'. Renevier and Nigay (2001) provided a clearer distinction between AR and AV systems. According to their definitions:

- In AR, interaction with the real world is augmented by the computer.
- In AV, interaction with the computer is augmented by objects and actions in the real world.

Bearing in mind the above terminological discussion, this paper will define the environment of this application and demonstrate why from an interpretative design point of view it is important to choose the suitable environment for each application (Drascic and Milgram 1996).

3. Description and Content of 'Tangible Pasts'

'Tangible Pasts' consists of 12 pages, which combine textual and visual information, three-dimensional models and animations in a single tangible interface, enabling the user to experience two prehistoric sites of the Minoan period, by turning the pages of the book and seamlessly moving between physical and virtual content. The book is divided into two sections: The first one regards the Minoan settlement at Zominthos, Crete; the three-dimensional models and the renderings presented in this section were produced as part of Papadopoulos' research project (Papadopoulos and Sakellarakis 2013) to simulate the impact of natural and flame illumination in the ceramics workshop of the building. The first page provides users with basic textual information for the archaeological site and the ceramics workshop, while the second includes two renderings of the settlement and the workshop as well as the marker that triggers the AR content (see methodology). This model employs basic illumination and textures due to the limitations of the AR platform used. However, it does not pose any constraints in understanding space and providing a sense of materiality. The third page provides a comparison between a photograph in the current state of preservation and the corresponding 3D model, while the fourth focuses on the potter's wheel with both textual information and an AR model of a potter's working bench, also including an animation of the turntable (Fig. 1).

Also, a digital museum for the finds unearthed in the ceramics workshop was constructed (Papadopoulos and Kefalaki 2010) in an attempt to work on different display combinations for the actual museum that will be erected in the future to house the finds of the excavation. In these two pages users can read about and also see in AR two special objects. The second part is related to Gournia, a Minoan town in eastern Crete, which is examined by Chrysanthi in relation to visitors' movement

and circulation patterns in archaeological sites. In this case as well, the principles of presentation are the same: basic textual information, photographs related to the site, and three-dimensional models, which were constructed by Frankland based on the principles of procedural modelling.

4. Methodology

4.1 Design

Our design methodology was based on the principles of iterative design (Buxton & Sniderman 1980). The 'Tangible Pasts' prototype was developed, tested and refined in order to improve its quality and functionality based on observations and users' evaluations. The prototype consists of a tangible element, a book, and several intangible elements in the form of three-dimensional models, animation, video and sound. As far as the physical element is concerned, we attempted to create a usable book that would facilitate reading, as well as the presentation of the textual and visual information and the interaction of the users with the virtual content in a seamless manner. During the first design cycle, we used A5 sheets, devoting half of the pages to textual data and half to visual information and markers. The typeface used was Papyrus, intending to create a readable calligraphic hand-crafted look that fitted the textured paper used to print the content. Texts' length was 170 words on average attempting to provide the basic information needed to contextualise the AR content presented. As far as the visual content is concerned, most pages included either two images of similar size or one larger image and the relevant marker. Regarding the latter, the limitations of the software used, made us employ preconfigured black and white markers, which are obtrusive elements in the process of reading.

Concerning the technical aspect of the prototype, 3dsMax 2012 was used for the production of the majority of the models and animations, while CityEngine was implemented for constructing alternative structural models of the Minoan houses at Gournia. The AR-Media 2.0 plugin for 3dsMax was employed to assign markers to the virtual models and configure the scenes according to the page layout of the book. The prototype was initially tested by ourselves within the available functionalities of the plugin, while at the stage of user evaluation,

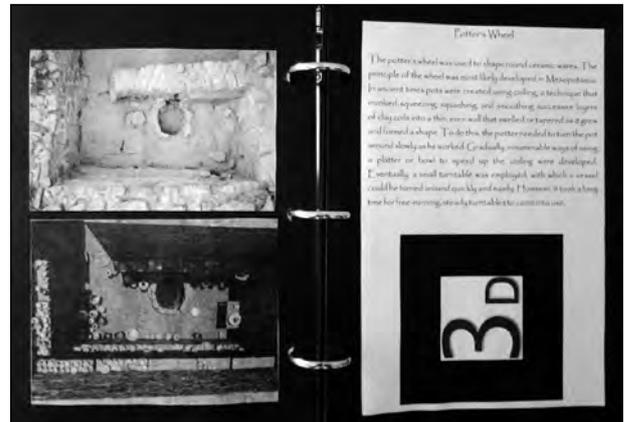


Figure 1. Pages 3 and 4 of the physical book.

the application was exported and presented by the standalone AR-Media Player developed by Inglobe Technologies. As far as the hardware is concerned, the workstation used had an Intel Xeon 2.67GHZ Quad CPU Processor, 512MB QuadroFX580 NVidia Graphics Card and 12GB RAM. The AR content was triggered by using a 1.3 megapixel webcam with auto-adjustment for low lighting conditions, positioned on a stand at the level of the head of a seated user, several centimetres behind his/her back. Lastly, the AR content was projected on a plasma screen and a keyboard was used for basic interactivity with the models.

Although AR-Media plugin could handle textures and a significant amount of polygons, the high-fidelity models already produced for another purpose had to be optimised for the needs of this prototype and the requirements of the development platform. All the physically accurate Mental Ray materials were replaced by standard textures, while the polygons of the scenes were significantly reduced, not only by optimising objects and surfaces, but also by removing unnecessary details and high-polygon objects. For example, in the ceramics workshop model, the amount of pottery was limited to such an extent, so as the reader/viewer would be able to understand the function of this space and the arrangement of objects within it. Also, an abstracted version of the architecture of the building was also created. In the case of Gournia houses, the models were produced particularly for this reason by employing procedural modelling, a script-based technique that uses 'shape grammars' to semantically describe 3D models (Müller et al. 2006). Based on this technique, we wrote a

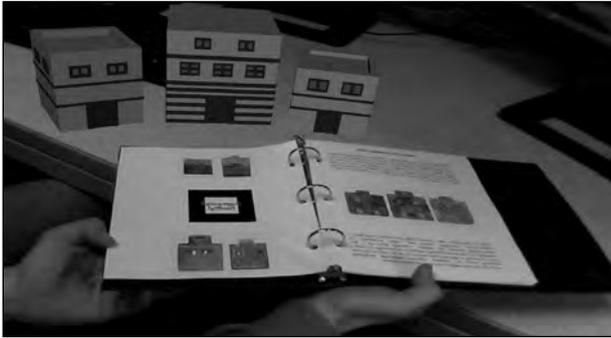


Figure 2. Augmented scene with an interpretation of the Minoan houses generated by CityEngine.

script that rapidly generated numerous 3D models that depicted differing levels of certainty in the archaeological data about the configuration and morphological features of the structures. The shape grammars were created based on construction rules and materials of Minoan domestic architecture (Preziosi 1983, Palyvou 1999, Shaw 2009). To present the different types of 3D models generated we used a pop-up animation approach where consecutive models pop-up from the pages and sink back into the pages giving way to alternative versions (Fig. 2).

4.2 Implementation and user-evaluation

It has been argued that the involvement of both expert-users and real-users in such early stage evaluations is 'the most promising approach to evaluate interfaces of a complex kind, such as cultural heritage virtual interfaces' (Karoulis et al. 2006). The evaluation of the prototype took place during the Open exhibition as part of the Visualisation in Archaeology Conference (ViA) at the University of Southampton (Fig. 3). ViA conference was an ideal place for conducting this evaluation, since a great number of students, scholars and professionals in archaeological visualisation were gathered during the course of this event, therefore providing a sufficient and varied in terms of expertise sample for testing the functionality and the overall value of the prototype (Fig. 2). The evaluation was based on questionnaires, observations and discussions with the participants. The purpose of the evaluations was to test the affective response of users to the tangible interface and their interaction with the MR cultural heritage content in order to evaluate our initial design. We were also concerned with the possible obtrusiveness of the virtual elements over the



Figure 3. Instance from demonstrating and conducting user evaluations of Tangible Pasts' first prototype at the VIA Conference.

physical and *vice versa*, trying to use the feedback to improve the design by leveraging physical and digital elements into a seamless experience.

The questionnaire was divided in three sections. In the first section, the participants were asked to give a general assessment of the prototype by describing the things that they liked the most and the least about this project. It was on purpose that we did not provide any predefined answers, so as to let users express their thoughts and describe any limitations or advantages as they see fit. For the qualitative analysis of our data we used a thematic analysis approach, which we found extremely useful and flexible. In the next two sections the users were asked to rank the usability of the prototype and its overall interpretive value. The last section about the demographics of the sample was also important to assess how different age groups, professions and people of varying education perceive such approaches to cultural heritage. All these three sections were necessary in order to draw some quantitative data for our research.

In parallel to the questionnaires, we observed users' interaction with the prototype. Observing a user's interactions with a prototype in a lab study is a typical HCI evaluation methodology. Although we did not employ these methods rigorously to evaluate our design against certain usability criteria, we did focus our attention on understanding how users chose to manipulate the book, how they responded to the triggering of the AR content and how their attention varied between the textual information and the analogue and virtual visual content. After testing the prototype, we conducted informal

interviews with several of the participants, during which they freely described their experience and expressed their concerns. This also allowed us to follow up on the behaviours we observed the users engaged in, helping us to determine their intentions and whether our prototype fulfilled them.

5. Results of the User-Evaluation

The first user evaluation questionnaire was completed by 22 people ranging from undergraduate students to professionals in various fields of archaeological visualisation, with their ages ranging from 20 to 65. Although the sample was relatively small to draw safe conclusions about several aspects of this prototype, it was adequate to perform a prototype evaluation and inform the next stages of the design.

The users were asked to mention three things that they liked the most and three things they liked the least about the project. From the first question we arrived at six principal themes representing participants' positive comments and revealing the strengths of the Tangible Pasts prototype. People commented positively on the interactivity that the prototype provides, mainly because they were able to move the 3D models in real time with their own hands, play with the animations and the 3D models giving them the ability to interact with the system and its content. This issue brings us to the second point, the tangibility of the interface. For example, Participant #19 commented: 'I'm the kind of person who wants to touch everything and therefore it's fun!' while quite a few mentioned that the physical book was an intimate object and friendly to the user who required no instruction of how to use the system.

In regard to blending real with virtual interpretive elements, participants also pointed out that they liked the fact that the prototype was not entirely computerised since it combined physical and virtual content. We also received many comments including words such as 'mysterious', 'intriguing' and 'magic', referring to the intuitiveness of the system to provide interpretation in an unexpected manner making the whole experience 'engaging', 'playful', 'entertaining' and 'feeling like an exploration'. In terms of the technology employed, participants liked the way technology was appropriated to

assist and stimulate interpretation via a common interpretive medium, a book, and referred to the use of technology as 'intelligent', 'innovative' and 'unexpected' having a 'low tech meet high tech feel'. In terms of the actual interpretive content, people commented on particular aspects of the 3D models and their interpretive value, the alternative reconstructions available and the fact that models were accompanied by textual interpretation in case someone wanted to explore in detail the archaeological sites or finds.

On the other hand we obtained a valuable insight into the constraints and flaws of the prototype in terms of its interactivity, design, technology and the actual interpretation provided. A serious issue picked up by the analysis of the questionnaire, as well as our observations, was that people tended to focus more on their interaction with the 3D models and ignore the textual and visual information provided in the book. This also explains participants' suggestions to use less text and bigger fonts, as well as to incorporate descriptions and explanations within the virtual display. The interpretation of this observation as a manifested tendency, especially in younger participants, to replace any analogue means of knowledge communication with more interactive and pleasing digital formats would only present one side of the same coin. The fact that the book contained a significant amount of analogue content was a design decision, since one of the main purposes of creating this interface was to shift the attention from the virtual and digital content to the physical environment, where users operate with the system while constructing their interpretations. This decision however, was not leveraged by the implications of the component technologies used to implement this application. More specifically, the use of the monitor as the display technology resulted by default in the shift of attention to the virtual environment (the monitor display) despite the fact that it depicted elements from the physical environment. The physical book and the user's actions in the physical space are merely used to interact with the system (Renevier and Nigay 2001). Thus, it could be argued that the interaction takes place in the middle of the virtuality continuum, in a MR environment which is difficult perceptually to define whether it stretches towards AV or AR. Nevertheless, had we used AR goggles as a display device our design decision would have been more

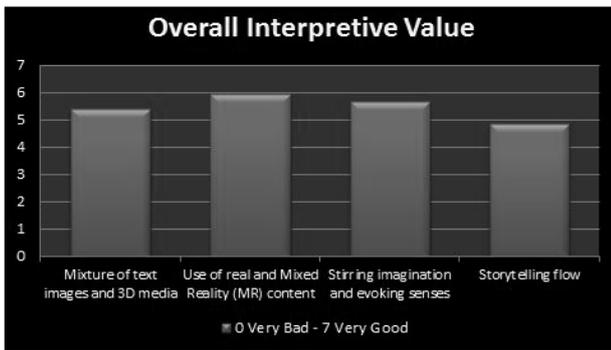


Figure 4. Table with the results of the evaluation on the prototype's interpretive value based on four criteria: 1. Effectiveness of mixed media cultural heritage information dissemination, 2. The use of real and MR content, 3. Stirring imagination and evocation of senses, 4. Assessment of the storytelling flow.

justified and effective in shifting the attention to the physical interface and the physical environment where the user handles a book with MR content.

A few participants with a background in computer graphics commented on the graphical/virtual design approach mentioning the simplicity of some models. However, as mentioned previously this was not a design issue, but a technical one since current available AR platforms cannot support high fidelity models with physically accurate lighting and textures. Because of that, we decided to exploit this technological limitation to evaluate people's responses to such 3D representations and address the long-debated discussions about photorealism and non-photorealistic rendering in archaeological interpretation (Frankland 2012). Finally, in terms of the interpretation we received some suggestions about how to enrich the narrative and include scenes from everyday life in a Minoan settlement.

Some of the participants pointed out the importance to populate the virtual scenes with people and action in order to contextualise the sterilised and static interpretations.

Participants were also asked to evaluate the overall usability of the prototype both in terms of interaction and its effectiveness in cultural heritage dissemination. More than half of the sample stated that they found it very satisfying, while a significant percentage said that it was extremely satisfying as an interactive tool that can improve the dissemination of cultural heritage knowledge (Fig. 5).

It is also interesting to mention that men were in overall more enthusiastic than women and that only women answered that they found it neither satisfying nor disappointing. Lastly, it was important to canvass their responses regarding the interpretive values of the prototype (Fig. 4). The use of real and MR content, as well as the stirring of imagination and the evocation of senses was rated with 6 out of 7, while the mixture of text, images and 3d media and the storytelling aspect was rated lower with 5 out of 7. As this prototype was mostly visual, since tactility existed through the book itself, but not actual tactility related to the archaeological features shown, it is quite odd that a great number of participants ranked quite high the imagination and the arousal of the senses. As far as the mixture of the various media is concerned, the evaluation with 5 out of 7 is reasonable as the text that accompanied the illustrations and the markers was quite long, and only few people read it through.

Our observations, also confirmed that participants were discouraged by the amount of textual information in the book. The majority of

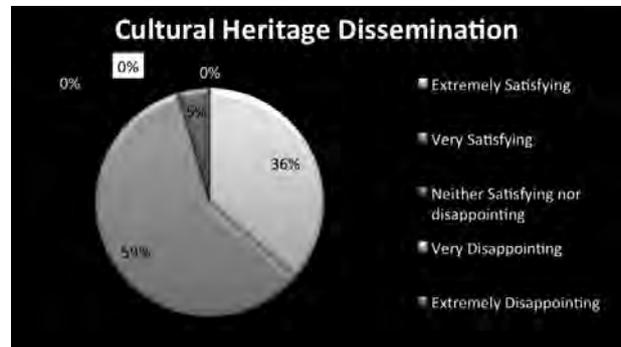
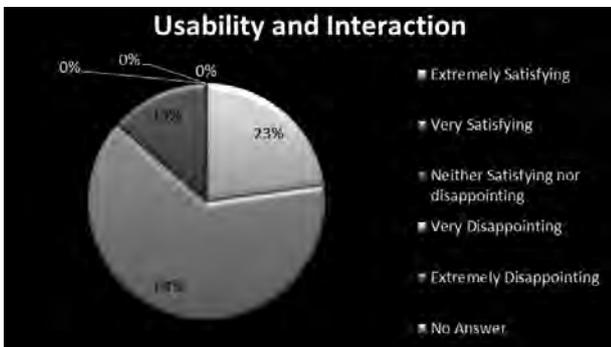


Figure 5. Tables with the results of the evaluation on the prototype's a. usability and interaction b. ability and effectiveness in disseminating cultural heritage content.

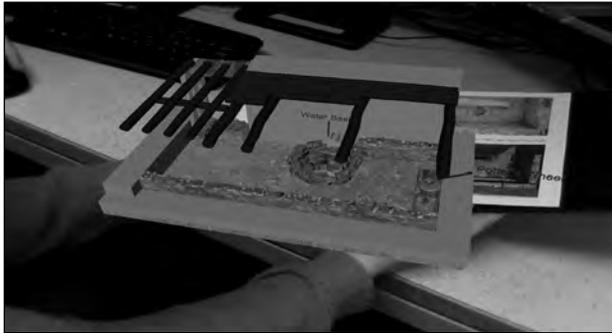


Figure 6. *The augmented scene of the workshop with added annotations.*



Figure 7. *The augmented scene of the spinning potter's wheel is enriched by the video of an ethnographic parallel of Cretan traditional pottery making.*

them were reading the first couple of sentences, while others were skipping the longest pieces. This was especially apparent for the pages where the virtual content was more impressive than the rest. In overall, it was noticed that younger people were more hesitant to read the analogue interpretations, due to their fascination by the technologies employed and the monitor display, while older participants were spending equal time to read the content and interact with the models, possibly because of their greater familiarity with more static pieces of information.

6. Refining and Re-evaluating the Prototype

The first user evaluation, as well as the observations and the discussions with the participants provided useful comments for the evaluation of our application in this initial stage. Their input was used for the improvement of the prototype in the second stage of development.

Participants' positive responses concerning the ability of the application to stimulate the

imagination and evoke the senses were a valuable find. As a next step we included sound, both in terms of narration and the sounds produced from the various features (e.g. potter's wheel), contributing to a more sensory engagement with the datasets, as well as enhancing the storytelling flow. In addition, we significantly reduced the textual information within the book, which was replaced by virtual annotations explaining various features in the 3D models (Fig. 6).

Finally, we incorporated a video presenting the process of traditional pottery making, running parallel to the potter's wheel 3D animation (Fig. 7).

The triggering of the markers, which were reduced in size to limit their visual obtrusiveness, is still implemented by the use of a web camera and a computer monitor, especially due to the high cost and the inherent limitations of new head mounted devices, i.e. goggles that currently exist in the market. In the next stage, we are going to implement our application with other platforms that enable image reference AR and use the actual images included in the book as markers, since they were only serving illustrative purposes in the application's current configuration.

We have already started testing the new version of the prototype to a different audience to see the responses of people that are not by definition fascinated by any means of archaeological visualisation. The second call for the evaluation of *Tangible Pasts* was made to everyone within the University of Southampton and was not limited to archaeology specialists or humanities students. Although it is quite early to quantify the results of this second round of evaluation, some of the participants provided valuable comments, which are briefly discussed below.

The majority of them positively commented on the newly introduced features of sound and video. They particularly liked the mixture of different media and the combination of conventional sources of information (e.g. sound) with virtual content. An interesting comment that was not pointed out in the previous evaluation was that the interaction with the book and its manipulation in front of the camera provided a sense of three-dimensionality, perspective and depth which is not perceptible by

using conventional means of representation. This may be due to the seated position of the viewers that allowed a more comfortable posture to move the book, in contrast to the first evaluation, where the configuration of the exhibition space only allowed a standing position, resulting in some difficulty to manipulate the book in relation to the camera and the computer monitor. However, it should be noted that a few of them found that the angle between the book and the camera as well as the constant adjustment of hands and posture in order to trigger the markers was uncomfortable and rather complicated. This also accrued from our observations during which we realised that users were often losing the markers from the camera's field of view and as a consequence the AR scene was starting over. This was rather annoying for the narration and the video content that had to restart each time. The difficulties that participants faced were expected, since they had to balance the familiar practice of reading with a completely unusual process of moving the book and the body to trigger the markers and make the rest of the content appear. These problems would have been overcome by employing AR glasses, which would make the objects appear on the book and the users to follow the familiar process of reading without having to adjust their position to enable the virtual content. Therefore, the digital workspace would turn into a physical locus.

Our observations also pointed out that all participants were reading through the information in the book, in contrast to the first evaluation in which participants showed a preference for the virtual content. It should be noted that from a total of 22 people, only a few read the information provided. This is understandable, considering the fact that the second evaluation was undertaken in a controlled and noise-free environment, conversely to the first, which took place in an exhibition space with constant presence of people and buzz.

7. Conclusions

The majority of the responses showed that the application of MR technologies for the improvement of users' experience and the enhancement of books' interpretive nature is a challenging pursuit. However, technological limitations, as well as non-affordable AR devices and enabling technologies have not taken this paradigm into mainstream

adoption. According to Gartner's hype-cycle Special Reports (Gartner 2010, 2011), where the maturity, business benefit and future direction of technologies is assessed, in 2010 AR was at the highest level of expectations, while in 2011, the underlying issues had made apparent that the wider adoption of this technology will delay for about five to ten years.

Within the archaeological computing communities there have been vivid discussions concerning the virtual imagery produced as part of our research and the modes and technologies we employ for public outreach. Our role as humanities researchers is to keep working on how to efficiently make these technologies accessible to the public, while building a critical framework according to which such technological advancements are exploited for the benefit of archaeological research and dissemination. People's responses seemed to encourage attempts of constructing aspects of the past and making them available to the public through novel computer methodologies. Our user testing and the feedback provided, showed that the continuous evaluation of such prototypes from a human and user-centred perspective, is an essential practice that can potentially prevent the risk of producing an application that is of little value to the users and the wider academic and cultural heritage community. Besides, we hope to have demonstrated that early stage user evaluations cannot only be used to assess technological components of CH applications but to significantly inform archaeological interpretive design *via* such user-centric approaches.

Acknowledgements

We would like to express our gratitude to Yannis Sakellarakis and Efi Sapouna-Sakellarakis for their support and for providing access to Zominthos material. Papadopoulos would also like to thank Psychia Foundation for the financial support. Finally, we thank the Archaeological Society at Athens, as well as the KE' and KΔ' Ephorates of Prehistoric and Classical Antiquities.

References

Azuma, R., Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre. 2001. "Recent Advances in Augmented Reality." *IEEE Computer Graphics and Applications* 21 (6): 34-47.

- Billinghurst, M., H. Kato, I. Poupyrev. 2001. "The Magicbook - Moving Seamlessly between Reality and Virtuality." *IEEE Computer Graphics and Applications* 21 (3): 6-8.
- Buxton, W., and R. Sniderman 1980. "Iteration in the Design of the Human-Computer Interface." In *Proceedings of the 13th Annual Meeting of the Human Factors Association of Canada*, 72-81. Canada: Human Factors Association of Canada.
- Champion, E. 2011. "Augmenting the Present with the Past." In *Playing with the past*, 157-176. London: Springer-Verlag.
- Drascic, D., and P. Milgram. 1996. "Perceptual Issues in Augmented Reality." In *Proceedings SPIE 2 653: Stereoscopic Displays and Virtual Reality Systems III*, edited by Mark T. Bolas, Scott S. Fisher, John O. Merritt, 123-34. California: SPIE-The International Society for Optical Engineering.
- Economou, M., and L. Pujol-Tost. 2007. "Evaluating the Impact of New Technologies on Cultural Heritage Visitors." In *Technology Strategy, Management and Socio-economic Impact, Heritage Management series, Vol. 2*, edited by J. McLoughlin, J. Kaminski, B. Sodagar, 109-117. Budapest: Archaeolingua.
- Frankland, T. 2012. "A CG Artist's Impression: Depicting Virtual Reconstructions using Non-photorealistic Rendering Techniques." In *Thinking Beyond the Tool: Archaeological Computing and the Interpretive Process*, edited by A. Chrysanthi, P. Murrieta Flores, C. Papadopoulos, 24-39. Oxford: Archaeopress.
- Gartner Press Release. 2010. "Gartner's 2010 Hype Cycle Special Report Evaluates Maturity of 1,800 technologies." Accessed June 2012. <http://www.gartner.com/it/page.jsp?id=1447613>.
- Gartner Press Release. 2011. "Gartner's 2011 Hype Cycle Special Report Evaluates Maturity of 1,900 technologies." Accessed June 2012. <http://www.gartner.com/it/page.jsp?id=1763814>.
- Karoulis, A., S. Sylaiou, and M. White. 2006. "Usability Evaluation of a Virtual Museum Interface." *Informatica* 17 (3): 363-380.
- Marco, de Sá, J. Antin, D. Shamma, E. F. Churchill. 2011. "Mobile Augmented Reality: Video Prototyping." In *CHIEA '11 Proceedings of the conference on Human factors in computing systems*, 1897-902. New York: ACM.
- Milgram, P., and F. Kishino. 1994. "A Taxonomy of Mixed Reality Visual Displays." In *IEICE Transactions on Information Systems* E77-D (12), 1321-1329.
- Müller, P., P. Wonka, S. Haegler, A. Ulmer, and L. Van Gool. 2006. "Procedural Modeling of Buildings." In *Proceedings of SIGGRAPH 2006 ACM Transactions on Graphics* 25 (3): 614-623.
- Noh, Z., M. Shahrizal, and Z. Pan. 2009. "A Review on Augmented Reality for Virtual Heritage System." In *Edutainment '09 Proceedings of the 4th International Conference on E-Learning and Games: Learning by Playing. Game-based Education System Design and Development*, edited by M. Chang, R. Kuo, Kinshuk, G.-D. Chen, M. Hirose, 50-61, Berlin: Springer.
- Palyvou, C. 1999. *Akrotiri Thiras. I Oikodomiki Techni*. Athens: Library of the Archaeological Society at Athens.
- Papadopoulos, C., and Y. Sakellarakis. 2013. "Virtual Windows to the Past: Reconstructing the 'Ceramics Workshop' at Zominthos, Crete." In *CAA 2010. Proceedings of Computer Applications and Quantitative Methods in Archaeology*, edited by Javier Melero, Pedro Cano, Jorge Revelles. Oxford: Archaeopress.
- Papadopoulos, C., E. Kefalaki. 2010. "At the Computer's Edge The Value of Virtual Constructions to the Interpretion of Cultural Heritage." *Archeomatica* 4: 46-51.
- Preziosi Donald. 1983. *Minoan Architectural Design*. Berlin: Mouton.
- Renevier, P., and L. Nigay. 2001. "Mobile Collaborative Augmented Reality: the Augmented Stroll." In *EHCI '01 Proceedings of the 8th IFIP International Conference on Engineering for Human-Computer Interaction*, 299-316, London, UK: Springer-Verlag.
- Rogers, Y., H. Sharp, and J. Preece. 2011. *Interaction Design: Beyond Human-Computer Interaction*. 3rd ed. UK: John Wiley & Sons.
- Shaw, J.W. 2009. *Minoan Architecture: Materials and Techniques*. Padova: Bottega d' Erasmo.
- Wright, P., and J. McCarthy. 2010. "Experience-Centered Design: Designers, Users and Communities in Dialogue." In *Synthesis Lectures on Human-Centered Informatics*, 3 (1): 1-123.

The Virtual Museum: a Quest for the Standard Definition

Laia Pujol

The Acropolis Museum, Greece

Anna Lorente

Telecom Bretagne, France

Abstract:

This paper aims to set the framework for the establishment of a standard definition of “virtual museum”. To that end, it will firstly examine its philosophical roots. Then, it will propose a framework of classification and it will review previous taxonomies and definitions. Finally, it will suggest a new definition of the term which is based on the philosophical and etymological meaning of virtuality, and which complies with the learning principles of the New Museology. This specific definition, together with the top down approach implicitly proposed by this paper, is expected to help unfold the potential of Information and Communication Technologies (ICT) for Cultural Heritage (CH) museums.

Keywords:

Virtual Museum, Classification, Definition

1. Introduction

According to V-Must, a Network of Excellence² funded by the 7th EU Framework Programme and purposefully devoted to the development of a specific research field about virtual museums (VM), these

“are personalized, immersive, interactive experiences that enhance our understanding of the world around us.”

The broadness of this definition, along with the diversity of the included typology³, gives an idea of the disparity of concepts gathered under this label. The VM arose spontaneously from a mixing of traditional museum practices, semiotic modes, market tensions and diverse technological possibilities (Lorente and Kanellos 2010). On the other hand, in most cases, the development was led by too technology-oriented perspectives (Kwon et al. 2003). As a consequence, it has become a complex notion that escapes from a shared definition. The same happens with the concept of Virtual

Reality (VR), which stands at the confluence of the historical development of different fields. Although several definitions and theoretical frameworks have been proposed in the CH domain (Barceló 2002; Champion 2005; Forte 2003), a standard definition has never been explicitly agreed. Instead, an assimilation with Virtual Heritage (VH) was spontaneously accepted and its use as a scientific simulation has been delayed several years⁴.

Polysemy of concepts is a well accepted characteristic of the Social Sciences and especially of the Humanities, which favours the advancement of knowledge in these fields. However, in the case of applied research it may be counterproductive: if we do not know what a VM is, it is difficult to determine how successful individual projects are, and perhaps impossible to create evaluative methodologies across different projects that support funding policies. The goal of this paper is therefore to set the foundations for the establishment of a standard definition of VM. On the one hand, the top-down approach requires a complete understanding of the field’s evolution, emphasizes the adoption of a conscious structure or planning, and proposes a specific definition based on explicit grounds or goals. With regard to the later, a definition based on the etymological

Corresponding author: pujol.laia@gmail.com

² <http://www.v-must.net/>

³ http://www.v-must.net/media_pubs/v-mustnet-deliverable-21

⁴ CAA 2012 is the first time a specific session is purposefully devoted to the simulative use of VR.

meaning of “virtuality” and which takes into account the common points between Information and Communication Technologies (ICT) and the current museological practices may help unfold the real potential of ICT for museums while ensuring they comply with the specific requisites of this context.

2. The Roots of the Virtual Museum

Due to a series of social and economic changes during the 20th century (Hooper-Greenhill 1994), the museum rethought its traditional role as a discipline-oriented, hierarchical, elitist institution aimed at the preservation of objects. The evolution in the definition of museum⁵ from 1947 to 1974 reflects this transformation since the text moved the emphasis from the institution and its objects to society and communication (Hudson 1998). However, the museum still functioned as a mass media (Schweibenz 1998). The progressive incorporation of ideas coming from the formal learning environment was gathered in the definition of 2007, in which education now preceded the study of objects as purpose of the museum. In practice, the so called New Museology⁶ applies the basic principles of Constructivism, which among others state that the adaptation of contents to each person, the adoption of an active role by the learner, and the exchange with other people are paramount for learning (Csikszentmihályi and Hermanson 1995; Falk and Dierking 2000).

Therefore, current museological practices and ICT share a common ground, since the main features of ICT -multimedia, interactivity, personalization- were already present in the development of the museum, although under slightly different terms and meanings (Pujol 2011). In this sense, it is important to be aware that although ICT allowed the accomplishment of utopias such as André Malraux’s “imaginary museum” (Kanellos 2009), they are just one particular way (based on computers) to continue developing the New Museology.

From an etymological and philosophical point of view, the meaning of the term “virtual” is related with the metaphysical debate and has slightly

changed over time. In the 19th and 20th centuries it was associated with specific non-visible elements or phenomena in Mechanics and Physics, respectively (Woolley 1994). In the 18th century the word was used in Optics, in relation to reflected images, as synonym of visible but non-material (Woolley 1994). These meanings are the result of the scientific revolution of the 17th and 18th centuries, which defined reality as material entity that is understood through experimentation. This is opposite to the previous medieval scholastic perspectives, which used the concept as synonym of potential, that is, the inherent qualities given by God to all entities, which allowed its intellectual understanding despite the existence of diverse empirical manifestations (Heim 1993). This definition was based on the original etymology of the concept, derived from the Latin term “*uirtus*”, a masculine quality that meant strength or power (Lévy 1995).

From this point of view, “virtual” is not opposed to “real” or “material”, but to “actual” (Lévy 1995) and helps to better understand the nature and capacity of several communication technologies or systems invented by humans. For instance, computers constitute the ultimate example of virtualization because they have become a universal machine, capable of potentially performing functions that previously corresponded to other devices (Lévy 1995). The same general notion can be applied to exhibitions, since the experience of the visit has been defined as a permanent dialog between visitors, objects and mediators (Falk and Dierking 2000). Hence, the brick and mortar museum already possessed the virtuality that is also present in the concept and the technological basis of the VM. In this sense, some authors (Deloche 2001) consider there is no opposition or possibility of replacement between the physical museum and its possible technological versions, because they all are aspects of the same entity: the “virtual museum” is a concept previous to technology, which designates the global museum field.

The opposite perspective (Welger-Barboza 2001) in the current museological debate about the VM considers that virtuality is a consequence of ICT applications, which aim at disseminating knowledge as wide as possible by means of digitalization. Therefore, the VM has no philosophical entity but arises from a change in the support. This position

5 http://archives.icom.museum/hist_def_eng.html

6 The term appeared officially for the first time in 1972, during the ICOM meeting in Santiago de Chile.

is so far the most common amongst museum professionals. Yet, it is also the most limited because it does not allow the full deployment of computers' potential. The progressive discovery of the communicational possibilities of ICT and of their similarities with the museum's modern postulates seems to be moving the field towards the first perspective.

3. Classification of VMs

In order to integrate all kinds of projects explicitly catalogued as VM by their authors, we propose their classification according to two orthogonal axes, which will be separately explained in sections 3 and 4.

The first axe (Fig. 1) follows Milgram and Kishino's (Milgram and Kishino 1994) definition of VR, and places the museum within the continuum between the real and the virtual worlds.

On the left side of the horizontal axe we find PC-based applications located in the physical world and aimed at enhancing the exploration of Cultural Heritage settings⁷. This specific museographical arrangement derives from the notion of VR as cognitive ecosystem, and has been meaningfully baptised by its authors (Forte, Pietroni, and Rufa 2002) as "the musealization of the virtual" to convey the idea that what makes the museum virtual is the introduction of this specific simulative technology. On the right side of the continuum we find museums that exist in the virtual world, either as totally immersive experiences⁸ or in the WWW.

The available classifications of museum online resources (Fig. 2) have evolved in parallel to their development. As a consequence, while in some cases there is a clear identification between types, in others they are included or they intersect, depending on the taxonomical criteria.

The first cataloguing was done in 1996 (Piacente 1996) and broadly disseminated thanks

7 For example, the Virtual Museum of the Ancient Via Flaminia (http://www.vhlab.itabc.cnr.it/flaminia/index_02.html).

8 For example, the Virtual Archaeological Museum of Herculaneum (http://www.museomav.it/index.php?option=com_content&view=article&id=93&Itemid=186&lang=en).

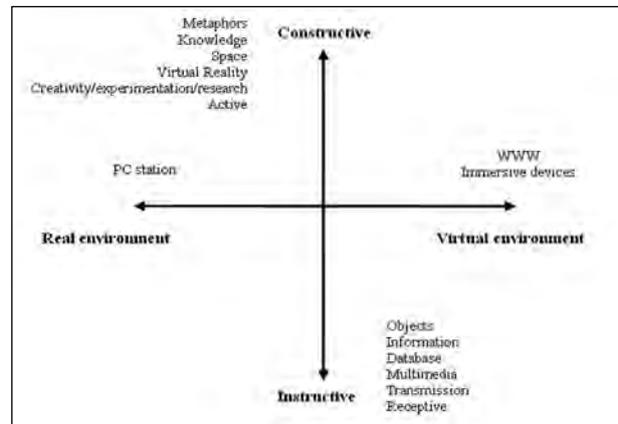


Figure 1. Classification of current VMs.

to its publication in "Museums & the Web" (Teather and Wilhelm 1999). The authors differentiate four different categories: electronic leaflets (museum information), reconstruction of the centre (copy of the physical exhibitions), true interactives (educational personalized resources) and big online databases. This classification was based on the goal or content of the website. A new criterion was introduced by Vidal in 2006 (Vidal 2006), who distinguished "showcases of real museums" (information and collections), "websites of online resources" (museums portal), and "virtual museums" (exclusively digital Art works). In this case, the emphasis was put in the relationship with the physical world.

Seemingly, Lorente and Kanellos (Lorente and Kanellos 2010), discriminate between "real museum's websites" (information and collections), "meta-museums" (museums portal) and "conceptual museums" (exclusively digital Art works). But they add "thematic museums" (works grouped under a topic) and use a different terminology that reflect their emphasis on the museological aim. Finally, Lauzikas (Lauzikas forthcoming) proposes a more detailed classification because he also takes technology into account. Thus, he distinguishes between "museum's website" (information about the museum), "virtual exhibition" (digitized museum collection), "virtualized real museum space" (panoramic photos or 3D scanning of real exhibitions), "online museums" (exclusively digital exhibitions), "VR projects" (simulation of the real or an imaginary world containing objects), and finally "virtual collection of references" (thematically structured collection of links to other web sites).

Piacente 1996 / Teather & Wilhelm 1999	Vidal 2006	Lorente & Kanellos 2010	Lauzikas 2010
Electronic leaflets	Showcase of real museums	Real museum's websites	Museum's web page
Reconstruction of the centre			Virtual exhibition
True interactives		Thematic museums	Virtualized real museum's space
			Online museums
	Virtual Museum	Conceptual museum	VR projects/Virtual Museum
	Website of online resources	Meta-museum	Virtual collection of references
Big online database			

Figure 2. Comparison between different classifications of online museum resources.

4. Interpretation

The second axe in Figure 1 reflects the communicational (or museological) paradigm of the VM, more constructive or more instructive depending on the general goal, the content, the interaction style and the role assigned to the user.

In the classifications presented in the previous section, the VM is associated with two different concepts. For Vidal, and Lorente and Kanellos, the “virtual” or “conceptual museum” contains works that have been created exclusively by computational means. This definition takes “virtual” as synonym of “non-material” or “digital”. This is why it is also called “digital museum”. Cary Karp (Karp 2005) provided an explicit definition for it, which is based on the latest definition of museum and the decision to include digital heritage as part of human intangible heritage (UNESCO 2003). Accordingly, the VM designates non-profit and non physical institutions, permanently accessible to the public, in the service of society and its development, that acquire conserve, research, communicate and exhibit exclusively digital-born heritage for the purposes of education, study and enjoyment.

The other possibility is represented by Lauzikas, who associates the VM with VR by means of the common term (“virtual”). This puts the emphasis on the spatial, simulative dimension and matches the oldest definition of VM (Tsichritzis and Gibbs 1991), which considered it was a networked 3D model of a real or invented museum setting containing “active” contextualized objects. This is not an uncommon perspective, and is usually supported by researchers working in the field of VR.

The only problem is that, as previously mentioned, in some cases it is difficult to distinguish the VM from VH applications⁹. According to different studies, the advantage of VR-based VMs over pure multimedia is that exploration is more intuitive because it uses familiar metaphors (Economou and Pujol 2007; Gordon 1999). Moreover, there are also early examples of VM (Loeffler 1994; Shaw 1994) that are based on Multi-user Virtual Environments (MUVE) and Object-Oriented Multi-User Dungeons (MOOs) in which users can communicate with each other from distance and even modify the virtual space (Gordon 1999). This fully interactive capacity places them closer to the “Constructive” side of the vertical axis in Figure 1, and also complies with the first philosophical stance presented in section 2.

Forte and Franzoni (Forte and Franzoni 1998, 202) proposed a definition that matches¹⁰: “By virtual museum we understand a computational environment consisting of a hypertext and hypermedia structure and an interface system containing metaphors that use a more or less intuitive graphic representation. This allows the visitor to perform actions and therefore to navigate within the context, to interact with it or even to modify it.”

Yet, in spite of experts’ classifications and nomenclatures, current uses show there is a third possibility. This understands “virtual” as synonym of “digitalized” (the philosophical stance presented in section 2), and due to the overwhelming universal spreading of the WWW, has associated the VM and

9 See, for example (Sylaiou et al. 2009; Sylaiou et al. 2005).

10 The translation is ours, from the original written in Italian.

online resources (Ivarsson 2009). With regard to digitalization, VMs correspond to multimedia hypertexts built upon databases of digitalized items, hence emphasizing the objectual dimension of the museum and keeping the VM subsidiary to the physical institution. These VM are located on the “Instructive” side of the vertical axis in Figure 1. The concept also strengthens the informational dimension of the medium. This taken to the extreme blurs the very same definition of museum as an institution preserving and disseminating material culture since it is also possible to find VMs where objects have been substituted by a theme¹¹, a heritage place (Bonfigli and Guidazzoli 2000; Bonfigli, Guidazzoli, and Liguori 2002) or a whole country¹² (somehow corresponding to Lorente and Kanellos’ “thematic museums” or Lauzikas’ “online museums”).

The notion of VM as a digitalization of real institutions can be found in early definitions, such as (Schweibenz 1998):

“The “virtual museum” is a logically related collection of digital objects composed in a variety of media and, because of its capacity to provide connectedness and various points of access, it lends itself to transcending traditional methods of communicating and interacting with the visitors being flexible towards their needs and interests; it has no real place or space, its objects and the related information can be disseminated all over the world.”

Seemingly, (Papageorgiou 2000) says:

“Virtual museums provide simulations of real museums in computers. They are virtual environments that represent a plan of the museums along with its exhibits and are presented over the internet to a wide range of users. The purpose they satisfy is to provide access to museums from long distances all day long. Thus visitors are able to take a tour in the museum without having to travel and without being restricted by the museum’s opening hours. What is more a virtual replicate of the

¹¹ For example, the Tax History Museum (<http://www.tax.org/Museum/default.htm>), the Virtual Shoes Museum (<http://www.virtualshoemuseum.com/vsm/>), or the Virtual Museum of Daily life (<http://muvi.cineca.it/>).

¹² For example, The Virtual Museum of Iraq (<http://www.virtualmuseumiraq.cnr.it/homeITA.htm>).

museum could provide more information to the visitor allowing him to examine the artefacts and receive even multimedia information about it.”

These definitions seem very close to Jamie McKenzie’s (McKenzie 1997), which has become a reference:

“A virtual museum is a collection of electronic artefacts and information resources-virtually anything which can be digitized. The collection may include paintings, drawings, photographs, diagrams, graphs, recordings, video segments, newspaper articles, transcripts of interviews, numerical databases, and a host of other items which may be saved on the virtual museum’s file server. It may also offer pointers to great resources around the world relevant to the museum’s main focus.”

Taken alone (as done by most researchers), this definition has two outcomes. On the one hand, it presents no evident distinction from a database. The association with the museum is done by means of the presence of a “collection” of artefacts. On the other hand, the fact that any digital content can be part of a VM is at the origin of a second assimilation, this time between VH and the VM. In this sense, the V-Must network’s website says:

“The term ‘VM’ is a general one that covers various types of digital creations including virtual reality and 3D”.

This justifies the gathering of so diverse projects and is very close to the most accepted definition of VH, which understands the concept as the digital recording, preservation or recreation of CH elements (Stone and Ojika 2000). However, McKenzie’s definition has to be understood in the context of his whole proposal, in which “learning VMs” were the key element of educational partnerships between local museums and schools (McKenzie 1997). Thus, the collection of digitalized items, which was selected and arranged by students, was not an end but a means for constructivist learning and community building¹³.

Coming back to the recent association between the VM and the WWW, it has been gathered by the

¹³ A similar approach can be found in (Worden 1997; Jones and Christal 2002).

latest definitions, such as the Wikipedia's, which also tries to keep the polysemy of the term by providing an all-encompassing definition:

“A virtual museum is a museum that exists only online. A virtual museum is also known as an online museum, electronic museum, hypermuseum, digital museum, cybermuseum or Web museum. The term used depends upon the backgrounds of the practitioners and researchers working in this field. As with a traditional museum, a virtual museum can be designed around specific objects (akin to an art museum, natural history museum, or archive), or can consist of new exhibitions created from scratch (akin to the exhibitions at science museums). Moreover, a virtual museum can refer to the mobile or World Wide Web offerings of traditional museums (e.g., displaying digital representations of its collections or exhibits); or can be born digital content.”

However, this does not contribute to solve the ambiguity of the term, and therefore is of little epistemological and methodological use in order to establish an operational definition of VM upon which to build a specific integrated framework for design and evaluation.

5. Definition

With that aim in mind, we would like to propose a definition for VM. “Digital museum”, the first possibility deduced from the analysis of online resources, is becoming a widely accepted term specifically for websites containing digitally created works. The third possibility, “virtuality” understood as the simple “dematerialization” of the museum (the second philosophical position) prevents it from exploiting other advantages of ICT apart from broad dissemination of the physical institution through the WWW. Conversely, the second possibility takes “virtuality” as synonym of performance (the first philosophical position), which not only is the main feature of computers but also complies with the learning principles (interaction) of New Museology.

Hence, taking into account which was the oldest notion of VM (Tsichritzis and Gibbs 1991), we would propose to define it as a virtual environment reproducing a space, which constitutes a reconstruction of a real place and/or acts as a knowledge metaphor, where visitors

can communicate between them, and explore, create or modify spaces and (digital or digitalized) objects (Worden 1997). Consequently, its location in the WWW is not mandatory, as supported by (Jones and Christal 2002), but just a matter of the aimed dissemination. Another consequence has to do with the role of the digital object. While in most VH examples, which are based in the second philosophical stance, it constitutes a simple surrogate of the physical object, that is, an “analogical substitute” (Deloche, 2001), in the proposed definition of the VM, which is based in the first philosophical stance, the object constitutes an “analytical substitute” (Deloche, 2001), that is, a dynamic and/or interactive image that can be manipulated and therefore accomplishes a scientific or learning function.

Which is the advantage of this concept of VM with regard to the real museum? The flexibility provided by ICT computational virtuality, which allows reconstructing different spaces and CH places; using different interfaces for different locations (Sylaiou et al. 2009); overcoming the limitations of the physical exhibition with regard to the number of objects and the approach to them (Bernier 2002; Sylaiou et al. 2005); offering an experimental platform for the design of exhibitions (Sylaiou et al. 2009); personalizing contents and formats (Bernier 2002; Bowen and Filippini-Fantoni 2004); establishing a real dialogue between museums and audiences; attracting new audience sectors (McDonald 2005); building community through the contribution to exhibitions; providing a greater sense of presence (Winn et al. 1999); and finally, as in the case of Virtual Learning Environments (VLE), using a constructivist approach, which benefits learning (Byrne 1996; Winn 1997), especially in the case of some specific subjects (Scanlon et al. 1998).

6. Conclusions

Believing in the importance of well established standard concepts for the development of (applied) research, this article aimed to set the framework for a standard definition of VM. It started by examining the theoretical framework that links the museum with ICT applications: because of an overlap of concepts and aims, ICT seem to accomplish some of the ambitions of the so-called New Museology. Next, the paper examined current examples and

uses, and suggested a comprehensive classification based on two axes: location (real vs. virtual world) and communicational paradigm (instructive vs. constructive).

It then reviewed previous classifications and definitions of the term. This led to the conclusion that the definition and communicational features of the VM depend on the meaning attributed to the term "virtual". The most usual is to understand it as a synonym of digitalization of real elements and provide a multimedia database of objects, where mainly unidirectional communication and strictly controlled learning processes take place. This definition places the VM as a subsidiary element with regard to the institution and evidences that the definition and implementation of the VM revolves around its conceptual distance to the physical visit. As a consequence, the real potential of ICT for museum communication and learning online cannot be completely unfolded: if we extrapolate the results from the well studied VLE, databases seem to have less "affordances" for learning, especially in the case of museums, where MOOs and MUVes reproduce the real exhibition and its capacities with a greater degree of analogy than a hypertext.

Consequently, we propose to use the term VM to refer to a digital spatial environment, located in the WWW or in the exhibition, which reconstructs a real place and/or acts as a knowledge metaphor, and in which visitors can communicate, explore and modify spaces and digital or digitalized objects. We believe this is the value of the VM, as a new museum resource that allows activities and dialogues in ways impossible both physically and institutionally for our good old brick and mortar museum. In this sense, the VM is deeply rooted in it, but it is something different, at the crossroads of distance learning, constructivism and technology, and should develop a totally new theoretical framework, based on the integration of a user-centred design with the three aforementioned fields. While it was not the scope of this paper, we hope that the present review will constitute a foundation upon which the new "bit and pixel" museum is solidly built.

Acknowledgements

The work presented here was funded by the Catalan Government's Department for Universities,

Research and the Information Society (FI, III Pla de Recerca de Catalunya 2001/2004), and by the French Institut Telecom (PhD Fellowship "Future et Ruptures").

References

Barceló, J. A. 2002. "Virtual Archaeology and Artificial Intelligence." In *Virtual Archaeology. Proceedings of the VAST Euroconference (Arezzo 24-25 November 2000)*, edited by F. Niccolucci, 21-28. Oxford: Archaeopress.

Bernier, R. 2002. The uses of virtual museums: the French viewpoint. In *Museums & the Web*, edited by David Bearman and Jennifer Trant. Pittsburgh: Archives & Museum Informatics.

Bonfigli, M.E., and A. Guidazzoli. 2000. "A www virtual museum for improving the knowledge of the history of a city." In *Virtual reality in Archaeology*, edited by J. A. Barceló, M. Forte and D. H. Sanders, 143-148. Oxford: Archaeopress.

Bonfigli, M.E., A. Guidazzoli, and M. Ch. Liguori. 2002. "From Museum to Mouseum. The Virtualisation of Daily Museums." In *Virtual Archaeology. Proceedings of the VAST Euroconference (Arezzo 24-25 November 2000)*, edited by F. Niccolucci, 213-215. Oxford: Archaeopress.

Bowen, J. P., and S. Filippini-Fantoni. 2004. "Personalization and the web from a museum perspective." In *Museums & the Web*, edited by D. Bearman and J. Trant, 63-78. Arlington, Virginia: Archives and Museums Informatics.

Byrne, Ch. M. 1996. "Water on tap: the use of Virtual Reality as an educational tool." PhD diss, College of Education, University of Washington.

Champion, E. 2005. "Interactive Emergent History as a Cultural Turing Test." In *VAST 2005: 6th International Symposium on Virtual reality, Archaeology and Intelligent Cultural Heritage. Short & Project Papers Proceedings*, edited by M. Mudge, N. Ryan and R. Scopigno, 120-125. Pisa: Eurographics Association.

Csikszentmihályi, M, and K. Hermanson. 1995. "Intrinsic Motivation in Museums. What Makes Visitors Want to Learn." *Museum News* 74 (3):35-62.

Deloche, B. 2001. *Le musée virtuel, Questions actuelles*.

The Virtual Museum: a Quest for the Standard Definition
Laia Pujol and Anna Lorente

Paris: Presses Universitaires de France.

Economou, M., and L. Pujol. 2007. "Educational tool or expensive toy? Evaluating VR evaluation and its relevance for Virtual Heritage." In *New Heritage. New media and cultural heritage*, edited by Y. Kalay, Th. Kvan and J. Affleck, chapter 19. London: Routledge.

Falk, J. H., and L. D. Dierking. 2000. *Learning from Museums. Visitor Experiences and the Making of Meaning*. Oxford: AltaMira Press.

Forte, M. 2003. "Mindscape: ecological thinking, cyber-anthropology and virtual archaeological landscapes." In *The reconstruction of archaeological landscapes through digital technologies. Proceedings of the First Italy - USA Workshop (Boston, Nov 1-3 2001)*, edited by M. Forte and A. C. Addison, 95-108. Oxford: ArchaeoPress.

Forte, M., and M. Franzoni. 1998. "Il museo virtuale: comunicazione e metafore." *Sistemi Intelligenti X* (2):193-240.

Forte, M., E. Pietroni, and C. Rufa. 2002. "Musealizzare il virtuale: il progetto di allestimento multi-interattivo per la Cappella degli Scrovegni." Accessed February 2013. <http://www.mlib.cnr.it/vhn/SitoScrovegni/MusealizzVirtuale.htm>.

Gordon, S. 1999. "The virtual museum - who needs it?" In *Archaeology in the Age of Internet. Computer Applications in Archaeology. 25th Anniversary Conference*, edited by L. Dingwall. Oxford: Archaeopress.

Heim, M. 1993. *The metaphysics of virtual reality*. New York: Oxford University Press.

Hooper-Greenhill, E. 1994. *Museums and Their Visitors*. London: Routledge.

Hudson, K. 1998. "The museum refuses to stand still." *Museum International* 197:43-50.

Ivarsson, E. 2009. "Definition and prospects of the Virtual Museum." Master Thesis, University of Uppsala.

Jones, J., and M. Christal. 2002. "The future of virtual museums: on-line, immersive, 3D environments." Accessed February 2013. http://created-realities.com/pdf/Virtual_Museums.pdf.

Kanellos, I. 2009. "Les musées virtuels et la question de la lecture: pour une muséologie numérique centrée sur le visiteur." *Revue des Interactions Humaines Médiatisées* 10 (2): 3-33.

Karp, C. 2005. "Digital Heritage in Digital Museums." *Museums International* 56 (1-2):45-51.

Kwon, Y. M., J.E. Hwang, T.S. Lee, M. J. Lee, J. K. Suhl, and S. W. Ryul. 2003. Toward the synchronized experiences between real and virtual museum. In *Papers at the Asia-Pacific Advanced Network Conference in Fukukoa, Japan*. Accessed February 2013. <http://www.qgpop.net/2003fukuoka/papers/B5-1.pdf>.

Lauzikas, R. in press. "Virtual museum between web page and Second Life." In *Fusion of Cultures. Proceedings of the XXXVIII Annual Conference on Computer Applications and Quantitative Methods in Archaeology*, edited by F. Contreras and F. J. Melero. Granada.

Lévy, P. 1995. *Qu'est ce que le virtuel?* Paris: La Découverte.

Loeffler, C. E. 1994. "The Networked Virtual Art Museum and other projects." In *The Virtual Reality Casebook*, edited by C. E. Loeffler and T. Anderson, 31-37. New York: Van Nostrand Reinhold.

Lorente, A., and I. Kanellos. 2010. "What do we know about virtual museums? A study about the current situation of virtual art museums." In *Proceedings of the Conference Transforming Culture in the Digital Age, April 14-16, 2010, Tartu, Estonia*, edited by A. Aljas, R. Kelomees, M. Laak, P. Pruulmann-Vengerfeldt, T. Randviir, P. Runnel, M. Savan, J. Tomberg, P. Viires, 208-21 Tartu: Pacific Arts Association.

McDonald, M. 2005. "The museum and the Web: comparing the virtual and the physical visits." Master Thesis, University of Virginia.

McKenzie, J. 1997. "Building a Virtual Museum Community." In *Museums & The Web Conference*, edited by D. Bearman and J. Trant. Los Angeles: Archives and Museums Informatics.

Milgram, P., and F. Kishino. 1994. "A Taxonomy of Mixed Reality Visual Displays." *IEICE Transactions on Information Systems* E77-D (12):1321-1329.

Papageorgiou, D. 2000. "Case Studies of Cultural Heritage

- in VR." In *Prospects of Augmented Reality in European Cultural Heritage Sites*, edited by M. Roussou, 14-28. Peania: Intracom, S.A.
- Piacente, M. 1996. *Surfs up: Museums and the World Wide Web*. Toronto: University of Toronto.
- Pujol, L. 2008. "Does virtual archaeology exist?." In *Layers of perception. Advanced technological means to illuminate our past. Proceedings of the Conference on Computer Applications and Quantitative Methods in Archaeology*, edited by A. Posluschny, K. Lambert and I. Herzog, 101-107. Bonn, Rudolf Habelt GmbH.
- Pujol, L. 2010. "Virtual Reality as a Learning Tool for Archaeological Museums". In *Beyond the artifact. Computer applications and Quantitative Methods in Archaeology, CAA 2004*, edited by F. Niccolucci and S. Hermon. Oxford: ArchaeoPress.
- Pujol, L. 2011. "Integrating ICT in museums." *Journal of Museum Management and Curatorship* 26(1): 63-79.
- Scanlon, E., C. Tosunoglu, A. Jones, P. Butcher, S. Ross, J. Greenberg, J. Taylor, and P. Murphy. 1998. "Learning with computers: experiences of evaluation." *Computers & Education* 30 (1/2):9-14.
- Schweibenz, W. 1998. "The "Virtual Museum": New Perspectives for Museums to present Objects and Information Using the Internet as Knowledge Base and Communication System." In *Knowledge Management und Kommunikationssysteme, Workflow Management, Multimedia, Knowledge Transfer - Proceedings des 6 Internationalen Symposiums für Informationswissenschaft (ISI '98)*, edited by H. A. Zimmermann, 185-200. Trier: Universität Trier.
- Shaw, J. 1994. "The Virtual Museum." In *The Virtual Reality Casebook*, edited by C. E. Loeffler and T. Anderson, 27-30. New York: Van Nostrand Reinhold.
- Stone, R., and T. Ojika. 2000. "Virtual Heritage: What Next?" *IEEE Multimedia* 7 (2):73-74.
- Sylaiou, S., F. Liarokapis, L. Sechidis, P. Patias, and O. Georgoula. 2005. "Virtual museums, the first results of a survey on methods and tools." In *In International Cooperation to Save the World's Cultural Heritage. Proceedings of the CIPA International Symposium*, edited by Sergio Dequal, 1138-1143. Torino: CIPA.
- Sylaiou, S., F. Liarokapis, K. Kotsakis, and P. Patias. 2009. "Virtual museums, a survey and some issues for consideration." *Journal of Cultural Heritage* 10:520-528.
- Teather, L., and K. Wilhelm. 1999. "'Web Musing': Evaluating Museums on the Web from Learning Theory to Methodology." In *Proceedings of the International Conference Museums and the Web*. edited by D. Bearman and J. Trant. Pittsburgh: Archives and Museums Informatics.
- Tsichritzis, D., and S. Gibbs. 1991. "Virtual Museums and Virtual Realities." In *Hypermedia & Interactivity in Museums, Proceedings of an International Conference*, edited by D. Bearman, 17-25. Pittsburgh: Archives & Museum Informatics.
- UNESCO. 2003. *Charter on the Preservation of Digital Heritage*. Accessed February 2013. http://portal.unesco.org/en/ev.php-URL_ID=17721&URL_DO=DO_TOPIC&URL_SECTION=201.html.
- Vidal, G. 2006. *Contribution à l'étude de l'interactivité. Les usage du multimédia de musée*. Bordeaux: Presses Universitaires de Bordeaux.
- Welger-Barboza, C. 2001. *Le patrimoine à l'ère du document numérique. Du musée virtuel au musée médiathèque, Patrimoines et Sociétés*. Paris: L'Harmattan.
- Winn, W. 1997. *The impact of three dimensional immersive virtual environments on modern Pedagogy*. Washington: Human Interface Technology Laboratory, University of Washington.
- Winn, W., H. Hoffman, A. Hollander, K. M. Osberg, H. Rose, and P. Char. 1999. "Student-built virtual environments." *Presence* 8 (3):283-292.
- Woolley, B. 1994. *El universo virtual*. Madrid: Acento Editorial.
- Worden, S. 1997. "Thinking critically about Virtual Museums." In *Museums and the Web*, edited by D. Bearman and J. Trant. Pittsburgh: Archives and Museums Informatics.

Can you Hack (the) Communication?

Hugh Corley

English Heritage, United Kingdom

Abstract:

Existing practice presents a far greater challenge to the adoption of innovative technology than does the rain, the wind and the mud. Archaeologists do not like change. With cheaper more robust consumer technology we must turn our attention to the greater challenge - successful adoption. Communication is fundamental to overcoming this challenge. In April 2008 English Heritage purchased the hardware and software necessary to implement the Intrasis archaeological information management system. This was only the beginning. Many of our procedures were incompatible with Intrasis others were simply out of date. The adoption of Intrasis presented the opportunity to review these procedures. Implementing this change has led to a higher quality record, and we are seeing faster assessments and analyses. Ultimately, the greatest cost was not the hardware or software, but the time spent adapting to change. This paper will discuss the integration of Intrasis into our working practices at English Heritage. It will explore how the system was introduced to our specialist teams, the impact, the costs and the benefits that were eventually realised. The hope is that the lessons we have learned will persuade other archaeologists to persevere to overcome the technical, organisational and human pitfalls inherent in such change. So you've discovered the greatest archaeological tool since the trowel but can you hack the communication required to get archaeologists to use it?

Keywords:

Communication, Change Management, Excavation Recording

1. Introduction

Fully digital field-data recording has yet to be adopted by the vast majority of archaeologists. As has been the case for many years, archaeologists operate a hybrid recording environment where some of the information is born digital and the rest is captured onto paper (May 2004b). As cited elsewhere (Andrew 2000; Backhouse 2006), this separates excavation data collection from its interpretation. Digital systems supported by appropriate procedures can enhance interpretation in the field, improve data quality and accelerate projects to completion. However, the benefits of any system will be difficult to realise without good communication. This paper will explore these issues by looking at the implementation of the Intrasis Archaeological Spatial Database System (Intrasis 2012) within English Heritage.

In April 2008, English Heritage procured the Intrasis system from the Swedish Nation Heritage Board. In advance of this a survey of the different systems was conducted and Intrasis
Corresponding author: hugh.corley@english-heritage.org.uk

was identified as the appropriate system because it met the data collection and structure (Cripps 2005) needs identified in the Revelation Report. Revelation highlighted the presence of copious ad hoc and bespoke archaeological systems within the organisation. Systems developed to meet specific methodological needs of the designer, but with no understanding of the needs of others or how data would be integrated between systems (May 2004a). The very aim of addressing this issue by adopting a single digital recording system reflected an awareness of the need for longer term thinking about our data collection and management strategies (Backhouse 2006). Key to accomplishing this was management support, not just verbal support and the finances needed to procure and operate Intrasis, but also the time that would be essential to review procedures to create an environment for any new system to succeed.

2. Procedural Review

The adoption of Intrasis provided an opportunity for a review and update of our working procedures. This review considered the role of Intrasis, how different teams relate to one another,

ensured that we had up to date procedures that were understood by the whole team and aligned to MoRPHE (English Heritage 2006). To be clear, this was not the altering of pre-existing practice and procedures to meet the needs of technology warned about elsewhere (Warwick 2009), it was revising, repairing and, in some instances, developing stronger procedures 'to support the complex needs and work patterns of archaeological teams in the lab and the field'(May 2004a). Although this did add to the transition required, it laid the foundation for successful implementation and better communication.

Review of procedures was accomplished through several formal and informal meetings between all of the teams involved. Different parts of an archaeological project from inception through to archive completion were considered in great detail to illuminate areas where practice was inadequate or needed clarification. Break-out groups were formed and different procedures were developed that enabled clearer communication and identification of the processes the teams undertake. As a result different checklists and written instructions were developed that clarify who is responsible for what and when and what tasks are required to move from one stage to the next. These have proven to be invaluable tools for our teams as projects are developed and worked through to completion.

3. Intrasis in the Field

It was clear from the outset that while Intrasis would be used in the field it would not be used in the trench. Initial attempts to use Intrasis in the trench proved impractical. As has been determined elsewhere (Warwick 2009), delivering what most of us would consider 'standard' IT becomes 'extreme' computing in archaeological conditions, and even before you add the dirt and mud.

To understand how Intrasis would work in the field it is worth looking at what data is collected on site and how it is used. The Revelation work discovered three types of information:

- Information needed to excavate and record the site;
- Information needed for onsite interpretation;

- Information needed for 'post-excavation' work (May 2004a).

It was with these types of information in mind that 'aide memiore' were developed. The 'aide memiore' are paper forms, designed to convey their impermanent state. These would allow the archaeologists to have the information they needed where they needed it on paper while still ensuring that the information for onsite interpretation and post-excavation work was captured in Intrasis. Although the recording might start on the paper it was not the primary record, but rather intended as a vehicle for capturing information before it is entered into Intrasis. To be clear, the data entry process is not a menial task, the entering of data into Intrasis is part of the capture and interpretative process.

Intrasis is capable of taking data that is born digital (from Total Station, Differential GPS & digital cameras) and data that is entered directly by users and capturing how they relate. Data (for example, description of a context) is entered into Intrasis on a locally networked, on site laptop by the excavator. During data entry, they must ensure that spatial and other data (for example, images, sample and finds information) are correctly related. Once they are satisfied that they have correctly completed the record they mark it as ready to be checked. The supervisor then checks this digital record to confirm it has the correct information (both tabular and spatial) and relationships.

Making certain that the system works in the field requires we first capture the born-digital data in Intrasis and that everything is safely backed up. To accomplish this we developed new procedures for downloading survey and photography equipment into Intrasis. Since the introduction of this kit, it has commonly been working practice that information is downloaded from the equipment in the evening back at the accommodation. However, this is considered bad practice and we wanted to ensure that it was done during working hours. To accomplish this, all survey and photography equipment now comes off site one hour before the end of the working day to have the data loaded into Intrasis. This also allows the network and the laptops to be properly shut down and packed away before the end of the working day.

To properly resource projects that are able to effectively complete excavations and leave the field with complete and checked record in Intrasis, we needed to adopt a 'front-loading' approach (Andrew 2000), with additional staff and time for our fieldwork. This additional resource for fieldwork has allowed our projects to leave the field with all contextual, finds and environmental data entered and checked in Intrasis. Additional supervision is provided to ensure all of the data is downloaded, checked and backed up as soon after a context is excavated as possible. As a result we have seen our projects move much faster to archive completion and into analysis in accordance with MoRPHE (English Heritage 2006)

4. The Cost of Change

As the previous section started to indicate, the cost of hardware and software is only a small part of adopting and implementing a digital recording system. Proper consideration of the needs of the hardware, software and its users should all be carefully considered as early in the process as possible. Failure to properly determine these costs will not only jeopardise the adoption of any system, but more critically could jeopardise any collected data.

Turning our attention to costs, it is worth beginning with the cost of the infrastructure for this type of system. We implemented Intrasis by using laptops inside site porta-cabins. Initially, we had hoped to run two concurrent projects each with 5 laptops (and 2 further laptops held back in case of failure). However, it became apparent that this caused bottlenecks and all 10 machines, would be needed on each site. It also became apparent that our finds and environmental staff were able to enter data directly into Intrasis without an 'aide memoire'. This meant that space would be required for 8 laptop workstations, network hardware and an onsite printer. We therefore needed to rent an additional office porta-cabin for our excavations and run network wiring through protective conduits for the remaining two laptops in the finds and environmental porta-cabin(s). The cost of all of this gear along with set-up and breakdown time must be taken into account.

5. Training

Easily the largest and longest running cost to implementing a digital system is training. All our staff receive training in the use of Intrasis as do all temporary excavation staff for each project. Making certain that everyone is confident to carry out all of the tasks that they are required to undertake is critical to successful adoption (Fisher 2010).

A core team of Intrasis users were trained in the use of Intrasis over the course of a week by the Swedish National Heritage Board. This team of 'super-users' take responsibility for training both their colleagues and the temporary excavation staff. A super-user is assigned to each project, responsible for addressing any issues and customisation of Intrasis that may be required. They are onsite for the first week of the excavation providing Intrasis training and make several other trips to site over the course of the project to address any issues that arise, and of course they are available to answer questions over the phone. Additionally, all staff are encouraged to phone the Intrasis helpdesk in Sweden, a service offered with the purchase of the system.

The training provided by the super-users offers a good introduction, however, those who invested time and really used the system became the most confident and capable users and peer instructors. As this confidence grew they were able to be more demanding of the system and ask for particular customisations.

6. Capturing Communication

One of the key elements that we hoped to capture with Intrasis was feedback; about data quality, the database itself, user experience and data analysis. By having Intrasis capture all of the data up to site archive completion, in a single system we hold all of this information alongside interpretations. By developing Intrasis to suit our needs we were able to feedback information about issues with the data as well as capture suggested improvements and customisations to ensure we are recording the correct data in the correct way.

Recording requests for alterations to the database became a critical part of the implementation process. Initially, we captured our recording

system as it existed prior to implementing Intrasis and minimised changes to only those that were unavoidable. However, once the system was put into use and users became more familiar and confident with the system further changes were necessary.

To manage these changes a change request system (CRS) was developed using Microsoft Sharepoint (Microsoft 2010). This basic system allows the collection of change requests, taken from personal communication and feedback within the different Intrasis databases. Changes were not only required to the data structure of Intrasis, but also clarifications in training material, procedures, documentation and suggestions for improvements to Intrasis.

It is worth exploring the impact of this system a bit further as it helps to understand how critical capturing this information can be to successful implementation. Firstly, capturing feedback about a system from its users can be extremely beneficial to determining where further training may be required. Also users who feel their concerns are being addressed to will put up less resistance. More important than this though, is the opportunity it provides to highlight methodological conflicts between different areas of archaeological investigation. Prior to the CRS, project managers or specialists would make a request for a change, which might get challenged or undone by a different project manager or specialist on the next project. Recording the change and documenting why it was required encourages more collaborative working by making colleagues with different methodologies and disciplines work together to capture the correct data. Another significant advantage to capturing changes and explanations behind them is that they will provide invaluable metadata to anyone reusing our data and examining our data structure.

7. The Digital Reality

The focus of the discussion has been on the ideal or how it should work. However, the reality is that despite the incredible potential benefits of digital recording, we encountered and still encounter a great deal of resistance and implementation issues. Resistance is inevitable, even among those responsible for implementing a system. At some point when instigating change on the scale of digital

recording you will run up against problem as unique as the individuals with them. Identifying these problems and determining appropriate resolutions is fundamental to successful implementation.

By capturing some data with 'aide memiore' we are arguably double-handling, particularly in the eyes of the person who has to write it down and then later enter it into Intrasis. Helping the excavators understand the benefits of this additional work is essential and something that we did not always get right. The benefit of this additional recording to the excavator must be a greater understanding of the site and how it all relates together that having a structured digital system can provide. This requires the excavator to have additional skills and time to query the system and look at different relationships in the data both tabular and spatial.

Despite hiring highly skilled and experienced staff we still confronted issues in use of the system. The time required to check data was routinely underestimated and concern about the amount of time this was taking arose. Anecdotal reports of supervisors checking paper rather than the primary digital records, were received. Excavators were reluctant to give up survey and photography equipment so it could be downloaded into the system before the end of the day. All of these changes to practice can be difficult to manage and we continue to be very reliant on our project archaeologists to follow the necessary procedures for the system to continue to work.

Digital systems are not a panacea that will solve all of our problems. 'Front-loading' a project can be hazardous. In my experience archaeologists are frequently running out of time at the end of a project. Providing them with more staff and/or more time without proper oversight will lead to more holes and a large backlog of data to process. Good and clear procedures can help avoid this, and making it clear what is expected of this additional resource is vital.

For Intrasis and our new procedures to demonstrate maximum benefit we needed excavations that ran for several months, not several weeks. The real benefit of this approach will be on sites that run for longer with staff that are trained and are using the system for several months.

8. Conclusions

Direct digital capture of excavation data is not a simple task. It is costly, tiring, aggravating and fraught with difficulty. So why do we want to do it? Why don't we just stick with things the way they are? Because properly structured digital datasets containing high quality data may save us all. In these times of shrinking budgets and re-aligned priorities we must all prove our value.

Digital datasets that can be semantically mapped will enhance cross searching sites alongside grey literature reports (Andrew 2000; Tudhope 2011). The great potential of searching across datasets could increase the pressure for data standards (Richards 2008) and, in turn, continue to increase the quality of excavation data. We must not only demonstrate that we can capture digital data, but that the digital data we capture is of the quality that we as a discipline should expect.

As suggested by Ellis (2012), tablet computers and particularly the iPad have now enabled direct, trench side digital excavation recording. Implementation in the research environment of Pompeii clearly demonstrates the potential for broader use of this technology and certainly appeared to encounter much less resistance than we experienced. The potential for all site staff to collect and check all site data digitally at the trench edge may have arrived. Replication of this project's success with iPads, will only be successfully if the full costs are considered, resistance is managed and appropriate procedures are developed and followed. Otherwise trench side computers will only ever be expensive toys for excavators to break.

Having now worked with several Intrasis datasets over the past few years, it is clear that digital errors are much easier to find and identify, than those buried in paper systems. Therefore, when moving to digital systems it is more important than ever that errors are identified and corrected as early in the process as possible. Far too often a 'harmless' shortcut in the field becomes a costly delay in later stages of the process (Backhouse 2006), these must be identified and avoided. Every stage of the archaeological process must consider how it can help to ensure that we produce the highest quality datasets, from the excavator through supervisors,

specialists, managers, inspectors and curators, everyone must start to develop skills so they can engage with digital data.

If we are to truly see the benefits of 'front-loading' our projects and include excavators in the interpretative process, we must acknowledge the costs. If site supervisors continue to expect to excavate, site excavators should expect to have time to explore datasets. Doing this will require more excavator and supervisor time away from the trench, not just entering and checking data, but contextualising the work they are doing and interpreting not just their small area in isolation, but the whole site.

Many of the lessons learned from the implementation of Intrasis are more broadly applicable to different database systems. More archaeologists need to develop the digital skills to quickly adapt to a variety of different systems, as new systems are still being developed and introduced. We need to find ways to become more familiar and at ease with a variety of different systems and work together to develop and adapt systems that ensure we are working in integrated and structured ways.

Acknowledgements

Any errors or misrepresentations are very much my own and, although these views are also principally my own, they are based on the work of the entire Intrasis Team. I would therefore like to thank for their help and support Tom Cromwell, Vicky Crosby, Andrew Lowerre, Sarah May, Fay Worley, Claire Tsang and Duncan H. Brown. In addition I would like to thank Karen Lund and Håkan Thoren for their training and helping to repair the system when I had adapted it so much it stopped functioning. Finally, I would like to thank all of my colleagues at Fort Cumberland and all of the excellent excavators we have employed on various excavations for their hard work and patience.

References

- Andrew, G., J. Barrett, and Lewis. 2000. "Interpretation not record: the practice of archaeology." *Antiquity* 74: 525-530.
- Backhouse, P. 2006. "Drowning in Data." In *Digital*

- Archaeology: Bridging Method & Theory, edited by P. and T. Evans Daly, 43-49. London: Taylor and Francis.
- Cripps, P., and K. May. 2005. "To OO or Not to OO? Revelations from Ontological Modelling of an Archaeological Information System." Paper presented at *Beyond the artefact - Digital Interpretation of the Past - Computer Applications and Quantitative Methods in Archaeology, 13-17 April, Prato*.
- Ellis, S. 2012. "Pompeii and the iPad: an update." Paper presented at *Computer Applications and Quantitative Methods in Archaeology, 26-30 March* University of Southampton.
- English Heritage. 2006. *Management of Research Projects in the Historic Environment: The Morpho Project Managers' Guide*. Swindon: English Heritage.
- Fisher, C., M. Terras, and C. Warwick. 2010. "Integrating New Technologies into Established Systems: a case study from Roman Silchester." In *Proceedings of the 37th Annual Computer Applications and Quantitative Methods in Archaeology Conference*, edited by B. Fisher, J. Webb Crawford and David Koller. Oxford: Archaeopress
- Intrasis. 2012. Intrasis: Intra-site Information System. Accessed 25 June 2012. www.intrasis.com/engelska/index_eng.htm.
- May, S., and V. Crosby. 2004a. "Holy Grail or Poison Chalice? Challenges in implementing digital excavation recording." In *Beyond the artefact. Digital Interpretation of the Past. Proceedings of Computer Applications and Quantitative Methods in Archaeology*, edited by F. Niccolucci and S. Hermon, 49-54. Budapest: Archaeolingua.
- May, S., A. W. Payne, K. Graham, T. Cromwell, J. L. Heathcote, D. Robinson, B. Attewell, P. Cripps, V. Crosby, C. Jones, E. Lyons, K. May, S. Reilly, K. Stonell Walker, J. Schuster, M. Walkden. 2004b. *Revelation: Phase 1 Assessment*. Swindon: English Heritage.
- Microsoft Sharepoint. 2010. Accessed May 2012. <http://sharepoint.microsoft.com/en-us/pages/default.aspx>.
- Richards, J., and C. Hardman. 2008. "Stepping back from the trench edge: an archaeological perspective on the development of standards for recording and publication." In *The Virtual Representation of the Past*, edited by M. and L. Hughes Greengass, 101-112. Farnham, Surrey: Ashgate Publishing Company.
- Tudhope, D., K. May, C. Binding, A. Vlachidis. 2011. "Connecting Archaeological Data and Grey Literature via Semantic Cross Search." *Internet Archaeology* 30.
- Warwick, C., C. Fisher, M. Terras, M. Baker, A. Clarke, M. Fulford, M. Grove, E. O'Riordan, M. Rains. 2009. "iTrench: a study of user reactions to the use of information technology in field archaeology." *Literary & Linguistic Computing* 24 (2): 211-223.

Identifying and Tracing Archaeological Material with RFID Tags

Ana María López, Ana María Salinas, Eduardo Pascual,
Guillermo Ignacio Azuara, Gloria Fernández
University of Zaragoza, Spain

Elena Gallego
Segeda Celtiberian Studies Centre, Spain

Francisco Burillo
University of Zaragoza, Spain

Abstract:

In this paper we propose to label archaeological materials with RFID (RadioFrequency IDentification) tags in order to identify them in an unambiguous way. In this way, it is possible to trace the relevant information associated to them, through all the phases of archaeological work, from fieldwork to museum storage. The system has been specifically designed to be integrated with the working procedures of the archaeological team working on the Segeda project (Segeda 2012). This technique improves archaeological work in several ways. First, it speeds up the process of collecting, saving, updating and duplicating the data associated with every piece of material. Second, it increases the information that can be stored with the material and can be retrieved without connection to the database. Third, it reduces human error in transcribing information. Finally, RFID labelling facilitates the process of localizing stored material and controlling its movement.

Keywords:

Information Management, Traceability, Radiofrequency Information

1. Introduction. Traceability in the Archaeological Work

When traceability is present in a productive process, it means that, at every step in the process, information about every product is collected and attached to it without ambiguity. We know all the production parameters of a specific item or group of items and also who is responsible for this information and when and where these data were collected. This knowledge is always available because it is precisely documented.

Archaeological research is divided in several steps. It is also a chain process that, in the context we work, starts at the excavation site where the findings are unearthed. These elements are moved to the laboratory in order to be cleaned and inventoried. In the next phase, the reconstruction of ceramic pieces is addressed. These pieces must finally be catalogued and sent to a museum. At every step, archaeologists must record information about the

Corresponding author: lopeztor@unizar.es

excavated artefacts. These data must accompany the items in order to assure their provenance, control the validity of the archaeological work and conclusions, and facilitate information exchange. Traceability is essential in the work of an archaeologist, usually following a recording manual where the data to be taken, the way the items are identified and the collection protocol are precisely defined (Parks Canada 2005, 129). The information is initially recorded in a field notebook, in paper or electronic format, and later transferred to an inventory or database. In any of these information repositories, the link between data and artefacts must be clear and traceable (Sands 2009).

As will be explained below, in a traceability system any item or group of items must be precisely identified to trace its evolution through the chain process. The identification code in archaeology can be directly attached to the element (alphanumeric codes handwritten or Quick Response (QR) codes stuck on it (Martínez-Moreno, González Marcén, and Mora Torcal 2011, 134-139)) or can be stored

with it (a paper card or sheet saved in the same container). The information related to each artefact that is going to be measured and recorded can be infinite, and must be selected in terms of relevance to the research. The aims of the excavation work will define which data from a huge number of possible observations will be collected and associated with the identification code. The gathered information can be used with a classification purpose following known and fixed criteria. But different elements can also be grouped on the basis of more subjective criteria, established dynamically in the course of the research process. The analysis unit can change from a phase to the next one and new data are incorporated into the description of these units.

Any traceability system must be integrated to the work procedure of the archaeological research group. In our case, this system is going to be implemented at the excavation site at Segeda in Spain.

This paper continues with an overview of the concept of traceability and a basic description of the radiofrequency identification technology and its main advantages. Next, the context where the system is going to work is introduced. In section 5, the design of the traceability system is detailed. Finally, the main conclusions and future steps are set out.

2. Elements of a Traceability System

The basic elements of a traceability system are the identification technology, the traceability steps, the management information system and the quality evaluation procedure.

To develop a traceability system, we must first choose the technology used to identify the objects. In order to trace the evolution of an object through a chain process, it must be precisely identified. The identification code can be manually written on a paper tag or on the element itself (these are the identification methods used so far at Segeda site). A different approach is the use of one-dimensional or two-dimensional code bars that are automatically read by means of a scanner (Martínez-Moreno, González Marcén, and Mora Torcal 2011, 134-139). In our system, radiofrequency identification (RFID) tags are used. These elements will be described later

in this paper. Also the level of identification must be defined. A code can stand for a single element, or a group of elements can share the same identification number. This identification level corresponds to the analysis unit established at every stage in the research.

Once the items are identified, we can control their progress in the traceability chain. The traceability chain defines the control points of the system:

- At which places the data are going to be collected.
- When the data collection takes place.
- Which data are going to be collected
- How the data are taken and transmitted to the information repositories (what is directly related to the identification technology)
- Who is responsible for the collected information.

The information storage repositories must be established and the way the information is saved defined. The data can be stored in one or several sites and formats. This redundancy guarantees the information recovery if one of the repositories fails and also makes it possible to read the data in different ways: by visual inspection, using a scanner or accessing to a digital database.

The collected data can be used to take decisions about the process, but generally, the purpose is to document and trace the activity done. Communication channels must be carefully defined. It is especially important when the traced artefacts are transferred to different locations or even studied by different institutions. Traceability can have different scopes. Backwards traceability describes the past history of every identified item. Internal traceability defines what processes follow the elements at one location. Forward traceability deals with the definition of the future situation of the elements: where the items are going to be moved, how they are going to be carried and when the transfer is planned. The documentation must be clear and precise and, in some cases, it is important to establish a role-based access control for different users.

Finally, the quality of the traceability system must be systematically measured in order to undertake improvement and correction actions. A group of measurable indicators must be defined to determine whether the traceability system meets the required performance criteria.

3. Radiofrequency Identification (RFID)

Radiofrequency identification, RFID, belongs to the group of automatic identification and data capture (AIDC) technologies, like code bars, smart cards or techniques of biometric identification. It is based on the use of a unique identification code for every element. This code is known as Unique Identifying Digit (UID) and it is stored in the internal memory of a RFID tag. In order to automatically identify one item, a RFID tag is attached to it. The UID that stands for this tag and for the object can be read and communicated to a control system by means of an RFID reader that interrogates the tag using an electromagnetic wave of the radio wave frequency band. The tag answers with its own radiofrequency signal. In this way, we know that one specific item has passed through the read range of the RFID reader in a known time interval. The process is outlined in Figure 1. No precise alignment between the reader and the RFID tag is required. If the reader is configured to scan for tags continuously, it is possible to control which elements have passed near the reader's position without any human interaction.

The basic elements of an RFID system are the RFID tag, the RFID reader and the RFID software known as middleware (Glover and Bhatt 2006). The RFID tag acts as the identification element as mentioned above. It is formed by an electronic chip, an antenna and a storage memory. This memory has always reserved blocks where the unique code UID is saved. These specific blocks cannot be modified so the integrity of the identification code is guaranteed. This basic memory can be extended to save more information. There is reserved space for storing access passwords, error detection and correction codes (as the Cyclic Redundancy Code CRC) or Protocol Control bits. The address of these data in the memory map is system-defined. Finally, the tag memory can have a bank, the user memory zone, where new data can be written, read and modified with the RFID reader. The format and size of the data

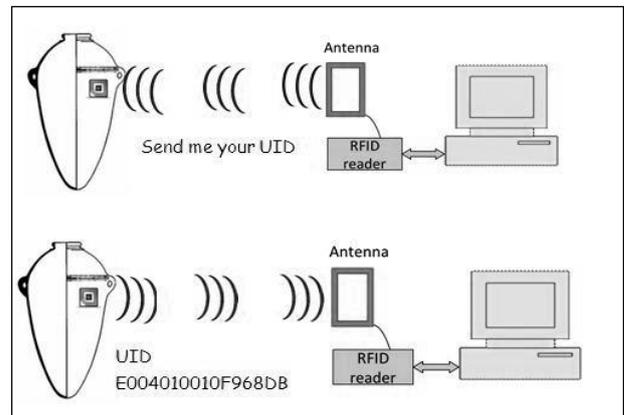


Figure 1. An RFID tag and a reader communicate by means of electromagnetic waves.

are user-defined. In this way, part of the traceability parameters can be dynamically saved inside the identification element and travel with the object identified. The tags can appear in multiple physical formats that can adapt to different applications and work environments.

The RFID readers have several functions. In most cases, they provide to the tags with the energy they need to transmit their radiofrequency signals (though there are active tags that have a battery). They communicate to the tags through radio-wave signals and get, for example, the Unique Code that identifies the item. This information is sent from the reader to the information system. The control system can be the origin of additional information that is saved in the user memory of the tag. The reader writes, reads, and modifies these data if the tag has a user memory bank.

The readers can be implemented in different formats. There are handheld readers and readers that are attached to computers. Some readers are simple and inexpensive, while others are more complex and robust devices that can work in very harsh environments. In some cases the tags are written inside the reader, like in RFID printers, but generally the tags just need to be inside the RFID reader read-range. The size of this communication area depends on the frequency of the radio-wave signals used and on the size of the antennas of readers and tags.

The third element of the RFID system is the RFID middleware or software that converts low

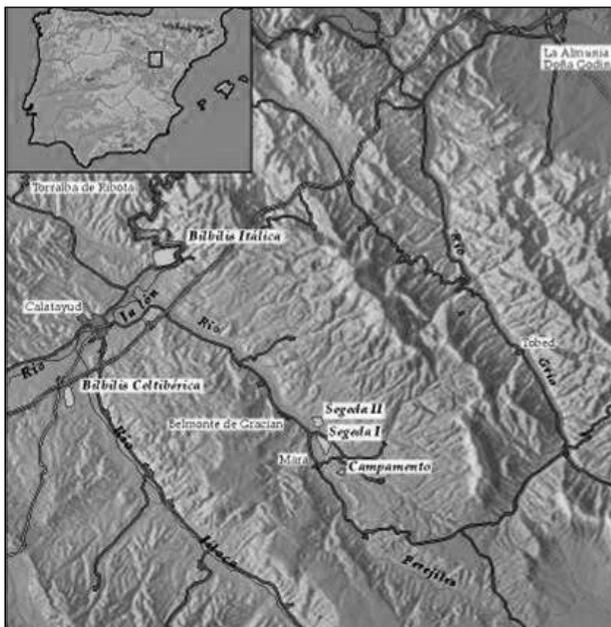


Figure 2. Segeda site is localized in the North-East of Spain.

level RFID hardware information from the reader into information useful to the system. The data can be used in an event driven process or merely can be saved in a system database. Though there are commercial middleware, the versatility of RFID systems is based on the possibility of design application specific programmes.

There are several parameters that determine the behaviour of an RFID system. We highlight here the importance of the frequency of the waves use to communication. There are several frequencies available for RFID systems: Low frequency (9-135 KHz), High Frequency (13.56 MHz) and Ultra High Frequency bands (433 MHz and 860-960 MHz). Some RFID applications, like animal identification, can only use one band (LF). The main performance difference is the increment of the reader range with frequency. If the elements to be tagged are metallic or have a high water content, only HF system can be used to prevent detection failures.

- RFID technology offers several advantages over others automatic identification technologies like code bars or biometric techniques when used in a traceability system:
- The read of the identification code is almost

instantaneous and can take place even without human interaction. It is not necessary to align precisely a scanner with the identification item (as in the case of bar codes or digital prints) to get the information. Being close to a RFID reader is enough.

- RFID tags contains a unique code (different from the case of code bars) There are no identical RFID tags.
- RFID tags can dynamically store extra information from the traceability process. These data can be modified at any time.
- RFID systems are flexible and versatile due to the possibility of frequency selection, the different physical support for the tags or the middleware design.

4. Segeda Site

The versatility of RFID technology makes possible to adapt the traceability system to the current working methodology of the group researching the archaeological site of Segeda.

“Segeda I” is located in the Northeast of Spain close to the village of Mara in Zaragoza province (Fig. 2). This site is situated close to the basin of the Ebro River, near its confluence with the Jalón River. Segeda was a celtiberian city-state, the largest one in northern of Spain, extending over 40 hectares. It played an important role in the Roman Wars in which the Iberian Peninsula was conquered. During the Celtiberian Wars (153 and 133 BC) Segeda was devastated by roman soldiers. It is possible to put the date 153 BC like “terminus ante quem” for all the materials found in this site (Burillo 2006, 203-240).

Archaeological work began at Segeda in 1998. Between 1998 and 1999 surface prospectations were undertaken to delimit the site. From 2000 year until now (2012) archaeological excavations have been carried out in different areas of the site (Fig. 3).

5. Design of the RFID based traceability system

As outlined above, the design of a traceability system implies the selection of an identification

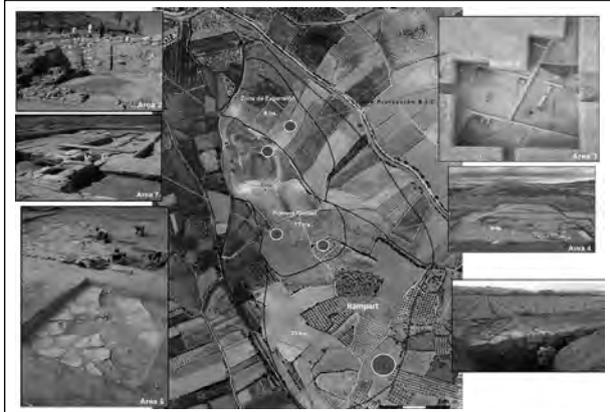


Figure 3. Excavation areas at Segeda site.

technology, the definition of the traceability chain, the information repositories and communication channels and, finally the way the system is evaluated. Our work is based on a number of premises:

- Full integration with the traceability methods used so far by the Segeda research group.
- Increase the automation of data collection and recovery
- Reduce uncertainty regarding provenance.

5.1. Identification technology

We use RFID tags operating in the UHF band to identify the findings of the site. These tags have a user memory of 512 bits, a size of 97 x 27 mm and are attached to an adhesive inlay. Their price is 28 euro cents each, for an order of 1000 units. Greater quantities orders allow for lower unit prices. The identification level resembles the traceability protocol used so far at Segeda. Initially, group identification is addressed. Fragments that are found together shared a RFID tag. Only when a whole piece is found or after reconstruction of a ceramic element, single identification level is used.

5.2. Traceability chain

The traceability chain consists of four steps (Fig. 4). The first place where information is collected is at the excavation site, at the moment an artefact is found. The artefacts are moved to the archaeological laboratory where they are cleaned and inventoried and, afterwards, follow a reconstruction

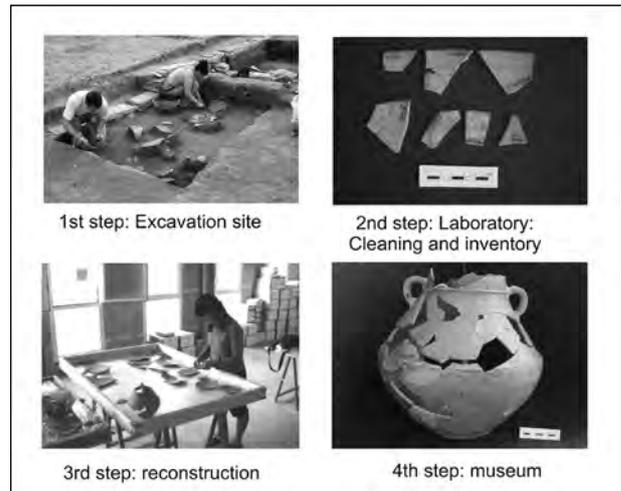


Figure 4. Check points of the traceability system.

process if possible. Finally, some of these pieces are transferred to a museum. These four control points will be described in more detail below. The data that are going to be traced and saved are the same ones that have been recorded so far by the Segeda research group.

5.3. Information repositories

The information is saved in four different formats: an electronic database, the RFID tag user memory, on the archaeological item itself with an ink mark, and handwritten on the RFID tag. This redundancy is intended to keep the old storage formats and the new ones working together in the test period. Not all the data are saved in the four formats.

The traceability system is intended to an internal use. All members of the research group have unlimited access to the information saved in the database and the RFID tags. If necessary, both repositories can be protected with an access password. A RFID tag can be locked so that only authorized users can read the stored information.

5.4. Evaluation of the traceability system

In order to evaluate the performance of the traceability system, several items must be measured. As initially the RFID system is going to work together with the old identification and control methods, it is possible to compare aspects like:

- Wrong data.
- Lost data.
- Time requirements.

Additionally, information from the users (archaeological researchers and collaborators) must be analysed. Opinions about the ease of use, the robustness of the devices or the clarity of the software must be taken into account to implement improvements to the system.

5.5 The traceability checkpoints

This section details the characteristics of each checkpoint of the traceability chain. First, the actions that must be carried out at each location are described. Then, we explain how these activities are modified because of the integration of the RFID technology. Finally, the required hardware is listed.

Step 1: Excavation site.

The first point for data collecting is the excavation site. When similar items are found in the same stratigraphic level and zone, they are bagged together. A paper card is used to write information about the date of recollection, the physical location and the material of the artefacts. This card is saved inside the bag. The data included in this card are shown on Table 1. These bags are sent to the archaeological laboratory.

EXCAVATION SITE	
Geographic information	<ul style="list-style-type: none"> • Zone • Lot • Area • Sector • Space • Stratigraphic unit • Square • Total station coordinates • Height
Chronological information	<ul style="list-style-type: none"> • Campaign • Date
Element information	<ul style="list-style-type: none"> • Material • Bag number

Table 1. Data to be collected at the excavation site. Common data are shaded.

Figure 5. Handheld RFID reader.



In the new system, the paper card attached to each bag is replaced by a RFID tag and the information saved in the memory of this tag. As this tag has a paper inlay, the handwritten information can be kept.

Some of the data to be saved at the excavation site are common to all the elements found on the same day. Some geographic and temporal data are known in advance, so we have decided to record them in a number of RFID tags before arriving to the field. If these tags are not used, the data can be modified because the tag memory is rewritable. In this way, recording work at the site is reduced. The data are provided by the user through the touch interface of a handheld RFID reader (Fig. 5). Then the tag is recorded just by clicking a button. The information saved can be checked or modified at any time.

The information is also stored in a local database inside the device. At the end of the day, it is transferred to the global traceability database. This database has the same structure as the one used so far by Segeda's research group. The only difference is that a new field, containing the tag UID, must be added to the tables of the database.

The hardware required at this location is an RFID reader, a computer and a printer (if the card

Campaign	Museum code	Area	Inventory number				Material		

Figure 6. Structure of the internal code or “sigla” used by Segeda group.

used so far is going to be created) before excavation and a handheld RFID reader during excavation. The RFID reader can be the same, as the handheld reader can communicate to a computer.

Step 2: Archaeological laboratory.

Once in the laboratory, the fragments are cleaned and marked with a number that is shared by all the elements of the same bag. This internal identification code, known as “sigla”, is the one used in inventory and contains information about the year and area of work (two digits), a code assigned by the museum to the excavation (two digits), the inventory number (a correlative number formed by four digits that differentiates the bags with similar data) and a character that indicates the material of the fragments inside the bag. Its structure is shown in Figure 6.

Besides, the content of the bag is analysed. The fragments that seem to be part of the same piece are packaged in a second bag. A new card is created for this new bag with the same information that contains the card of the original bag. This new bag is saved inside the initial bag (Fig. 7).

In our system, the inventory number should be the new information that must be added to the RFID tag. However, some of the data that formed part of the “sigla” are already recorded in the tag memory so, it is not necessary to save the whole code. Further, the RFID software can automatically calculate the codes that are going to be written on the fragments.

When the elements of the bag are classified, it is the moment to confirm their material and correct it if necessary. The number of sherds is counted and added to the RFID tag and also the number of elements not identified. If it is necessary to duplicate the tag, the RFID reader just reads the original tag and records its data in a second tag at the same time that saves it in the database. Table 2



Figure 7. When some fragments are identified as belonging to the same ceramic piece they are bagged together. The new bag is saved inside the original one. Both bags contain cards with the same data.

gathers the information to be collected at this step of the traceability system.

The hardware elements required to implement the traceability system in the laboratory are the same that those used in the excavation site.

Step 3: Reconstruction.

The third control point is the reconstruction laboratory that can be the same or different than the previous one. The aim is to join fragments in order to obtain a new element. This piece will be identified with a new number and a new card. The previous classification in bags makes the task easier. The unidentified artefacts are left in the original bag whose information card is destroyed.

The RFID system must create a new tag to be attached to the piece, but also it is going to update the tag of the original bags that is not going to be eliminated.

ARCHAEOLOGICAL LABORATORY: Cleaning and inventory	
Original bag (tag created at the excavation site)	<ul style="list-style-type: none"> • Inventory number that completes sigla code. • Number of fragments identified to belong to a piece <ul style="list-style-type: none"> • Number of amorphous
Secondary bag (tag duplicated from the original one at the laboratory)	<ul style="list-style-type: none"> • All the data from the original RFID tag • Number of fragments.

Table 2. Data to be collected at the archaeological laboratory. Cleaning and inventorying phase.

The new RFID tag specific for the reconstructed piece must store the code that identifies it. In the inventory and database the link between this code and the “sigla” of the original shards is created. In the tag memory the numbers that identify the source bags are recorded. This data will help to follow backwards traceability without consulting the inventory or database.

At the same time, the existent tag is updated. The information added indicates that a number of fragments from the bag are part of a piece with a known number. In this way forward traceability is fed back. All the new information is gathered in Table 3.

When a new piece is reconstructed, we leave the group identification level (one tag for a group of fragments) and assume single item identification level.

The required hardware is the same as the used in the previous phase.

Step 4: Museum transfer.

When an artefact is transferred to a museum, it must be accompanied by a record including all the relevant information in the format set by this institution. Assuming the museum agrees to use the RFID system, it may be necessary to add more data, but not to modify the fields already stored. We would have to adapt the RFID application

to visualize the information fields following the museum requirements. The RFID tags belonging to the traceability system should be updated with the date of the departure of the piece. At this moment, there is no contact with any host museum.

5.6. Software interface

The software application developed to read and write information from the RFID tags has been created using the programming language Visual C#. This language eases the creation of visual interfaces that simplify the user operation. Besides, the RFID devices can be controlled with a set of functions written in this language that can also easily communicate with databases. The basic events that can be controlled with the application software are tracing tags, reading and writing tags and deleting tags (Fig. 8).

To interact with an RFID tag, we must first be sure that it is inside the reader range. So, before writing or reading a tag it is necessary to trace which tags are near the reader.

The UID’s of these tags are listed on the interface (Fig. 9). In this way we can select the tag we are going to work with. Choosing the read menu, all the data stored inside the RFID tag are displayed in an interface that has been designed in order to resemble the Filemaker recording files used so far, as shown on Figure 10.

ARCHAEOLOGICAL LABORATORY: Reconstruction	
Existing tags are updated.	<ul style="list-style-type: none"> • Identification code/s of the piece/s created from fragments from this bag.
New piece RFID tag	<ul style="list-style-type: none"> • Identification code of the piece. • Identification code (sigla) from the bags that have contributed to the piece.

Table 3. Data to be collected at the archaeological laboratory. Reconstruction phase.

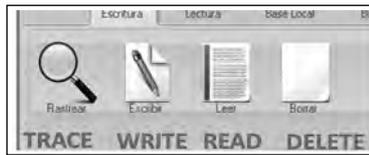


Figure 8. Top Left. Basic actions of the RFID software.

Figure 9. Left. After selecting the tracing option, the list of tags are shown on the screen.

Figure 10. Top Right. Screen of data associated to an RFID tag.

All the important data are included and can be completed or modified from a touch screen or from a keyboard. When the write option is clicked, the data are added or updated in the tag memory, but also in the database. The field that guarantees that the information saved in the RFID tag is the same as in the database is the unique identification code of the tag. This number is read from the tag when any action is done on it and compared to the codes identifying any record in the database. If an UID is not present in the database, a new record is created. If a tag is deleted, the associated record in the database is not eliminated but marked as deleted.

All these actions are carried out almost instantaneously. The software also verifies that the data to be recorded agree with the established format and sends an error message otherwise.

6. Conclusions

In this paper, a traceability system, based on RFID technology, is presented. This control system has been specially designed to be integrated to the

research work developed around the Segeda site in Spain, though it can be adapted to a different archaeological research group and different artefacts. If the artefacts are large enough, the RFID codes can be directly attached to them. The system prototype has been partially implemented and successfully tested in the reconstruction laboratory step. The whole prototype is going to be tried out in summer 2012.

The use of RFID technology eases and automates the process of data collection. The information can be updated simultaneously in the RFID tag and in an electronic database without ambiguity. The storage capability of the RFID tags makes it possible to access to the relevant information of the artefacts without a network connection to the database. It is not possible with other automatic identification technologies like code-bars, because their storage capability is very low. Datamatrix and Quick Response codes have greater storage possibilities, but once the information code is created there is no possibility of adding or modifying the data saved. The memory of

an RFID tag can be rewritten more than a hundred times.

Information transmission by means of radio-wave signal does not require a perfect alignment between RFID tags and readers. An RFID reader placed at the exit door controls which elements leave or enter a room. An RFID reader can even scan items inside a closed drawer or box, without the need to open it to see what is inside.

Finally, traceability implementation forces us to clarify and thoroughly understand what happens at every step in the process and what information is really relevant. It helps to optimize the collection and management of data, reduces uncertainty and mistakes, and optimizes our knowledge of the present and past locations of the artefacts.

References

Burillo, F. 2006. "La Ciudad Estado De Segeda I." In *Segeda y Su Contexto Histórico: Entre Catón y Nobilior*

(195 Al 153 a.C.): Homenaje a Antonio Beltrán Martínez, edited by F. Burillo, 203-240. Zaragoza: Centro de Estudios Celtibéricos de Segeda.

Glover, B., and H. Bhatt. 2006. *RFID Essentials (Theory in Practice)*. Sebastopol: O'Reilly Media, Inc.

Martínez-Moreno, J., P. González Marcén, and R. Mora Torcal. 2011. "Data Matrix (DM) Codes: A Technological Process for the Management of the Archaeological Record." *Journal of Cultural Heritage* 12 (2): 134-139.

Parks Canada. 2005. *Parks Canada Archaeological Recording Manual: Excavations and Surveys*. Version 1.0. Quebec: Parks Canada.

Sands, R. 2009. *Artefact Studies, Artefact Data and Information Technology: A Point in Time Review*. Dublin: UCD Schools of Archaeology.

Segeda 2012. "Segeda Foundation Website." Accessed July, 2012. <http://www.segeda.net>.

Matera Città Narrata Project: a Multimedia and Multi-Platform Guide for Mobile Systems

Eva Pietroni

CNR Institute of Technologies Applied to Cultural Heritage, Italy

Abstract:

The aim of Matera Città Narrata (Matera Tales of a City) project is the creation of a digital platform able to support the public before and during the visit of Matera (South Italy, Basilicata Region, Unesco World Heritage since 1993), through the access to cultural contents both in remote and on mobile devices, on and off line, while attending places, sites, itineraries. Several contents have been developed: 3D reconstructions of the city and of its landscape in eight chronological phases, movies, multimedia, virtual reality environments, audio guides, declined according to different communicative formats implemented for different and multi-platform devices. Matera Città Narrata is not a descriptive traditional guide inspecting monuments and mentioning who made them and when. On the contrary it tells the stories that took place in those ancient sites, real fragments of life, myths, events, characters, memories, times.

Keywords:

Mobile Technologies, Storytelling, Usability, 3D Reconstructions, Intangible Heritage

1. Discovering a World Heritage Rupestrian Site Through the Use of Mobile Technologies

Matera Città Narrata project has been realized in 2009-2011 by CNR ITABC and it was supported by the Regional Promotion Agency (APT) and the Basilicata Regional Government. Its aim is to provide the tourists visiting the city with cultural contents related to the places, the history and the intangible heritage in order to enhance their cultural experience and encourage them to remain and spend a longer time in this wonderful city and in the surrounding landscape.

Matera seems, or better is, an enormous sculpture, a city modeled in the calcareous rock of the deep canyon called Gravina, where thousands of natural caves have been used by men and women to find protection and organize their life, from 12.000 years ago until now (Fonseca et al. 1999). The territory in fact is plenty of rupestrian sites: prehistorical settlements (paleolithic and neolithic age) rupestrian churches, caves used as houses, pastoral villages excavated in the calcareous rock, jazzo's (spaces destined to animals and surrounded by a dry stone wall), medieval manor farms, rupestrian monasteries.

During the Neolithic Age (around 8000 – 5000 B.C) man did not live in the grottos. After the discovery of the agriculture, he began to live in fortified and sheltered villages encircled by moats that had the function to isolate and protect the group of primitive huts. Here the community lived, worked, and cultivated the land. The grottos close to the village were destined to sanctuaries and cemeteries. Later men began to occupy and live in grottos in a more systematic way, especially to find protection from the barbarian danger during the first medieval age (Fonseca, 1979). In the same period many pre-existing caverns were modified or hollowed out so as to be dedicated to worship and they were often decorated with notable mural paintings. In the Rupestrian Churches Park, (about 8,000 hectares in Matera and Montescaglioso territory) more than 150-200 rupestrian churches have been counted and today their discovery and identification is still in progress. Most part of these rupestrian sites continued to be used in the successive times, up until the half of the 20th century; the grottos were used as dwellings, animal shelters, places of worship and burial grounds and for religious ceremonies.

The most famous areas of the city are called "Sassi" (Sasso Barisano and Sasso Caveoso) where the frontages of the houses are built with calcareous bricks obtained excavating the inner caverns, just to protect the entrances. In most parts of the Sassi the

Corresponding author: eva.pietroni@milib.itabc.cnr.it

streets are located on the rooftops of other houses (Laureano 1993) and such a spatial configuration stimulated famous writers, film directors, artists to bring to Matera their art. Coming to Matera and discovering its caves is something like to re-discover the wormth of the mather, an ancestral feeling.

People lived inside the Sassi until 1950s, with animals and in really bad conditions. At this time the Italian Government decided to evacuate the inhabitants of Matera, defined “national disgrace”, deporting them forcefully in the new buildings designed by famous Italian architects (Quaroni, Piccinato, De Carlo, Fiorentino, Lenci), built outside the ancient city. So Matera became a laboratory of the most advanced tendencies of the Neorealist urbanism linked with Italian Rationalism. In the new districts people started to live in much better conditions but they lost that ancestral status based on communal life all around open spaces. Today, after a period of abandon, the Sassi of Matera have been reevaluated. In 1993 the city became Unesco World Heritage and many caverns have been restored (even if some work is still needed) to host craft activities, hotels, shops; also some families have come back to live in their ancestor’s caves (Fig. 1).

In 2009 the Regional Promotion Agency (APT), in collaboration with CNR IRAT, collected some statistic data on tourist presence in Matera and on the time they usually spend in the place in relation with the huge amount of sites offered by the context. What came out from this investigation was that until today the visit of Matera has been usually included in wider package tours (from the Adriatic and Ionic coasts) and people visit this place in few hours, really an insufficient time to enjoy and understand the place and its extraordinary cultural heritage. Moreover the visit is often accompanied by local physical guides, (some sites and excursions cannot be found and experienced without them). This is why the Basilicata Region decided to promote and support this project for which CNR ITABC has been involved as coordinator. From the begin we tried to plan a system that could not enter in competition with local guides (many people base their job and subsistence on this activity), and that could be perceived as an added value to what is already existing in terms of cultural offer. We didn’t want to make a traditional guide, describing



Figure 1. Matera, view of the Sasso Barisano and Gravina river, (2010).

monuments and telling who made them and when. This kind of contents remains in the background while a novel approach in storytelling was the real issue we decided to follow (Antinucci, 2004 and 2007). For instance we decided to orient a great effort to the 3D reconstruction of the landscape and of the urban evolution in eight different ages, from two millions years ago until today: this is the first time such a work has been developed and presented to the public and we believe it is something very important in order to understand and communicate the essence of Matera. Moreover this kind of contents cannot be transmitted and shown to the public by physical guides without the aid of something visual. For the same reason we decided to involve local people, inhabitants, children or artists to give their particular and original interpretation of Matera’s story and life, through their own memories and experiences.

2. Usability and Inclusivity

The contents, organized in an online database, are accessible, for free, from a variety of mobile technologies.

The project is quite pioneering as the state of the art in this field is still scant and our first attempt has been to create something new not only, and not necessarily, from a technological point of view but also and especially in the general infrastructure, in the design of the applications, in its robustness and possibilities to be maintained by the local administrations and, of course, in the

choice of the cultural contents, as of the proper communicative formats. This is the right approach to let such a project survive, to make it accessible and usable for more than a brief initial promotional launching. In the last years many projects realized to provide the tourists with cultural guides on mobile devices are died after few months from their opening. Sometimes this was due to the extremely sophisticated technological solutions that have been employed, requiring for instance the combination of several inter-dependent services (wi-fi, GPS localization, bluetooth) that make the general infrastructure very delicate and fragile. We wanted to built something more robust, that could be able to work even in absence of a GPS signal or a wi-fi connection (we provide these systems but visitor are not totally dependent from them). In fact the urban spaces of Matera are often very narrow and places are located one above the other, so GPS could not be always precise (Wagner et al.2005; Takacs et al. 2008).

Thus we tried to calibrate and fit the project to the real needs of the people visiting Matera. If solidity and stability of the system are fundamental, accessibility and facility for public to use the platform are imperative. One of the keyword of the project is “inclusivity”; this means that all kinds of visitors can enjoy in some way of the available cultural contents, using the more or less advanced technology they own. Contents, in fact, have been adapted to different formats that can be accessed as multimedia contents on a website (www.materacittanarrata.it declined for desktop systems and also for smartphones) or as stand alone applications to be downloaded on smartphone and tablet (iOS, Android, Java). Also a simple audio guide has been implemented, for old generation call phones, and a .pdf version of the guide that can be printed and brought while visiting the places. We also provided a wi-fi infrastructure to cover the most attended places in the Sassi neighbourhoods (6 wi-fi hotspot in total, each one with a range of about 100 mt). Here people can easily access the website for free or download applications and contents in local, using them afterwards while continuing the visit. In the places not covered by the wi-fi infrastructure users can access the on line contents using the telephonic connection according to their pay plan. The project, coordinated by ITABC and realized also through the collaboration of private companies, has



Figure 2. Sign put in front of the monument with all the information to access the cultural contents.

been concluded and now it is open to public. At the moment all the contents are given for free, APT Basilicata and Matera Municipality will support the costs in the next months and then they will decide how to manage the services in the future. Our effort, in the digital platform design, was oriented also in keeping the costs of maintenance as lower as possible.

3. Contents and Accessibility

Before visiting Matera the user can access the website from home, school or every other static base and he finds cultural contents organized in different multimedia sections. He can explore on line or download contents in different formats (*.mp3, *.pdf, stand alone applications for smartphones or tablet) in order to use these materials during the successive visit of Matera, without the need to connect again to the internet. Standalone applications for smartphone have been implemented in different versions for the most common operative systems: iOS, Android, Java.

While visiting Matera the user can use his device to access contents in the following ways:

1) "Database" access while moving through the site: the visitor stops in front of a monument and finds a sign with its name and its ID (a 3-digit number), together with a toll free number or a URL. He has some possibilities (Fig. 2): he can connect to the internet (the place is in a wi-fi area or he uses his own pay plan) he can access the web site and, keying in the ID, he reaches information about the monument and other related contents. He can also download a stand alone application if he owns a smartphone, in order to avoid to connect to the internet in the successive steps of the visit.

On the contrary, if he has not a connection to the internet, he can call the toll free number indicated on the sign. In this case a synthetic voice (IVR, Interactive Voice Response) will answer asking to choose the preferred language (Italian, English, German); after it will ask the monument ID and finally if the user's preference is for an audio guide (2-3 minutes) or a MMS (30"- 40" audio explanation and 5 images). After this third choice the proper content will be sent.

2) "Guide" access while moving through the site: the user is in front of the monument and he has just received the information about it. After this and in relation with this, the visitor can enjoy thematic itineraries, witnesses, video clips with personal stories or memories told by local people related to that monument. The website has been realized in Flash (for desktop platforms) and in html (for mobile phones). It follows a specific narrative style and only accredited authors can apply updates and changes in order to preserve the general approach. But we also added a separate section open to external contributors from local cultural operators, in which texts, images, movies can be uploaded.

Moreover APT Basilicata, in collaboration with Matera Superintendency, is now planning to open a local multimedia urban centre in the undergrounds of Piazza Vittorio Veneto (a strategic square of Matera from which the main itineraries through the Sassi and up to the Civita develop), where the tourists will find information about the project. In these caves some large projections will give the public a

suggestive overview of the most original contents that have been realized: 3d reconstructions, stories by local inhabitants, performances and comments left by the artists, visions of the sites before and after 1950 and so on.

In the urban centre visitors will also have the possibility to find a fast internet connection in order to download cultural contents from the web site to their own devices before continuing the visit.

So the general aim is to realize an integrated multi-disciplinary and multi-dimensional cultural project, a virtual heritage network able to ideally connect contents, places, ages, authors and users, real worlds and virtual cyberspaces. According to this approach (and following a more general tendency of the actual time) the border line between on and off line contents access become more and more blurred, as people begin to live in a continuous interchange with local and remote experiences.

4. Sections and Technical Development

In the following paragraph the main contents are going to be explained.

Guide:

The guide is one of the main sections of the project, available in all outputs. As mentioned before, our goal was not to create a guide following the traditional descriptive approach of monuments and sites. Each place/story is told through a movie (or, in the case of the iPad application, through a slide-show with an audio comment) composed by 10-15 images; if the user is in front of the real place he doesn't need an oversized multimedia format. The guide can be accessed in two modes: single sites or itineraries. In single site mode you can access one site at a time, by choosing it on a map or by keying in its ID. Itineraries provide also a narrative connection among several sites. There are four of them: two in the Sassi neighbourhoods, one up to the Civita and one in the territory outside Matera, along the Gravina.

Fly on the territory:

This section (available only in the iPad application) aims at the real time 3D exploration

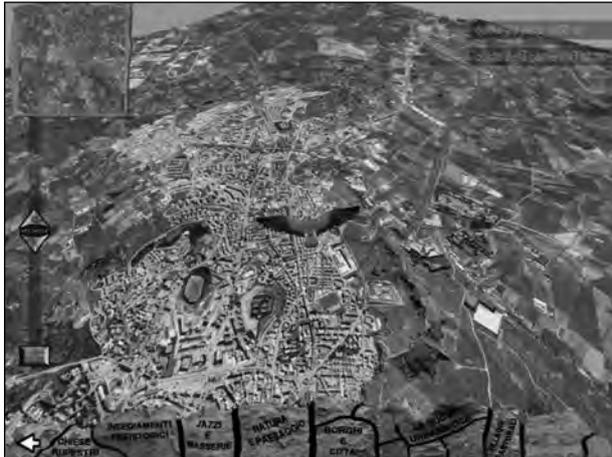


Figure 3. Fly on the territory main interface, iPad application.

of the territory all around Matera. It is an “holistic” vision of the context where the user has the possibility to visualize and locate the main typologies of cultural heritage distributed all over the area (Fig. 3).

The DEM was generated from topographical surveys with different resolutions: 1:10.000, 1:2.000, 1:500 (thanks to Ufficio Sassi of Matera) and its final resolution is 2 mt. It is managed through many LODs (level of details), in fact for the real time visualization in the iPad 1 (considered as basic standard) we needed to limit the geometry to 20-25.000 polygons. The 3D model, divided in many tiles according to the level of detail adopted, was mapped with Quickbird satellite photos (80 cm resolution), and with an IGM map where the location and identification of all the topographical themes is easier. The virtual exploration is controlled by a third person camera: moving and tapping the



Figure 4. 3D reconstruction of Matera in the Norman period (IX-XI century AD), (rendering in Vue by Raffaele Carlini & Stefano Borghini).

iPad the user can guide a Lesser Kestrel (a typical little hawk living in Basilicata) flying all over the territory; it can loose or gain altitude and speed, turn left and right acting the proper animations. The user decides the direction to apply to his movement but the bird keeps a certain degree of autonomy. In fact the camera has not a rigid constraint and a fixed distance from the avatar-hawk but a flexible behavior that produces a nice effect.

The user can modify the textures of the territory choosing the map or the satellite image, he can also visualize some thematic layers, as 3D icons located on the landscape: prehistorical sites, rupestrian churches, shepherd’s settlements, manor farms, new urban districts, towns near Matera and so on. Some of these icons are interactive and let the user access to more detailed information (movies) about specific sites (Fig. 3).

The GIS project integrates also data connected with the reconstruction of the potential ancient landscape (Forte 2008) of Matera, (vegetation and soils map, geological and paleo-botanic data) that we propose in another section of the project.

Browse through the Ages:

The section, available in all outputs, is dedicated to the evolution of the landscape and of the urban context of Matera through the time. This evolution process is described through video clips based on 3D rendering showing the context in eight

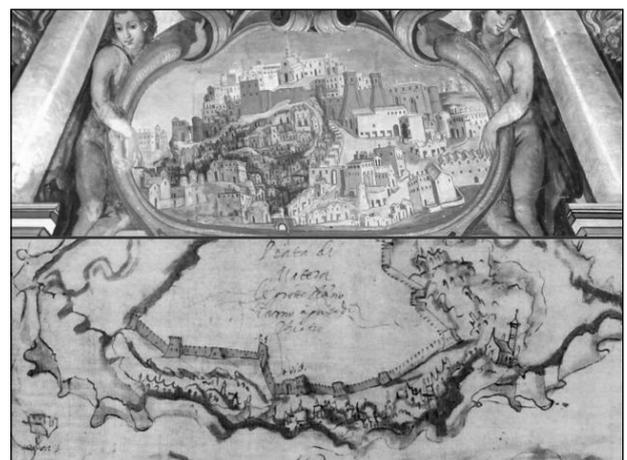


Figure 5. iconographic sources used for the reconstruction of Matera during the Renaissance period and info-graphics.



Figure 6. web site with the 3D reconstruction of Matera seen from Belvedere della Murgia in different ages: actual, pleistocene, neolithic and norman period.



Figure 7. Historical photos are compared to show how the Sassi have been changed through the time: in 1949 people were still living there, after 1950 they were moved in the new buildings and the Sassi were abandoned and closed.

different ages: the emersion of lands from the sea more than two millions years ago, the terrain and “Gravina” canyon formation, the birth of the first neolithic entrenched settlements (5000-4000 B.C.), the diffusion of life in natural caves, the construction of the “built” and fortified space in the Middle Ages, the expansion from the Civita to the Sassi suburbs, from Sassi to the planar area where the modern and contemporary city has developed.

A great effort has done to model in 3D Matera and its landscape in eight chronological phases. The

main purpose of these reconstructions is to show the general transformation and the growth of the whole shape of the city rather than focusing on specific buildings or monographies.

The 3D models have been realized in 3D Studio Max and Vue, starting from the topographical data given by the Urbanistic Department of Matera Municipality (CAD data mainly), from historical documentation, iconography, literary citations (Figs 4 and 5) (Volpe 1818, Gattini 1882, Tropeano 2003, Demetrio 2009) and finally from the archaeological investigations and studies carried on by CNR IBAM and the Basilicata Superintendency.

No further topographical acquisitions have been made by our team. A critical info-graphic apparatus was also elaborated in order to highlight, on the actual images of the town, the interpretative process followed and the correspondences between the 3D reconstructions of the past and what is still visible today.

Virtual Exploration through the Ages:

The 3D reconstructions of the past chronological phases have been used also in this section, available in the web site and in the iPad application, in order to create an “augmented” experience. The user can explore some views of the urban and natural landscape and observe them in their process of historical evolution and transformation. Six main views (indicated on the map with orange icons), are panoramas in 3D, reconstructed in eight chronological phases, from 2 millions years ago until today. In each of them the real actual image and the 3D reconstructions of the past ages are perfectly overlapping and it is possible to switch among them, perceiving and understanding the whole chronological progression. A cursor allows to modify the transparency of the active level in order to have more possibilities of comparison with the actual landscape (Fig. 6).

Other views consist in historical images (most of them belonging to Giuseppe Buonsanti’s archive) taken inside the Sassi areas (blue icons on the map). Every couple of images shows the life and the place before and after the depopulation of 1950s or its successive restoration (Fig. 7).

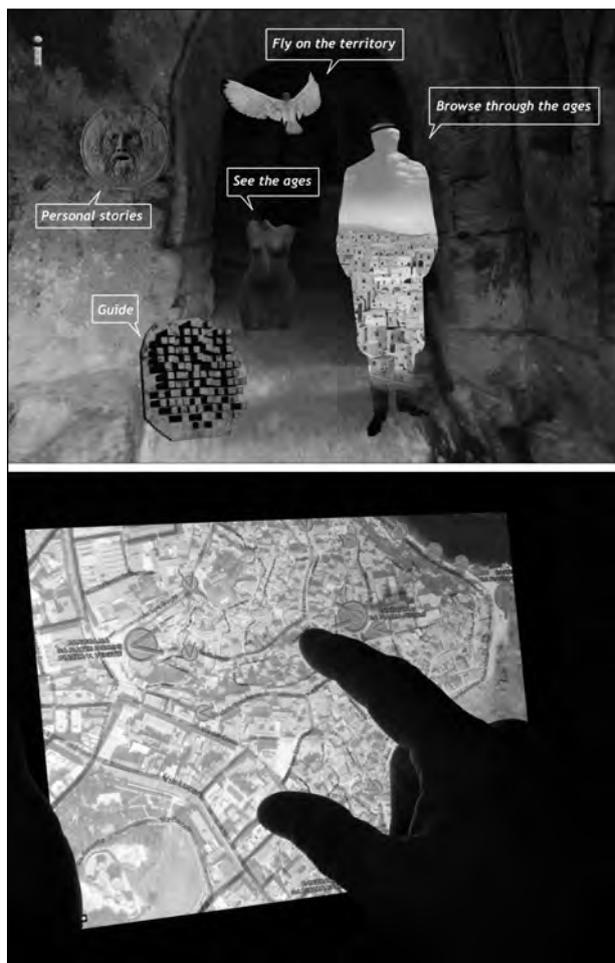


Figure 8. iPad application developed in Unity 3D (programmer Claudio Rufa, E.V.O.CA.): main menu and map - interface.

Personal stories:

A lot of people have been involved in this project: they have narrated the places of the city and of the territory on the basis of their own memories and encounters, their own visions, discoveries and life experiences. Also many children from the local schools of Matera have been involved; finally ten well-known artists gave their contribution, bringing their art to Matera.

In our application these witnesses are fixed in video clips, each one two- three minutes long. The section is available in the web site and in the iPad application. In order to avoid to make the iPad application too heavy, we included in local only ten personal stories (witnesses) while the other ones

are available through a wi-fi internet connection. This section is of fundamental importance for us as it allows to multiply the points of view, the expressive registers and the cognitive and aesthetic perspectives from which the city is considered.

Finally some numbers: 86 movies, 30 slideshow, 8 chronological phases reconstructed in 3D, 6 panoramas for “augmented” visualization, a VR environment, 67 sound tracks. A flash multimedia web site, a html website for cellphones, MMS, audio guides in IVR, guides in .pdf, an iPad application (1GB), multi-platform applications for smartphone (among which a Treasure Hunt “O’ Munaciedd” for children realized by CNR ITD in Palermo available for iPhone), a wi-fi dorsal in the Sassi neighbourhood.

The iPad application is one of the most advanced output we realized (1 GB, AppStore, category Education) . For this application we used Unity 3D for iOS as developer platform. The main reasons that led us to this choice are: the short development time consuming in relation with the complexity of the application, the strength and stability of the engine and its well written documentation. The target of this application was the iPad first generation but it runs also on iPad2 and iPad3 (Fig. 8).

5. Conclusions

Despite the Basilicata region is one of the richest territory in Italy, but also in the world, in terms of Cultural Heritage, it is still partially unexplored in comparison with the wider touristic and cultural circuits. The absence of proper infrastructures for mobility (railways, airports, public transports) causes problems of accessibility to the cultural and naturalistic sites. This situation is due also to the fact that the Italian Government has always limited the economic and cultural investments in this territory (except for the petrol extraction) and thus it still remains quite “marginal”. For this reason one of the primary need of the project was to give a new evidence and coherence to these cultural paths.

“Matera Città Narrata” project aims at enhancing the cultural experience of Matera and of its territory, (Unesco World Heritage since 1993) providing contents about the physical and intangible heritage of this unique city, where men established since 12.000 years ago, living in



Figure 9. the first three images are referred to the smartphone website interfaces (Netagency and Imagimotion s.r.l.). The last image is related to the application “O’ Munaciedd” developd by CNR ITD-Palermo.

caves. We didn’t want to confuse the instruments (new technologies subjected to fast obsolescence) with the final purpose: cultural dissemination, innovative models for the Cultural Heritage fruition. For this reason we decided to invest mainly on the capitalization of the intangible heritage, in order to strengthen the unicity of this territory. The actual generation of people living in Matera that are about 50 -60 years old is the last one carrying on the direct experience of the life within caverns, they are depositary of the long and uninterrupted experience of rupestrian civilization. Today children and young people start to look at this patrimony in a different way as their life and uses have changed, they speak about Sassi as something belonging to their ancestor. In the next years Matera’s Sassi are going to become a “musealized” site, more and more open to tourism. We cannot know if in the future they will risk to loose the identity that actually can be still strongly perceived.

The main components of the platform are: 1) the web site, accessible in remote from home or school but also adapted for smartphone (Fig. 9), where it is possible to enjoy narratives, to make specific researches or to download different materials useful during the visit of the city; 2) cultural contents and applications for mobile devices (old style mobile phones, smartphones, iPad) adapted to different operative systems.

Every user can reach cultural contents in a simple way, he can choose the communicative format he prefers which will be supported by the technology he owns.

Another imposing aspect of the project has been the 3D reconstruction of Matera in different ages, realized by CNR ITABC, that obliged us to collect, study and interpret a huge amount of historical and iconographic sources.

The project has been concluded from a technical point of view and the applications are already available in the web. Now we are waiting for the opening of the documentation centre in the undergrounds of Piazza Vittorio Veneto (in charge of APT Basilicata, Matera Superintendency and CNR IRAT) that will represent a further opportunity to support and promote the project.

In conclusion we hope the project can produce social and cultural benefits and in particular:

- to improve the quality of the cultural and touristic services offered in the territory;
- to improve and enforce the presence of italian and foreign visitors in Matera and support them providing cultural information that can be accessed through a variety of devices and formats;
- to involve the people living in Matera in the processes of promotion and communication of their regional culture;
- to contribute to the creation of new development models in the domain of cultural heritage, in order to improve the citizens’ life quality;
- to determine preservation actions of the diffused, tangible and intangible heritage, as a consequence of the enhancement of the cultural and productive identity of the citizens.

A fundamental activity is the public’s monitoring (thanks to CNR IRAT) that has been concluded in these days. The elaboration of the final results is now in progress and unfortunately it is not available at the moment of this writing. We are going to refer about it in a next paper that will be presented together with the colleagues of CNR IRAT.

The future maintenance and management of the system will be responsibility of APT Basilicata

and Matera Municipality. We really hope they will be able to support, enforce, improve, update the platform and its contents also in the next years.

Acknowledgments

The project involves many territorial public institutions: the Basilicata Region, the Superintendency for Environmental and Architectonic Heritage of Basilicata Region, the Regional Direction of Ministry for Cultural Heritage (MiBAC), the Municipality of Matera, the Park Authority, the Provincial Administration of Matera. Also other CNR Institutes were involved: ISTC, IRAT, IBAM, ITD-Palermo. Some private companies collaborated for the creation and development of the contents: Virtual Heritage Italia Network, E.V.O,CA., Dinamo Italia, HSH, Netagency, Imagimotion. I would like to thank my colleagues Stefano Borghini, Raffaele Carlani, Nicola Montesano, Claudio Rufa, Augusto Palombini, Sofia Pescarin, Bartolomeo Trabassi, Marco Di Ioia, Sara Zanni, Belen Jimenez.

References

Antinucci, F. 2011. *Parola e immagine, storia di due tecnologie*. Roma: Laterza.

Antinucci, F. 2004. *Comunicare il Museo*. Roma: Laterza.

Antinucci, F. 2007. *Musei Virtuali*. Roma: Laterza.

Bres S., and Tellez. B. 2009. "Localization and augmented reality for mobile applications in cultural heritage." In *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. <http://www.isprs.org/proceedings/XXXVIII/5-W1/>.

Cameron F., and Kenderdine S. 2007. *Theorizing Digital Cultural Heritage: a critical discourse*. Cambridge: MIT

Press.

Demetrio, R. 2009. *Matera. Forma et Imago urbis*. Bari: Laterza.

Fonseca, C.D., R. Demetrio, and G. Guadagno. 1999. *Matera*. Bari: Laterza.

Fonseca, C.D. 1978. *Habitat – Strutture – Territorio*. Galatina.

Forte, M (eds). 2008. *La Villa di Livia, un percorso di archeologia virtuale*. Roma: Erma di Bretschneider.

Gattini, G. 1882. *Note storiche sulla città di Matera*. Napoli: Stab. Tip. Perotti.

Gibson, J.J. 1979. *The Ecological Approach to Visual Perception*. Hillsdale: Lawrence Erlbaum Associates.

Laureano, P. 1993. *Giardini di pietra: i Sassi di Matera e la civiltà mediterranea*. Torino: Bollati Beringhieri.

Takacs, G., V. Chandrasekhar, Y. Xiong, R. Grzeszczuk, W.C. Chen, K. Pulli, N. Gelfand, T. Bismpiagiannis, and B. Girod. 2008. "Outdoors Augmented Reality on Mobile Phone using Loxel-Based Visual Feature Organization." In *Proceedings of ACM international conference on Multimedia Information Retrieval*. New York: ACM.

Tropeano, M. 2003. "Il Parco Archeologico, Storico, Naturale delle Chiese Rupestri del Materano." *Geologi dell'Ambiente* Suppl. 1.

Volpe, F.P. 1818. *Memorie storiche profane e religiose sulla città di Matera*. Napoli: Stamperia Simoniaca.

Wagner, D., T. Pintaric, F. Ledermann, and D. Schmalstieg. 2005. "Towards Massively Multi-user Augmented Reality on Handheld Devices." *Pervasive Computing, Lecture Notes in Computer Science* 3468: 77-95.

Evaluating Virtual Museums: Archeovirtual Case Study

Sofia Pescarin
CNR ITABC, Italy

Alfonsina Pagano
University of Lugano, Switzerland

Mattias Wallergård
Lund University, Sweden

Wim Hupperetz and Christie Ray
University of Amsterdam, Netherlands

Abstract:

November 2011 saw the opening of the exhibition “Archeovirtual” organized by CNR ITABC and V-MusT Network of Excellence, in Paestum, Italy, under the general direction of BMTA (BMTA is the Mediterranean Expo on Archaeological Tourism directed by Ugo Picarelli and organised by Salerno Province). The event, part of a wider European project on virtual museums, was a rare opportunity to show many different projects about Virtual Reality and Cultural Heritage. During the show, four types of evaluation tools were employed to examine user behaviour and usability of the interfaces, plus understanding the gap between user expectation and experience. First analyses revealed that the impact of interactive applications on the user seems to depend on the capability of technology to be “invisible” and to allow a range of possibilities for accessing content. To achieve this, virtual museums need a more integrated approach between cultural contents, interfaces and social and behavioural studies.

Keywords:

Virtual Museum, Evaluation Methods, Social and Behavioural Interaction

1. Introduction

Evaluation activities are highly useful since they can improve and enhance the research domain. This is true also of the newly born field of Virtual Museums. Developing a good framework for analysis and evaluation can have a significant impact on the production phases of a virtual museum so that certain outcomes, such as knowledge exchange, cognitive improvement, and cultural heritage communication can be better achieved. Up to now, the field has produced insufficient extensive studies on the effectiveness of such virtual applications analysed on the basis of cross-comparative methods of evaluation; nothing really explicative has been done that has taken into account all of the complex variables that are inherent in a VM project. We still have to rely only upon visitors’ surveys and researches focused on web sites or user interface design analysis are provided by laboratories and

Corresponding author: sofia.pescarin@itabc.cnr.it

cultural institutions. Consequently, it is very difficult for us to build up a reliable and effective grid of indicators that would be helpful for analysing, studying and communicating with specialists and academics, thus, limiting the potential for a real improvement of the research. Therefore, our main goal is to study the criteria and parameters that could be used as reference for generally evaluating virtual museums, considering what kind of method, if one exists, we should adopt.

To move towards solving these issues, a European project focused on virtual museums, V-MusT.net (www.v-must.net), formed an entire work package dedicated to qualitative and quantitative evaluations through a range of interactive laboratory experiments. Although a previous attempt had been carried out in this field during the exhibition “Building Virtual Rome” 2005, in Rome [1], the results of that study did not reach enough detail, due to a lack of strategy



Figure 1. Entrance of Archeovirtual exhibition, Paestum, Italy.

and complexity required for evaluating and comparing different types of digital applications. A second attempt at VM evaluation was conducted in November 2011 during the exhibition of virtual archaeology, “Archeovirtual” (www.archeovirtual.it), held in Paestum, Italy, within the Mediterranean Expo of Archaeological Tourism (Fig.1).

Throughout this paper, we will present the framework in which V-Must acts, the most relevant methods of evaluation of VM applications, discuss their peculiarities and weaknesses, and present the strategy that we used to formulate our own approach to evaluation. The latter is the result of a synergy between laboratories, each with different experiences and methods.

2. V-Must.net Context and the VM Evaluation

A significant number of projects and applications used in virtual heritage are often born and die in digital labs, without being experimented or tested with real, final users. These “prototypes” – that is how they actually have to be called, since a “product” is not verified until it enters the market and starts to be used – are the result of experts, technicians, and academic knowledge, combined together to give birth to multidisciplinary and avant-garde outputs. Since most of the time these experiments concern the collective cultural heritage and aim at reaching the widest audience, it is necessary to consider how the target audience will experience the VM. For example, what if the “common” user could not operate them properly?

Could it be a problem of content understanding or a usability issue? Furthermore, how can we make them available and usable for all kinds of target groups? To begin answering these questions, the V-MusT, Virtual Museum Transnational Network (www.v-must.net), was funded as a European project and representing a Network of Excellence dedicated to Virtual Museums with the goal of making their next generation:

- More communicative in terms of availability of their connection to certain locations and for the type of information, tailored on the final user;
- More effective referring to the audience’s satisfaction and learning benefits;
- More sustainable in terms of reusability and exchangeability of content units on different medium;
- More accessible in terms of platforms (pc, tablet, iPad, mobile phone, site-specific projection, etc.) and user interfaces (desktop-based interaction, device-based interaction, natural interaction);
- More visible in public spaces and on line.

Here we have to make a specification: for “virtual museum” we mean the virtual reconstruction of environments and exhibitions, which contain cultural heritage objects either damaged or destroyed. From a theoretical viewpoint, however, the new term stands as a new model of communication that aims at creating a personalized, immersive, interactive way to enhance our understanding of the world around us. The term VM is an acronym that covers “various types of digital creations” that are available “online, in museums and on heritage sites”².

Thus the four-year long European program, including 18 partners from 13 different countries, aims to provide the heritage sector with the tools and support needed to develop virtual opportunities that are educational, enjoyable, long lasting and easy to maintain. To achieve the goals set forth in V-MusT.net, is fundamental to develop a grid of indicators, which we can use to guide evaluation or refer to when designing a virtual museum. Until now, few

² Definition available at: www.v-must.net/virtual-museums/what-virtual-museum.

- Stereoscopic 3D animations with device-based interaction - *Alejandria, el sueno de Alejandro*⁴; *Apa*⁵; *The Turnback time cave*⁶;
- Virtual reconstructions of buildings and artefacts with desktop-based interaction - *Spatial analysis of the Cathedral of Santiago de Compostela*⁷; *Plassac – La Villa du lle siècle*⁸, *AR-Etruschi*⁹; *Iconography of the Medical School of Salerno*¹⁰;
- Virtual reconstructions of archaeological sites with natural interaction - *Acquae Patavinae*¹¹; *Etruscanning 3D*¹²;
- Augmented reality with device-based interaction - *AR for the Presentation of CH Objects in Museums*¹³;
- Mobile solutions for cultural contents with device-based interaction - *The Lombards on Ipad*¹⁴; *Rome MVR*¹⁵.

During the exhibition, visitors could hear stories from Etruscan nobles, visit ancient Bologna with a character named Apa, experience the dream of Alexander the Great visiting Alexandria, explore the architectural intricacies of the Cathedral of Santiago de Compostela, and much more.

The approach to VM evaluation that was developed for Archeovirtual 2011 used four tools for evaluation: observations, visitor questionnaires, user interviews and VM developer interviews. These tools were selected to achieve our predetermined goals of evaluation, described earlier, and took into account the broad range of VM applications and formats included in the Archeovirtual exhibition.

4 Info at <http://elbranchito.es>

5 Info at <http://cineca.it>

6 Info at <http://virtualware.es>

7 Info at <http://videalab.udc.es>

8 Info at <http://plassac.gironde.fr>

9 Info at <http://noreal.it>

10 Info at mariapiadib@gmail.com; capone@unisa.it

11 Info at <http://aquaeptavinae.lettere.unipd.it>

12 Info at <http://regolinigalassi.wordpress.com>

13 Info at <http://fraunhofer.de>

14 Info at d.depadova@gmail.com; mara.pontisso@gmail.com

15 Info at <http://altair4.com>

We adopted a strategy that included a variety of evaluation tools to reflect the range of purposes we aimed to achieve with our evaluation, which included collecting both quantitative and qualitative results of visitor and VM developer responses.

To gather information on the user demographics and the length of user interaction with the VM applications, a non-interactive approach of *observation* was employed, yielding quantitative results about the audience. Observation was also used to identify technical issues and environmental requirements that could improve the presentation of the VM applications.

Visitor questionnaires also provided qualitative results about the users, including their familiarity with computers, their experience of the immersion and interaction, and the ease at which content is accessed in the VM applications.

In order to gauge the difference between the kinds of experience that VM developers intended and the actual user experience, we conducted *brief interviews* with both the VM developers and the users who experienced the VM applications at Archeovirtual. This allowed us to compare the responses of the developers and users to identify whether or not the expectations of the VM developers we met by the experiences of the users.

4. Preliminary Results

During the three days evaluation conducted at Archeovirtual, we were able to collect 99 observation reports, 215 written questionnaires, 49 user interviews and 25 VM developer interviews. While these numbers are substantial and produced a significant amount of data to analyse, they fell short of the results we were hoping to achieve for each of 14 VM applications. It was our hope to collect 10 observation reports, 10 questionnaires and 3 interviews for each VM application for each day of the exhibition. Unfortunately we were not able to reach our target participation numbers due to a limited number of attendees, the insufficient number of questionnaires available, and the need for evaluation supervisors to be in the field, both to conduct the surveys, and protect and control the VM presentation spaces.

Archeovirtual results helped first to define the visitor's profile by comparing and overlapping the data obtained by all the type of surveys. The *average user* was a female (199 women and 163 men inquired), adult, between 20 and 35 years old, mainly academic or working as archaeologist and engineer. She was European, mainly coming from the surroundings of where the exhibition was held. She visited the show with school, alone or with his family, spending minimum 10 minutes in each projects' area. She had some knowledge about Archaeology and Cultural Heritage and used a lot the computer not only for professional jobs but also in the free time.

For what concerns observations, outcomes pertained the *visitors' behaviour and attitudes*. About the overall involvement, visitors seemed really participative since they comments at the applications (35% much, 22% a bit, 33% no) and interacted a lot with them (34% much, 31% a bit, 25% no). They seemed to understand what they are experiencing, as they did not ask for more information. About the projects themselves, visitors did not seem to have problems accessing the contents (65% of them declared to do not having problem of accessibility) so demonstrating quite good usable systems without particular technical concerns (only 12% of them declared to have encountered technical problems; 19% to have had usability constraints).

Regarding visitor questionnaires, what came out about *users' expectations and experience* was that content of applications appeared innovative, well structured and captivating. These comments were directly linked to the positive aesthetical experience had by visitors who considered the projects for their pleasant interfaces, inviting and creative in the design and motivating by the environment. Nevertheless, contrasting results arose about the usability of applications: practical, simple and undemanding projects allowed to gain medium/high sense of immersion but a with low "perception" of interactivity. This is a strange result also because of the feedback coming from the observations: as said above, that users interacted a lot with the applications without encountering so much problems of usability and technical issues! Probably that was due to a misunderstanding on the meaning of "usability" by both observers and visitors.

Referring to brief interviews, what was pointed out was that the most appreciated features in applications were: realism, contents, storytelling, 3D reconstructions and interactivity. Users also affirmed that almost all of the applications were easy to use; they positively met the starting expectations; in definitive, they were considered a powerful way to enhance the learning of cultural contexts.

Comparing then experts and developers on the involvement evocated by applications, some contrasting results concerned both the sense of immersion, considered high by users and medium-low by experts/creators and the interactivity of the systems, considered low by users and medium-high by experts/creators.

5. Comparison of Evaluation Approaches: Etruscanning 3D

The Etruscanning 3D project, a European project (Culture 2007) aimed at improving collaboration in VM development, produced a VM application, which virtually reconstructed the Regolini Galassi tomb and allowed users to experience the tomb through the use of a natural interaction interface. The Virtual Reconstruction of the Regolini Galassi Tomb, as it is called, has been presented in association with a joint Etruscan exhibition presented in the Netherlands at both the Allard Pierson Museum, in Amsterdam, and at the Rijksmuseum van Oudheden, in Leiden, but it was also presented at the Archeovirtual 2011 exhibition in Paestum, Italy. During all presentation periods and in each location, the Virtual Reconstruction of the Regolini Galassi Tomb was evaluated using a variety of evaluation approaches to achieve a range of purposes, as presented in Table 1.

During the evaluation of the application at the Allard Pierson Museum (APM) and the Rijksmuseum van Oudheden (RMO), two primary approaches were used in parallel: observation and visitor interviews. Separately, at the APM, we also ran a short A-B Study on the integration of the application into the wider Etruscan exhibition, and at the RMO a group of 21 archaeology students from the University of Leiden completed a questionnaire about their experience and opinions of the application. At Archeovirtual, we also used observation and visitor interviews, as well as a short written survey that was completed by

both the visitors to the exhibition and the developers of the VM application.

The approaches to observation and user interviews conducted at the APM and RMO were done using a consistent, pre-set methodology, which employed the use of template observation forms and open interview questions. The purpose of the observation approach was to observe the user demographics, the environment into which the application was presented, including visitor flow, lighting and safety, and to determine the technical functionality and reliability of the application. Observation was conducted 16 times between the APM and RMO, using the observation templates and ensuring there was no interaction by the observer with the application or its users. The user interviews were conducted 10 times between the APM and RMO using a set of pre-defined questions that provided qualitative data about the user experience of interaction and immersion in the application, as well as the museum context into which the application was presented and some user demographic information. The results of the observation and interview approaches were returned to the VM developers throughout the evaluation period so that new versions of the application could be created to remove any observed technical issues and improve the interaction technique.

An A-B Study was also conducted at the APM using two groups of participants, 8 in each, to examine the integration of the application into the wider Etruscan exhibition, but also to evaluate the content, the use of interaction and immersion, the added value of including the application with the exhibition. Participants of this study were divided into two groups; Group A would experience the application before the exhibition, while Group B would visit the exhibition before experiencing the application. Throughout the study, participants would fill in a series of questionnaires, which would later be analysed and compared to determine the merits of embedding the VM application into the exhibition.

In addition to the other evaluation approaches used at the APM and RMO, a class of 21 archaeology students from the University of Leiden also completed questionnaires based on the user interview template, but tailored for an expert

audience. This approach was used to collect data from an audience which already has knowledge of the Etruscan civilization and the traditional archaeological method used for presenting reconstructions, but the student responses also extended into their personal experiences with the interaction and immersion used in the application, as well as how the VM application was integrated with their experience of the exhibition and the physical museum.

While the approaches for VM evaluation at the Archeovirtual were similar to those applied at the APM and RMO, the methodology and the purpose of the approaches used were different and yet complementary to the techniques used and results obtained at the APM and RMO. As with the approaches used at the APM and RMO, the purpose of observation at Archeovirtual was also intended to observe the visitor demographics, to monitor the environment, including visitor flow, lighting and safety, and to assess the technical functionality and reliability of the application. The results of observation at Archeovirtual and at the APM and RMO clearly identified the impact that the physical environment had on the quality of the VM presentation, which may not have been fully understood if we had only conducted observation in one location. These results were confirmed by the application's users in their responses to questions about their experience in both the questionnaires and the user interviews from all evaluation locations.

The product of those cumulative approaches for the evaluation of the Virtual Reconstruction of the Regolini Galassi Tomb are a blend of both quantitative and qualitative results. While observation in all evaluation locations and the written surveys conducted at Archeovirtual provided quantitative results for visitor demographics and the technical reliability of the application, the questionnaires and interviews used at the APM and RMO, both in the A-B Study and with the University of Leiden archaeology students, provided a significant amount of qualitative data, offering greater insight into the opinions and attitudes of the application users. Naturally, using such a range of evaluation approaches requires an equally wide range of approaches to data analysis, which can be challenging and time consuming, however, clearly defining the goal or purpose of the evaluation

ARCHEOVIRTUAL	APM / RMO
<i>Dates of Evaluation</i>	
17-20 Nov. 2011	16 Oct. 2011 - 15 Mar. 2012
<i>Methods Used</i>	
Observation, Interview, Written Survey	Observation, Interviews, Questionnaires
<i>Results</i>	
Observation - 13	Observation - 16 (15APM, 1RMO)
Interview - 7	Interviews - 10 (4APM, 6RMO)
Written Survey - 34	A-B Study - 16 (all APM, 2 groups - 8 Group A, 8 Group B)
	Questionnaires - 21 (ULeiden Archaeology Students)
<i>Purpose</i>	
Observation - User demographics, Environment, Technical functionality	Observation - User demographics, Environment, Technical functionality
Interview - User demographics, Interaction and immersion	Interview - User demographics, Interaction and immersion, Museum context
Written Survey - User demographics, Usability, Content, Interactivity, Experience, Expectation	A-B Study & Questionnaires - User demographics, Content, Interaction and Immersion, Museum/Exhibition context, Added value of VM to museum/exhibition

Table 1. Comparison of evaluation methods used at Archeovirtual and APM/RMO.

and selecting the most appropriate approach to evaluation that will yield the type of results desired (quantitative or qualitative) can reduce the amount of evaluation and data analysis required. Although some of the results of the methods used to evaluate the Virtual Reconstruction of the Regolini Galassi Tomb may have overlapped, the differences in the range of approaches applied also amounted to a wider range of user responses, giving a more complete understanding of the impact of the application than just one evaluation method would have been able to provide.

6. Assessment of Evaluation Methods

Aside from both timing and organizational constraints, several interesting and useful considerations emerged from the Archeovirtual evaluation experience. Primarily we can affirm that among the collected data, the results were occasionally missing the convergence of pairing results, which means that despite having the same questions asked, using the same investigative tools, we experienced different answers. It could be either a problem of data collection, or a methodological issue in conducting the interviews, or perhaps it is simply a natural, yet unpredictable outcome. One of the reasons for this might be that the visitors at the exhibition were quite heterogeneous, ranging in age and experience. Furthermore, contrasting results appeared when we investigated usability issues, which may be a problem of meaning (visitors do not fully get or misinterpret), of incapacity of interviewees (to explain what usability is about) or of a misunderstanding during the observation sessions. Some results were also not discernible,

possibly due to the unpredictable development of the evaluation experience. Regarding the four evaluation tools we applied, we can comment that:

Observation sessions, in their completeness, are difficult activities to conduct, since it is not always clear whether what the observer highlights and recognizes corresponds to the real intentions and attitudes of participants. The majority of the time, the observer is detached from the situation, relying upon his own perceptions, which may be inaccurate or influenced by personal experience, a lack of strategy in facing the activity, or by distractors, which may separate him from the established starting observation goals. The observation method inevitably involves human factors such as the psychological state of the participant, his emotional reactions at certain external inputs, and individual abilities and preferences developed by the participant. What has to be first highlighted in the Archeovirtual case is that, despite the observers were facilitated in addressing observations, thanks to a predefined protocol, they may have missed or misinterpreted something. That could be due to the fact that they did not have a strong academic background in human psychology or study of collective behaviours. The attitudes and behaviours of visitors could also depend on their individual characters and psychological qualities, therefore, not knowing the participants in advance makes the reading their behaviours and personal features much more complicated.

Visitor questionnaires generally allow participants to speak about themselves and directly express their own opinions and points of view. In the Archeovirtual case, these surveys saw problems of interpretation: visitors reported the difficulty in fully understanding some questions or words of the provided protocol, like “immersion”, “involvement”, “usability” and “interaction”. Nevertheless, the staff gave them the further details requested in order to let them answering correctly. Moreover, some participants refrained from answering certain questions because of their incapability to express an opinion or because they attended only part of the exhibition.

User interviews are always double-face tools, since generating valid and reliable data is definitively non-trivial. That is due to the fact that

the participant's response is subjective, context-dependent and dynamic over time. Due to a lack of time to get into depth with each interview and reach the desired result numbers, not all Archeovirtual visitors answered all of the questions included within the protocol, which was influenced by the special attention that each user gave to some questions over others, dwelling on certain topics. In addition, the staff was not always about to follow the consequentiality of the protocol, since they preferred to go along the logic of the speech dictated by the interviewee. From a content point of view, sometimes suggestions influenced the responses of users, either formulating various hypotheses while answering or anticipating issues; thus, participants were guiding the interview through their responses. We also noticed that younger visitors were full of enthusiasm and open to new learning experiences, contemplating the use of new virtual technologies in supporting Cultural Heritage, while those visitors who were a bit older, and perhaps more experienced, revealed an opposite attitude.

VM developer interviews, most of the time, helped us in understanding the starting goals of the VM application, the project's design and the crucial steps, which would bring the "product" away from the laboratory to meet with the final users. In the Archeovirtual case, the developer's aims did not always fit with user experience, such as grade of sensory immersion in the application, type of content interaction, quality of user interfaces, and so on. These differences may be a result of the communicative level of the project set up at the beginning of its design. Some applications that are born in digital labs, where merely experts in the field and specialized technicians worked on them, contribute to the creation of complex and stratified content/interface/technology.

7. Conclusions

In conclusion of our collective efforts, we cannot report at this point final statements in the field of interaction/relations among virtual environments, virtual content, museums and final users. Alternatively, it is possible to provide some suggestions to start creating a grid of indicators useful when designing and producing a virtual museum - in its multiple forms.

From a *methodological point of view*, the combination of several evaluation tools to collect both quantitative and qualitative data is proved to be worthy. This will give investigators the opportunity to study the effects of virtual environments and exhibitions towards the public in depth, including the possibility to target specific user groups and develop ad-hoc strategies for evaluating and analysing their social behaviours. The methods selected for evaluating VM would ideally be dictated by the desired kind of results or evaluation outcomes that are to be achieved. With this in mind, it is important to evaluate VM using a range of techniques in order to get the most complete understanding of the impact of the VM.

From a *content perspective*, instead, the kind of information a user expects from a virtual museum experience relates to the application development process, the storytelling or narrative which ties together the application, the accuracy of the virtual and 3D reconstructions, and the possibility to interactively query different content using the system. Another important issue to take into account when projecting a virtual museum and its final impact on the users, it is to carefully consider the use of interactive or non-interactive interfaces. The former seem positively affecting the feedback of users if the technology employed remains "invisible" to them, allowing them to be engaged with the system but, at the same time, not being intrusive, overriding and complicated. Moreover, the technology employed has to allow for a wide range of possibilities in content access. Furthermore, it has been proven that a sort of "guide," mainly similar to a physical guide or a multimedia figure can enhance the visitor experience. In the first case, the aid of a real person - who knows the application, how it works and what its story is about - guarantees the effectiveness of the user's satisfaction while he is "playing."

In the end, to provide more objective criteria to create a complete grid of indicators for VM evaluation, further studies should be undertaken and focused towards advanced digital technologies affecting human behaviour; the general "schema" of interconnectedness between computing inputs and human feedback is not completely clear yet. More evaluation and statistics are indeed necessary and should be required by project leaders and decision-makers at the beginning of the work.

Acknowledgments

We would like to thank staff and colleagues who contributed to this work: all V-MUST partners, all Archeovirtual staff - technicians, sponsors, students and trainees; Francesco Vallecoccia, trainee at VHLab, CNR ITABC (University of Lugano) who has helped to process data. Specifically Etruscanning 3D: Visual Dimension (D. Pletinckx), CNR ITABC (E. Pietroni), Softlogic (C. Rufa). For Archeovirtual: U. Picarelli (BMTA director), Leader SAS, Salerno Province.

References

- Forte, M., S. Pescarin, and L. Pujol Tost. 2006. "VR applications, new devices and museums: visitors' feedback and learning. A preliminary report." In *The 7th International Symposium on Virtual Reality, Archaeology and Cultural Heritage - VAST*, edited by M. Ioannides, D. Arnold, F. Niccolucci and K. Mania, pages, 64-69, Crete: Eurographics.
- Harper, B. D., K. L. Norman. 1993. "Improving User Satisfaction: The Questionnaire for User Interaction Satisfaction Version 5.5." In *Proceedings of the 1st Annual Mid-Atlantic Human Factors Conference, 1993*, 224-228. Virginia Beach.
- Hassenzahl, M. 2006. "Hedonic, emotional and experiential perspectives on product quality." In *Encyclopedia of Human Computer Interaction*, edited by C. Ghaoui, 266-272. Hershey, USA: IGI Global.
- Hassenzahl, M., M. Burmester, F. Koller. 2003. "AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. [AttrakDiff: A questionnaire for the measurement of perceived hedonic and pragmatic quality]." In *Mensch and Computer 2003: Interaktion in Bewegung*, edited by J. Ziegler, J. and G. Szwillus, 187-196. Stuttgart, Leipzig: B.G.Teubner.
- ISO FDIS 9241-110:2006. *Ergonomics of human system interaction - Part 110: Dialogue principles*. Switzerland: International Organization for Standardization (ISO).
- Jordan, P. 2002. *How to Make Brilliant Stuff That People Love ... and Make Big Money Out of It*. New York: John Wiley and Sons Ltd.
- Kirakowski, J., and M. Corbet. 1993. "SUMI: the Software Usability Measurement Inventory." *British Journal of Educational Technology* 24 (3): 210-212.
- Law, E., V. Roto, M. Hassenzahl, A. Vermeeren, and J. Kort. 2009. "Understanding, Scoping and Defining User Experience: A Survey Approach." In *Proceedings of Human Factors in Computing Systems conference, 719-728*. New York: ACM.
- Nielsen, J. 1994. "Heuristic evaluation." In *Usability Inspection Methods*, edited by J. Nielsen and R. I. Mack, New York: John Wiley & Sons.
- Rubin, J., D. Chisnell. 2008. *Handbook of usability testing*. New York: Wiley Publishing.

Extracting Scar and Ridge Features from 3D-scanned Lithic Artifacts

Eitan Richardson, Leore Grosman, Uzy Smilansky and Michael Werman

The Hebrew University of Jerusalem, Israel

Abstract:

We propose a method for computer analysis of 3D-scanned lithic artifacts. It provides a detailed description of the scars and ridges that were left on the surface of the stone tool while it was produced by knapping. Traditional documentation and research methods are based on hand drawn images of stone tools and manual measurements. Our algorithm enables automatic, objective and precise documentation of the tool surface, together with a quantitative analysis of the scar and the ridge networks. The proposed algorithm detects ridges on the surface of the scanned object, segments the surface to scars and calculates features that can be used for clustering and classifying lithic artifacts. We demonstrate our method on various lithic artifacts, present the extracted features and compare the results to traditional hand drawings.

Keywords:

Scars and Ridges, Lithic Artifacts, Segmentation, 3D Scanning

1. Introduction

Much of what is known today about prehistoric periods derives from the study of stone artifacts manufactured during the Paleolithic period that spans over 99.5% of human history. The gradual evolution of stone technology from rough and simple forms to highly sophisticated and refined objects marks the cognitive evolution of the human brain, as well as the development of manual, technical and social skills. Prehistoric stone tools were manufactured by knapping flakes from a stone core and creating a scarred surface. The surface that is created after removing a flake is called a *scar* and the boundary of the scar a *ridge*. As more flakes are removed, ridges become borders between adjacent scars on the lithic artifact (Fig. 1).

The most important and difficult task of the archaeologist is to decipher the mute messages embedded in the shapes and forms of the lithic artifacts and thus learn about different aspects of the lithic industry. Of prime importance are questions related to the production technology of the artifacts. The very nature of the irreversible process of knapping implies that the producer of a stone artifact had a clear idea not only of the final product he/she had aimed at, but also of the chain of operations which should be followed to achieve the desired goal. Hence, the mental process is engraved in the morphology of artifacts retrieved

Corresponding author: eitan.richardson@gmail.com

from prehistoric sites, which include the waste and the final products of knapping - the tools. This is the reason why a major effort in prehistoric research is devoted to the study of knapped stone morphology, aiming to deduce from the observed features the chain of actions and considerations involved in the production (Bar-Yosef and Van Peer 2009; Boeda et al. 1990; Sellet 1993). The most prominent morphological feature of the stone tool surface is the distribution and hierarchy of scars and ridges which decorate it. They provide the only direct evidence for the succession of flake removals. This is why the study of the apparent scars and the ridges that form their boundaries was always considered of supreme importance in prehistory. Scars and ridges provide direct clues for the various phases in handling the raw material of the knapper and also hint to his/her technological skill.

In the present work we utilize a high precision optical 3D scanner that bridges between archaeology and computer graphics/vision methods enabling computer-based analysis of the morphology of prehistoric stone tools. This work relies on previous work that was aimed at positioning and extracting measurements of lithic artifacts using 3D-scanning models (Grosman et al. 2008, Grosman et al. 2011). We present an algorithm for identifying the ridge and scar pattern on the surface of the object. It enables the introduction of quantitative and intrinsic measures such as the number of scars, their surface area, depth and mean curvature.

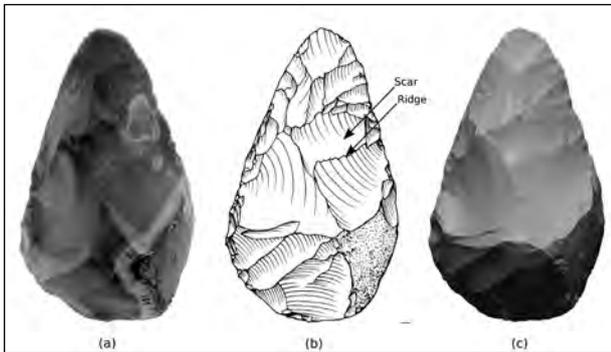


Figure 1. Lower Paleolithic Handaxe from the site NBA (after Grosman et al. 2011). (a) A photograph of the stone tool (b) A traditional hand drawing with indication of a scar and a ridge (c) Rendering of the 3D-scanned tool.

The presentation is arranged along the following lines: Section 2 of the paper discusses prior work, both in the field of prehistoric archaeology and in computer vision. In section 3, we provide a detailed description of the scar segmentation algorithm. The extracted features are described in section 4 and some results are presented in section 5, including comparison to hand drawings and to other mesh segmentation methods. In the last section, we discuss the results and possible future work.

2. Related Prior Work

Most of the archaeological research which focused on scars and ridge patterns has been rather intuitive and descriptive. Several attempts were made to explain prehistoric assemblage variability by trying to quantify the extent, nature and variability of the reduction chain (e.g. Davis & Shea 1998; Eren et al. 2005; Eren & Sampson 2009; Kuhn 1990).

Attempts to quantify flake scar patterning were based primarily on two dimensional projections (e.g. Conrad et al. 2004), drawing the scars manually and subjectively, thus failing to capture the complexity of the 3D nature of stone knapping. Recently Clarkson et al. (2006) suggested a 3D analysis of flake scars using a 3D measurement tool. The 3D data was recorded manually and the scars were described in terms of directed lines (vectors) which connect the extreme points of the scar – a rather incomplete description which fails to convey the complexity and detail of the network of scars and their characteristics.

Scar detection, from the computer graphics/vision point of view, belongs to the intensively studied field of image segmentation. In recent years, with the increased availability of three-dimensional data, there has been much research in the area of 3D surface segmentation. One of the challenges in surface (or “mesh”) segmentation is to define the criteria for evaluating the quality of the resulting segmentation. In most cases there is no objective principle which establishes a correct or preferred segmentation. In (Shamir 2008), the author defines two types of mesh segmentation methods, which aim to achieve two different goals. *Part-type* segmentation methods aim to segment the 3D object into meaningful parts, for example, segmenting a 3D human figure into head, arms, legs etc. The goal in *surface-type* segmentation is to segment the surface of the object into patches according to some criteria. The problem of segmenting the scanned lithic artifact belongs clearly to the latter type. In addition, the underlying knapping process, which ideally creates well-defined surface scars, should provide proper criteria for the success of a given segmentation: it should be consistent with the expert archaeological evaluation.

Surface-type mesh segmentation algorithms are based on various surface attributes. Some examples are surface curvature and geodesic distances, which we discuss in more detail in section 3. Other possible attributes are the normal directions of the triangular faces, the dihedral angle between adjacent faces and the planarity and convexity of the segments. A popular family of segmentation methods is Watershed (Mangan and Whitaker 1998), a region-growing method that starts with local-minima points of some attribute function of the mesh and grows regions until local-maxima points are reached. In section 5, we compare the results of our method to those of Watershed.

Methods for detecting ridges and valleys on a surface embedded in 3D draw on techniques developed for the related problem of edge detection in images. In our work, we use an existing ridge-detection method (Yoshizawa et al. 2005). Applying a ridge detection method to the surface of a scanned stone tool does not solve the scar segmentation problem by itself, the method does not detect all the scar-ridges, and on the other hand, it detects many false ridges. The reason for this failure is that

ridges on stone tools have a specific archaeological definition – they are the intersection of two scars. In our method, we utilize ridge detection as an intermediate step towards the scar segmentation as described in section 3.5.

3. Detecting Scars and Ridges

3.1 Formal definition of the segmentation problem

The main goal of our algorithm is segmenting the surface of the scanned tool into scars. The stone tool is scanned using a structured-light 3D scanner, generating a triangular mesh (Fig. 2), which is a surface in three-dimensions defined by small triangles – a set V of n vertices $v_1 \dots v_n$ in R^3 and a set F of m triangular faces $f_1 \dots f_m$, each face defined by an ordered set of three vertices. The goal of the segmentation task is to divide the set F of m faces into k disjoint sets $F_1 \dots F_k$, such that each F_i corresponds to a single scar on the surface of the object. It should be noted that the number of segments k is not an input to our algorithm – it is detected automatically. Once the mesh has been segmented into scars, the ridges can be defined as the borders between scars (the polygons connecting the ordered set of vertices that form the border of each segment). For rendering purposes, these polygons can be smoothed to form a smooth ridge curves.

3.2 Outline of the algorithm

Our method is based on the importance of high-curvature ridges for identifying borders between scars and on the other hand on the need for a robust method that will ensure the spatial consistency of the scars.

The surface curvature value at each point is first calculated. This value measures how strongly the surface bends at each point. Curvature ridges are then detected. These are continuous lines on the surface which follow local maximum points of the curvature value. The curvature-ridges mostly coincide with the actual scar-ridges. The curvature values and ridges are then utilized in a clustering process which breaks the surface of the object into small sub-scars. The clustering process is done in a way that ensures that a single cluster will not cross a high-curvature area or a ridge. In the next step, the border between each adjacent pair of clusters is

analyzed, clusters with a weak border between them are merged together. At the end of this process, the small sub-scars have been merged to become complete scars. A final optimization step is then performed to fine-tune the segmentation. At the end of the optimization step, the final scars are defined and the border between each adjacent pair of scars is defined as a ridge.

The list below outlines the main steps in our algorithm. Each step is further detailed in the sections below. Figure 3 shows the intermediate results.

1. Preprocess the mesh (3.3)
2. Estimate the maximum principal curvature at all vertices (3.4)
3. Detect curvature-ridges (local-maxima lines of the principal curvature) (3.5)
4. Perform initial geodesic clustering into sub-scars, using random seeds (3.6)
5. Merge adjacent clusters that have a weak border (3.7)
6. Fine tune the segments using graph-cut optimization (3.8)

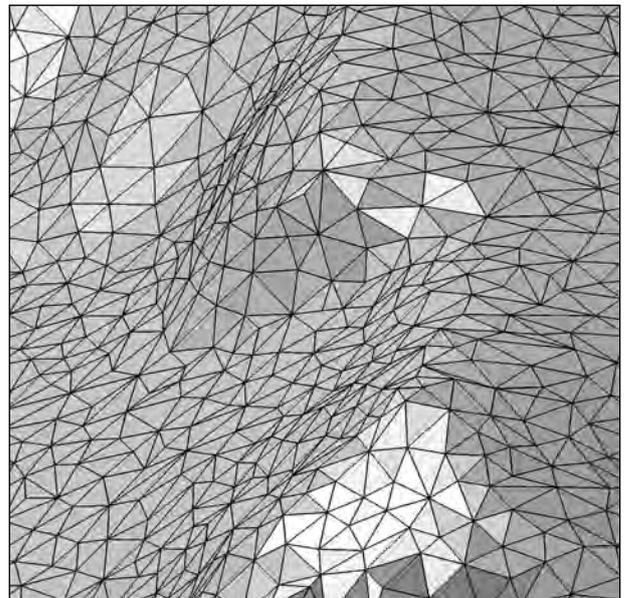


Figure 2. A close-up look on a scanned object, represented as a three-dimensional triangular mesh.

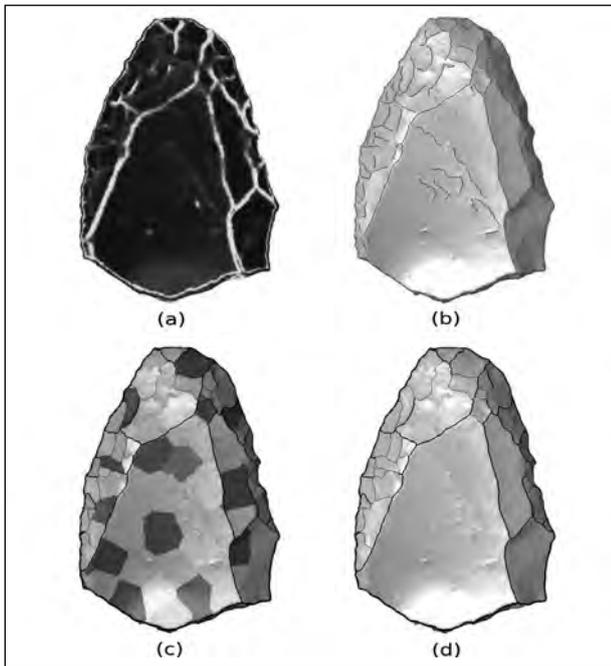


Figure 3. Intermediate steps results. (a) Estimated maximum principal curvature (b) Detected ridge lines (c) Geodesic clustering (segments indicated by colors) and post-merging (indicated by black borders). (d) The final scar-ridges after graph-cut optimization.

7. Define the ridges as the borders of the segments (3.9)

3.3 Mesh preprocessing

Artifacts can be scanned at different levels of detail, or resolution. As a pre-processing step, we apply a mesh simplification technique (Siek et al. 2002) to bring the resolution of all processed tools to a common value. In addition, we use Principle Component Analysis (Jackson 1991, Jolliffe 2002) to position the object in 3D so that the main axis of the tools is parallel to the x axis etc. This step is required for the proper calculation of the object's dimensions.

3.4 Principal curvature estimation

The principal curvature values and directions at each point on a smooth surface embedded in a three-dimensional space are mathematically defined as the two eigenvalues (k_1 , k_2) and corresponding eigenvectors of the shape tensor at that point. The principal curvature on a triangular mesh is estimated

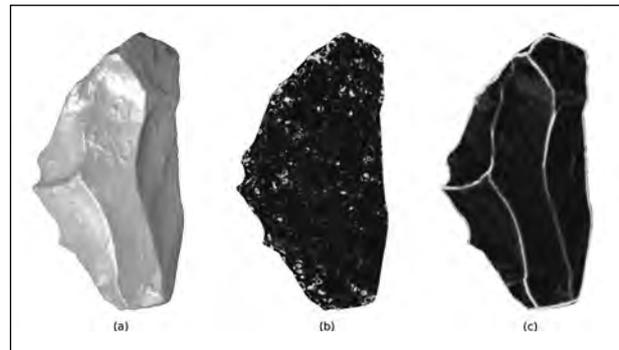


Figure 4. (a) The original scanned object. (b) The minimum principal curvature value - k_2 (not informative for detecting the ridges.) (c) The maximum principal curvature value - k_1 .

using (Cohen Steiner and Morvan 2003). As can be seen in Figure 4(c), the maximum curvature value (k_1) is high along the ridges.

3.5 Curvature-ridges detection

We apply a ridge-detection method (Yoshizawa et al. 2005) to detect curvature-ridges on the surface of the object. Curvature-ridges are defined as local-maximum points of the maximum principal curvature value (k_1) in the principal curvature direction (Koenderink 1990). We use the term “curvature ridges” to differentiate these ridges from the stone artifact ridges or “scar ridges”, which are the goal of our method. The differences are:

Curvature ridge detection identifies many minor ridges, which are caused by imperfections in the shape of the object. Most of these ridges are irrelevant to our work and are considered false detection.

In some cases, parts of the boundary of a scar will not be detected as a curvature-ridge. Our goal is to detect the entire boundary of each scar.

Note that (Yoshizawa et al. 2005) also detects valley lines on the mesh, which are points of negative curvature. Our method does not use the valleys. Figure 5 shows the detected ridge lines.

3.6 Geodesic clustering

As a first step towards segmenting the scars, we perform initial clustering of the mesh faces.

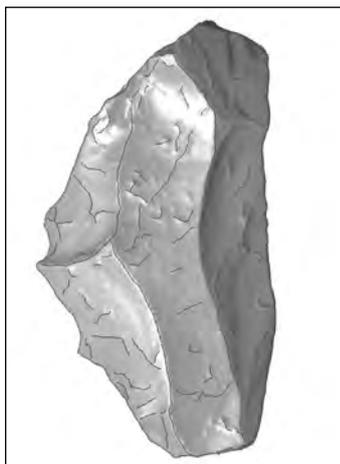


Figure 5. Detected curvature ridges.

If we consider the original mesh as a clustering in which each face forms a cluster, our goal is to reduce the number of clusters from an order of 100K to several hundred. In addition, we require clusters to be connected component and we want the borders between adjacent clusters to be consistent with the curvature-ridges.

A clustering process assigns to one cluster elements that are close to each other according to some distance function. Figure 6 demonstrates three vertex distance functions. The Euclidean distance causes points on opposite sides of the mesh to be close and therefore is not suitable for mesh clustering. The geodesic distance measures distances on the surface of the mesh. We use the Fast Marching method (Sethian 1999) to estimate the geodesic distance between a starting point to all other points on the mesh. We use non-uniform edge weights, based on the maximum curvature value at

each vertex, thus increasing the geodesic distance between a pair of points on the two sides of a high-curvature ridge, as can be seen on the right column of Figure 6. This encourages low-curvature areas to be clustered together and the border between clusters to coincide with the curvature-ridges.

Our clustering method proceeds along the following line: For each vertex we maintain three properties - a boolean flag indicating whether it is a part of a cluster, the assigned cluster (label) and the geodesic distance value to the segment center (or seed) point. We start with all vertices being unoccupied and having an “infinity” distance from a seed. We choose a random vertex (random seed) on the mesh and calculate its weighted geodesic distance to all other vertices. We then mark the geodesic neighborhood of the seed (all vertices whose geodesic distance from the seed is smaller than a predefined threshold) as occupied. We assign to the new seed all vertices that are geodesically closer to it than to their current seed. We then choose the next random seed that is not occupied and does not lie on a high-curvature ridge. The process continues until there are no available vertices.

Note that although the above method clusters the set of vertices and not the faces, it is easy to define a clustering of the faces based on the vertices clustering. We use majority voting of the clusters of the three vertices. In rare cases, the three vertices of a triangular face will belong to three different clusters. In this case, we choose the cluster of the first vertex.

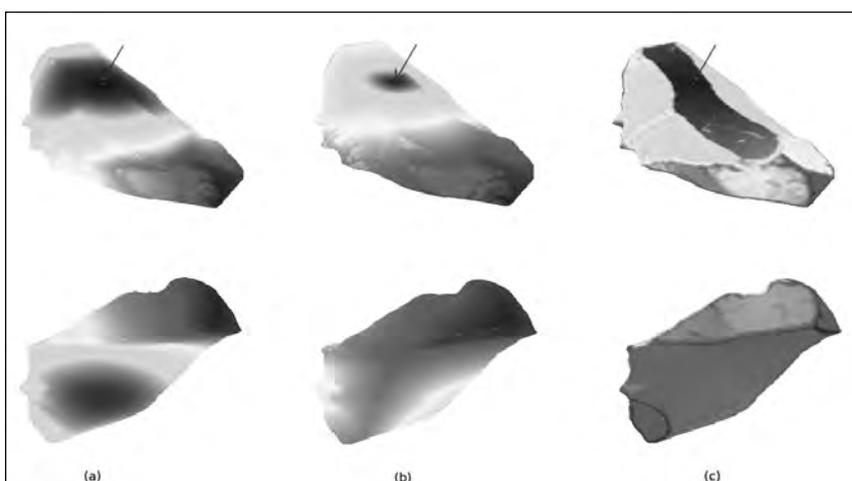


Figure 6. Distance functions. Each column shows the distance between the point marked by a red arrow in the top image to all other points on the object, according to a distance function. The distances are indicated using blue-to-red color coding. The objects on the bottom row are rotated to show the other side. (a) Euclidean distance. Note that points on the other side of the object are close to the marked point. (b) Geodesic distance. (c) Geodesic distance with a weight function based on the maximum curvature.

3.7 Post merging

The geodesic clustering phase generates sub-scar clusters. These clusters (or segments) represent over-segmentation of the surface of the mesh, however, the segments do not cross the scar-ridges. The border between a pair of adjacent segments contains points that have equal geodesic distances to the two seeds. These borders fall into two categories - borders that strongly depend on the randomly chosen seeds of the clusters, and stable borders that do not depend on the exact location of the seeds. On flat areas (or areas with negative curvature), the points in which the geodesic distances from the two seeds are equal strongly depend on the selected random seed. However, if the two seeds fall on the two sides of a high-curvature ridge, the dependency will be much lower. Randomized Cuts (Golovinskiy and Funkhouser 2008) is a method that utilizes this feature to identify meaningful borders, by running the clustering algorithm multiple times and keeping only the repeating (or stable) borders. We found that in our problem, it is sufficient to identify the meaningful borders by calculating the median maximum-curvature value along the border and merging adjacent segments with a low-curvature border. Figure 7 shows the results of the geodesic clustering and post-merging.

3.8 Graph-Cut optimization

Graph-cut is a familiar method for image and for mesh segmentation (Boykov and Jolly, 2001). We use graph cut as a final post-processing segmentation phase. Because we want to improve the quality of the border between the segments in addition to improving the segmentation, we use graph-cut at the individual triangular faces level. The graph-cut algorithm solves the min-cut optimization problem of partitioning a graph into two disjoint groups of vertices while cutting edges with a minimal total weight. In our case, we use a variant of graph-cut (alpha-expansion), which allows partitioning the graph into a larger number of segments. The input to graph-cut is the weighted face-adjacency graph. Each face forms a node in the graph and an edge connects each adjacent pair of faces (creating a 3-regular graph). We now add a node for each segment label and an edge between each face-node to its initially-assigned segment. The weights of the edges connecting face-nodes are the

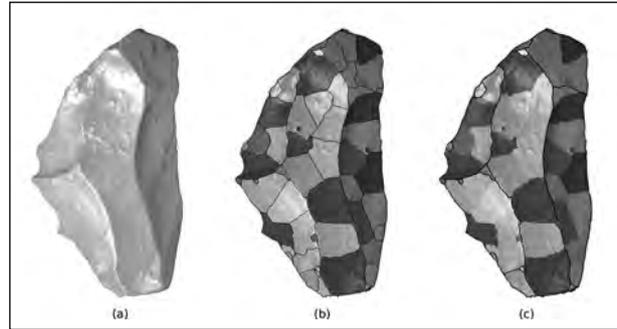


Figure 7. Geodesic clustering and post-merging. (a) The original object. (b) The geodesic clustering result. (c) Post-merging: The colors indicate the original geodesic clustering and the black borders indicate the result of the post-merging (random borders removed).

smoothness terms. They define the cost of cutting the mesh into two segments at each edge. The weights between the face-nodes to the segment-nodes are the *data term* and represent the cost of changing the original assignment of a face to a segment. In our method, we base the smoothness term on the maximum-curvature and on the detected ridges. An edge that passes through a detected ridge line will have zero smoothness term (thus encouraging graph-cut to segment the mesh at these edges). Other edges will have a weight that is proportional to $1/\max\text{-curvature}$, so that the cost of partitioning at a flat area is high.

3.9 Defining the ridges

The final scar ridges are the borders between scars. We define the border between each adjacent pair of scars as a ridge line. We assign a weight to each ridge line, based on the median of the maximum-curvature along the ridge line. This weight can be used to visually emphasize major scars by drawing them with a thicker line. In addition, we apply a smoothing function to the ridge lines. Figure 8 shows the final ridges, which are the result of the graph-cut optimization and the ridge definition phases.

4. Feature Extraction

As presented in section 1, one of the goals of our work is to calculate features that can be used for automatic classification and clustering of scanned artifacts. Clustering and classification algorithms require features that are descriptive and discriminative (can differentiate between classes

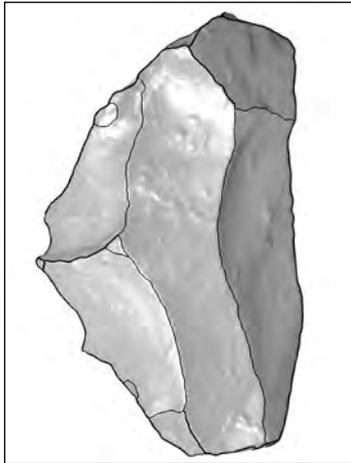


Figure 8. The final scars and ridges.

of objects). (Tangelder and Veltkamp 2008) define several categories of 3D shape matching methods, for example, global descriptors and local features. Both these methods “take into account only the pure geometry of the shape”. In contrast, graph based methods “attempt to extract a geometric meaning from a 3D shape using a graph showing how shape components are linked together”.

In the case of lithic artifacts, proper segmentation of the surface into meaningful scars allows generating a graph-based descriptor, which is rich and descriptive on one hand and with a reasonable dimensionality (k) on the other hand. Figure 9 shows the scar adjacency graph diagram super-imposed on the drawing of the object. Note that each scar node can contain several attributes of each scar, such as:

- Area of the scar,
- Parameters of best-fitting plane, planarity value,
- Mean normal direction,
- Histogram of curvature values (1D or 2D),
- Shape of scar (use some shape feature) / of ridge.

In addition, edges can hold information about the borders between scars, for example, the mean curvature value along the border. Usage of graph-based descriptors require a definition of a metric and some graph matching algorithm such as (Melnik et al. 2002) in the classification or clustering phase.

In addition to the scars graph-based descriptor, global features such as the ones described in

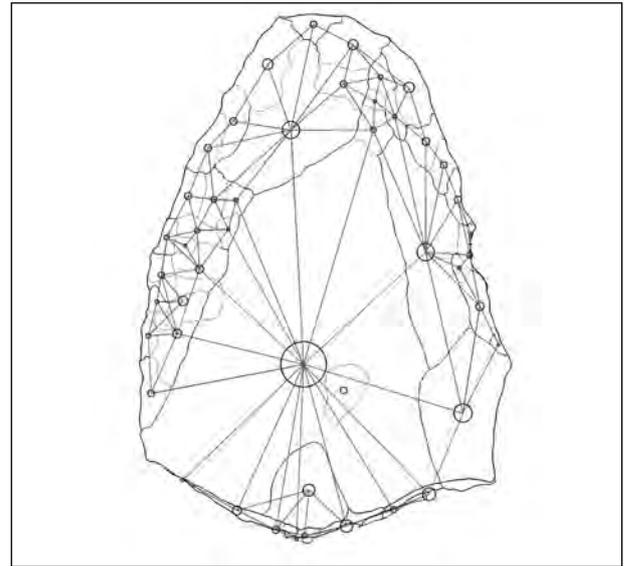


Figure 9. Lithic artifact represented as an adjacency graph of the scars. Size of the scar nodes in the drawing (blue circles) is proportional to the diameter of the scar. Edges (red) connect adjacent scars. The edges connecting to main scar on the back of the object are not shown.

(Tangelder and Veltkamp 2008) can be calculated for the entire object:

- The 3D shape histogram, as defined by (Otagiri et al. 2002),
- Histogram of the principal curvature (1D or 2D),
- Histograms of normal distances from centroids,
- The volumetric moments of inertia (Zhang and Chen 2001).

5. Results

Figure 10 shows the output of our method on 13 different stone tools. The figure shows the scar ridges (the borders between the detected scars).

Next, we compare our results to hand-drawings (Fig. 11). As can be seen, our method created a segmentation which is quite similar to the segmentation that is expressed in the hand-drawing, although in some cases, our method broke a single scar to two or more parts. This is further discussed in section 6.

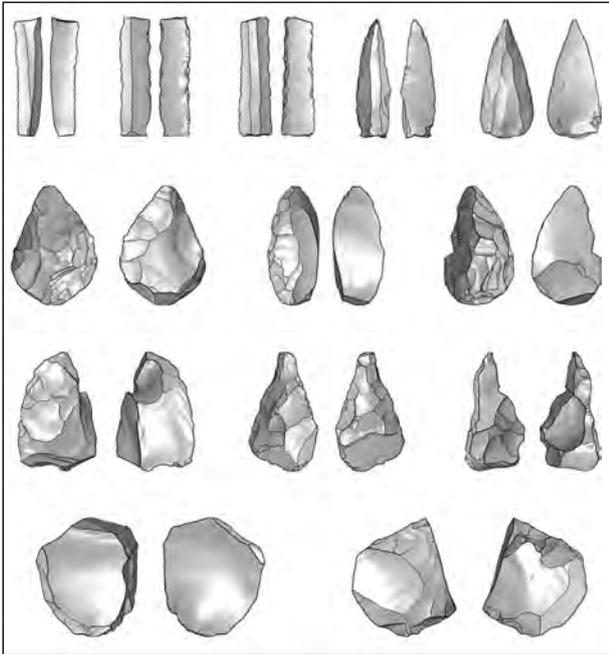


Figure 10. Scar segmentation results.

We compared the performance of our method to Watershed (Mangan and Whitaker 1998), which is considered to be one of the more robust mesh segmentation methods. We ran an implementation of Watershed that uses the principal curvature value as a height function. As can be seen in Figure 12, the out-of-the-box performance of Watershed is quite good, but compared to our method, it produces less accurate ridges and fails to segment some of the scars.

6. Discussion and Future Work

The surface scars and ridges of stone tools are highly valuable for morphological study and for documentation. We presented a method for automatic detection of the scars and ridges on 3D-scanned lithic artifacts and for extraction of quantitative global and scar-related features, which can be used for clustering and classification of prehistoric artifacts. As presented in section 5, the scar segmentation result is consistent with prior methods, however, it is not always identical to the manual segmentation, which is expressed by the hand drawing. For example, our method might segment a scar into two parts because of some irregularity of the scar surface, while in the manual segmentation, the archaeologist can decide to define

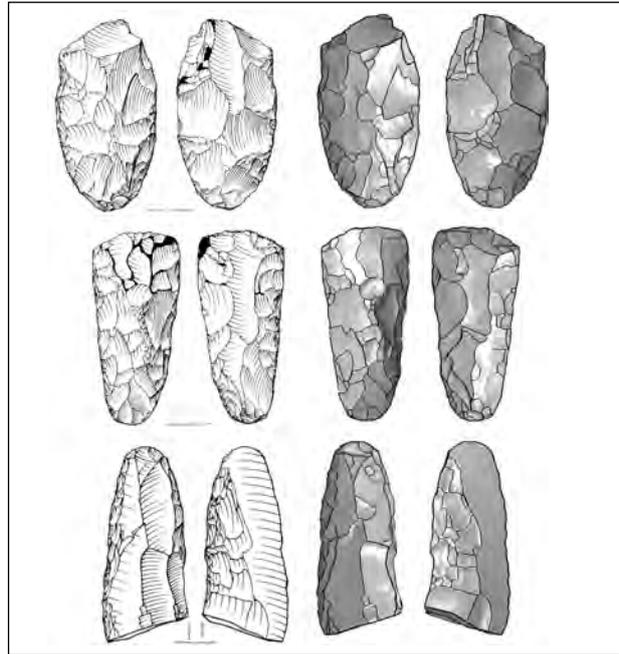


Figure 11. Segmentation results compared to traditional hand drawings (our results shown on the right.) Tools are from the Neolithic site of Abu-Gosh, excavated by Ianir Milevski, Drawn by Leonid Zeiger, Israel Antiquity Authority.

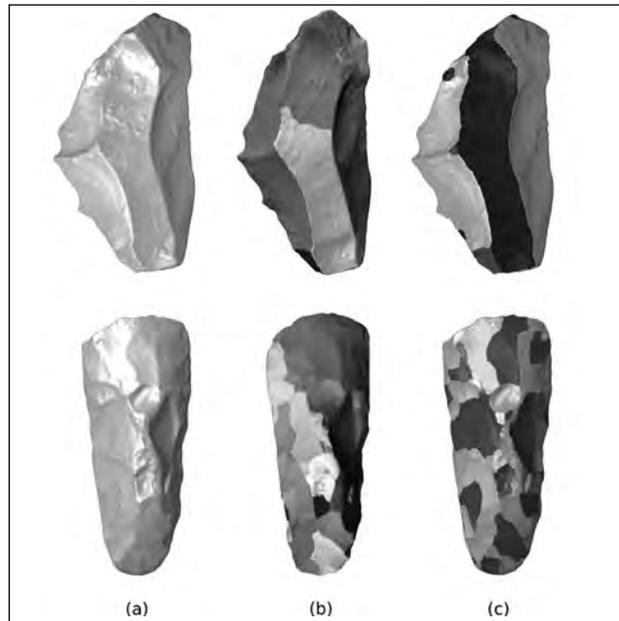


Figure 12. Segmentation results compared to Watershed. (a) The original objects. (b) Watershed results. The result is less consistent with the actual scars (e.g. on the right side of the object on the top row, at the top of the object in the second row). (c) Our method.

a single scar due to some higher-level archaeological considerations such as the position of the point of percussion. For the purpose of quantitative research (for example, clustering and classification of tools from several assemblages), which is the main goal of our work, we believe that the quality of our scar segmentation is sufficient, since it is performed in an objective and consistent manner on all scanned objects and it is mostly consistent with the archaeological scars and ridges definition.

We are now integrating the automatic segmentation algorithm in a computer code which will be used for documenting lithic tools in archaeological reports and articles. The purpose is to enable the archaeologist to mark selected scars and ridges on the 3D model, and produce 2D views of the artifacts which include the marked features. The markings should reflect the degree of confidence or importance attributed to the feature, such as i.e., using thicker lines to delineate the more outstanding ridges. Moreover, it will enable the archaeologist to introduce his/her judgment in cases of archaeologically uncertain situations as for example when two neighboring scars should be merged to a single one.

Future work can aim at enhancing the scar segmentation by introducing constraints or regularization terms on the shape of the scars (for example, requiring each scar to be convex) or by analyzing the surface of each scar. Detection of the conchoidal ripples on the surface of the scars can be performed by means of Fourier analysis. This can enhance both the quantitative description of the scar and the generated drawing.

As a next step towards utilization of the quantitative scar and ridge features, we plan to apply machine learning methods to automatically cluster and classify stone tools from several sites and from several periods.

References

Bar-Yosef, O., and P. Van Peer. 2009. "The Chaine Operatoire Approach in Middle Paleolithic Archaeology." *Current Anthropology* 50 (1): 103-131.

Boykov, Y.Y., and M.-P. Jolly. 2001. "Interactive Graph Cuts for Optimal Boundary & Region Segmentation of Objects in N-D Images." In *IEEE International Conference on Computer Vision*, 105-112. Washington: IEEE.

Clarkson, C., L. Vinicius, and M. Mirazon Lahr. 2006. "Quantifying Flake Scar Patterning on Cores using 3D Recording Techniques." *Journal of Archaeological Science* no. 33 (1): 132-142.

Cohen-Steiner, D., and J.-M. Morvan. 2003. "Restricted Delaunay Triangulations and Normal Cycle." In *Proceedings of the nineteenth annual symposium on Computational Geometry. San Diego, California*, 312-321. New York: ACM.

Davis, X.J., and J.J. Shea. 1998. "Quantifying Lithic Curation: An Experimental Test of Dibble and Pelcin's Original Flake-Tool Mass Predictor." *Journal of Archaeological Science* 25 (7): 603-610.

Eren, M. I., Manuel Dominguez-Rodrigo, Steven L. Kuhn, Daniel S. Adler, Ian Le, and Ofer Bar-Yosef. 2005. "Defining and Measuring Reduction in Unifacial Stone Tools." *Journal of Archaeological Science* 32 (8): 1190-1201.

Eren, M. I., and C. G. Sampson. 2009. "Kuhn's Geometric Index of Unifacial Stone Tool Reduction (GIUR): does it measure missing flake mass?" *Journal of Archaeological Science* 36 (6):1243-1247.

Golovinskiy, A., and T. Funkhouser. 2008. "Randomized Cuts for 3D Mesh Analysis." *ACM Transactions on Graphics* 27 (5): 1-12.

Grosman, L., O. Smikt, and U. Smilansky. 2008. "On the Application of 3-D Scanning Technology for the Documentation and Typology of Lithic Artifacts." *Journal of Archaeological Science* 35: 3101-3110.

Grosman, L., G. Sharon, T. Goldman-Neuman, O. Smikt, and U. Smilansky. 2011. "Studying Post Depositional Damage on Acheulian Bifaces using 3-D Scanning." *Journal of Human Evolution* 60 (4): 398-406.

Jackson, J. E. 2004. *A User's Guide to Principal Components*. London and New York: John Wiley & Sons.

Jolliffe, I.T. 2002. *Principal Component Analysis, 2nd Edition*. New York: Springer.

- Koenderink, J.J. 1990. *Solid Shape*. Cambridge: The MIT Press.
- Kuhn, S.L. 1990. "A Geometric Index of Reduction for Unifacial Stone Tools." *Journal of Archaeological Science* 17 (5): 583-593.
- Mangan, A.P., and R.T. Whitaker. 1998. "Surface Segmentation using Morphological Watersheds." In *Proceedings for IEEE Visualization 1998 Late Breaking Hot Topics*. Washington: IEEE.
- Meignen, L., J.-M. Geneste, and E. Boëda. 1990. "Identification de Chaines Operatoires Lithiques du Paleolithique Ancien et Moyen." *Paleo* 2: 43-80.
- Melnik, S., H. Garcia-Molina, and E. Rahm. 2002. "Similarity Flooding: A Versatile Graph Matching Algorithm and its Application to Schema Matching." In *International Conference on Data Engineering*. In *Proceedings of the 18th Conference on Data Engineering*, 117-128. Washington: IEEE Computer Society.
- Ohbuchi, R., T. Otagiri, M. Ibato, and T. Takei. 2002. *Shape-Similarity Search of Three-Dimensional Models using Parameterized Statistics*. In *10th Pacific Conference on Computer Graphics and Applications*. Washington: IEEE Computer Society.
- Sellet, F. 1993. "Chaine Operatoire; the Concept and its Applications." *Lithic Technology* 18 (1-2):106-112.
- Sethian, J. A. 1999. *Level Set Methods and Fast Marching Methods Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science*. Cambridge: Cambridge University Press.
- Shamir, A. 2008. "A Survey on Mesh Segmentation Techniques." *Computer Graphics Forum* 27 (6): 1539-1556.
- Siek, J., L. Lee-Quan, and A. Lumsdaine. 2002. *The Boost Graph Library: User Guide and Reference Manual*. Boston: Addison-Wesley.
- Tangelder, J., and R. Veltkamp. 2008. "A Survey of Content Based 3D Shape Retrieval Methods." *Multimedia Tools and Applications* 39 (3): 441-471. doi: 10.1007/s11042-007-0181-0.
- Yoshizawa, S., A. Belyaev, and H.-P. Seidel. 2005. "Fast and Robust Detection of Crest Lines on Meshes." In *Proceedings of the 2005 ACM Symposium on Solid and Physical Modeling*. Cambridge, Massachusetts. 227-232. New York: ACM.
- Zhang, C., and C. Tsuhan. 2001. "Efficient Feature Extraction for 2D/3D Objects in Mesh Representation." In *IEEE International Conference on Image Processing*, 935-938. Washington: IEEE.

Retracing Prehistoric Population Events in Finland Using Simulation

Tarja Sundell, Juhana Kammonen, Martin Heger, Jukka U. Palo and Päivi Onkamo

University of Helsinki, Finland

Abstract:

Due to Finland's low population density, geographical location and gradual settlement history, the country has been a genetic isolate. Genetic diversity is still slightly reduced among the Finns, especially in the eastern part of the country. This diversity reduction and the specific 'Finnish Disease Heritage' are explained by founder effects and bottlenecks and there is archaeological evidence for profound fluctuation of population sizes in the past. Here, we apply forward-time population simulations including archaeologically justified population size estimates to detect past demographic events. Our results indicate that simulation models beginning with serial founder effects have clearly reduced genetic diversity in early prehistory. Compared with our previous simulations, the gender-specific migration brings the simulated genetic diversity closer to the observed contemporary genetic diversity in Finland. Additionally, the simulation scenarios with temperate constant gene flow produce Y chromosomal diversity measures similar to those observed in present-day Finnish population.

Keywords:

Prehistoric Population, Population Simulation, Population Bottleneck, Stone Age, Early Metal Period, Finland

1. Introduction

The first pioneers settled in the Finnish region circa 11,000 BP following the retreat of the continental ice sheet after the last glaciation. According to archaeological finds, the region has been continuously inhabited ever since. Finland's prehistoric population events and settlement history have been studied by a number of archaeologists (e.g. Siiriäinen 1981; Meinander 1984; Nuñez, M. 1987; Huurre 1990, 2001; Lavento 1997, 1998, 2001; Carpelan 1999a, 1999b; Edgren 1999; Halinen 1999, 2005; Nuñez & Okkonen 1999; Mökkönen 2002, 2011; Pesonen 2002, 2005; Takala 2004; Rankama & Kankaanpää 2008; Saipio 2008; Tallavaara, Pesonen & Oinonen 2010; Tallavaara & Seppä 2012).

On the other hand, geneticists have carried out numerous studies concerning the population genetics and the origins of Finns (e.g. Nevanlinna 1972, Cavalli-Sforza et al. 1994; Sajantila et al. 1995, 1996; Lahermo et al. 1996, 1998; Lahermo 1998; De la Chapelle & Wright 1998; De la Chapelle 1999; Kittles et al. 1998, 1999; Savontaus & Lahermo 1999; Norio 2000, 2004; Peltonen et al. 2000; Varilo et al. 2000, 2003; Kere 2001; Uimari et al.

2005, Lappalainen et al. 2006, 2008; Lappalainen 2009; Salmela et al. 2006, 2008; Service et al. 2006, Hedman et al. 2007; Hannelius et al. 2008; Jakkula et al. 2008; McEvoy et al. 2009; Palo et al. 2009). Finns are, when compared to West European populations, an outlier (Lao et al. 2008). Especially in the male-mediated Y chromosome diversity is reduced and there are significant differences between eastern and western Finland. The genetic diversity reduction and the specific 'Finnish Disease Heritage' (Norio 2003a, 2003b, 2003c), the enrichment of rare endemic genetic diseases in Finland, is explained by a profound founder effect or a bottleneck in the past.

Adequately preserved organic remains from the Mesolithic and Neolithic yielding DNA are practically non-existent in Finland due to the naturally acidic soil. Thus, we cannot obtain direct evidence of haplotypes or diversities in DNA of the ancient inhabitants. Therefore, we combine knowledge from two different fields of science to reconstruct prehistoric demographic events: population size estimates based on archaeological data and population genetic simulations. Disentangling past population events is a complex task that benefits from the synthesis of

Corresponding author: tarja.sundell@helsinki.fi

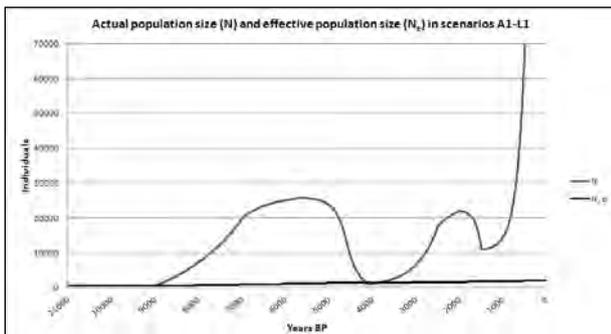


Figure 1. The actual total population size (N , dark grey) and theoretical effective population size (N_e , black). The effective population size has been calculated as the harmonic mean of the actual population size.

archaeological and genetic data. In this contribution we build upon our previous work (Sundell et al. 2010) and evaluate the effects of an initial founder effect vs. serial founder effects, internal and gender-specific migrations as well as migration waves and a Neolithic population bottleneck on the Finnish gene pool.

1.1 Archaeological background

When assessing population size the principal assumption is that the archaeological signal, evidenced by e.g. radiocarbon date distribution, correlates with the population size: the stronger the detected archaeological signal is, the larger the population that left the signal has been. This does not yield an absolute value of the population size at a given time but, instead, estimates the relative population sizes between consecutive time periods.

The archaeological records from all of the excavations conducted in Finland have been documented in the Registry of Ancient Monuments, a national database of the National Board of Antiquities (http://kulttuuriymparisto.nba.fi/netsovellus/rekisteriportaali/mjreki/read/asp/r_default.aspx).

According to the Registry there are (in June 2012) 10,150 dwelling sites on mainland Finland that belong to the Stone Age (ca. 8800-1800 calBC) while only 556 belong to the Early Metal Period (1800 calBC -1 AD). There seems to be a clear difference between the numbers of settlement sites in inland culture when turning from the Neolithic to the Early Metal Period (Lavento 1997, 2001; Saipio 2008).

The Early Metal Period dwelling sites are usually smaller and spatially differently distributed than in the previous period. The small number of structures found and the smaller sizes of sites could suggest temporary sites, possibly associated with hunter-gathering population and their seasonal settlement model. According to the archaeological signal, the number of inhabitants severely declined from the periods when large villages were settled in Saimaa, ca. 2500-1800 calBC (Lavento 2001). Although it can be argued that Early Metal Period dwelling sites are more difficult to find, the taphonomic processes (Surovell and Brantingham 2007; Surovell et al. 2009) would, however, have more impact on older Stone Age sites and should cause a sampling bias favouring the Early Metal Period sites.

Despite the uncertainties in defining what constitutes a dwelling site, the difference between the numbers of Stone Age and Early Metal Period sites appears substantial. The above-mentioned facts may thus reflect a genuine decrease in population size, in genetic terms a bottleneck. The archaeological evidence (see above) suggests that this may have been especially notable in eastern Finland between the Late Neolithic and the Early Metal Period.

In addition to the discrepancy between the numbers of dwelling sites from preceding eras, archaeological evidence tells about other demographic processes which can be used as prior knowledge in the simulations, e.g. possible migration from neighbouring areas. The spread of Typical Comb Ware (TCW) culture into Finland ca. 4000-3500 calBC is apparent and brings significant changes into archaeological finds. Together with a novel style of ceramics, there is a notable rise in the frequency of new raw materials such as flint and amber as well as a new house building tradition (e.g. Vuorinen 1982; Meinander 1984; Carpelan 1999a; Halinen 1999; Pesonen 2002; Edgren 2007). Furthermore, the population reaches a maximum at that time (Tallavaara et al. 2010).

The population peak is also evident through a Bayesian spatial analysis of archaeological record from the time period (Kammonen et al. 2012). Moreover, the weaker archaeological signal following the population peak at 4000-3500 calBC indicates a subsequent decline of the population. This has been

rendered as proof of a plausible Neolithic population bottleneck (Lavento 2001, Sundell et al. 2010) in Finland. Additionally, the spread of Corded Ware (CW) culture (ca. 2900-2300 calBC) has influenced the repertory of archaeological finds in the country (e.g. Edgren 1999; Carpelan 1999a). Again, a new style in ceramics together with new artifacts and new settlement patterns as well as burial practices.

1.2 Population genetic simulations

Population genetic simulation tools can be used to study the effects of population demography on genetic diversity over centuries or even millennia. The data sets simulated under different population historic scenarios can be contrasted with the observed genetic variation at the studied loci and thus support or refute certain scenarios.

In our simulation, each individual carries a mitochondrial genome and, if male, also a Y chromosome. These genomes are affected by similar evolutionary forces as natural chromosomes: they are transmitted from generation to generation, they are prone to mutate and their frequencies drift by chance. However, when contrasted to autosomal chromosomes, these molecules are haploid and uniparentally inherited, and as a consequence have lower molecular effective sizes. Hence they respond more rapidly to changes in population demography than autosomes. In addition, characteristic demographic processes e.g. population growth and decline, population subdivision and migration, can be incorporated into the models.

Our simulations begin at 11,000 BP when the first postglacial pioneers settled the country. The population growth is simulated according to the strength of the archaeological signal (Tallavaara et al. 2010). We employ two archaeologically justified bottlenecks: one at 4100-3800 BP and another at 1500-1300 BP. We split the population into geographic sub-populations, add gender-specific internal migration as well as migration waves and constant gene flow from neighbouring populations, compatible with archaeological phenomena.

2. Materials and Methods

We used simuPOP (Peng & Kimmel 2005), a forward-time population genetics simulation

environment. Forward simulations can be used to create virtual populations which are then simulated through entire simplified model histories. SimuPOP consists of a number of components from which users assemble a suitable simulator. These components are operated through Python script files.

The simulation moves forward in ten-year-steps. The simulated population is age-structured i.e. generations overlap. Reproductive ages are set at 20-60 years for males and 20-40 years for females. The maximum lifespan is 60 years and the natural mortality rate is 15 % per ten years. The mating of individuals as well as the selection of those individuals who die is based on random sampling. Each mating produces 1 to N offspring according to a zero-truncated Poisson distribution ($\lambda = 2$). Offspring is produced until it, together with the surviving part of the population, reaches the population size at the next simulation step.

The simulations focus on mitochondrial and Y chromosomal variation. Mitochondrial DNA (mtDNA) is maternally inherited and simulated as a 631 bp DNA sequence, corresponding to HVS-I and HVS-II of the mtDNA control region. Any nucleotide at any position can mutate creating new variation. The mutation rates (μ) follow the 2-parameter model of Kimura (1980), distinguishing between transitions ($\mu = 0.00000128/\text{locus}/10 \text{ yrs}$) and transversions ($\mu = 0.00000064/\text{locus}/10 \text{ yrs}$). The Y chromosome is simulated as a haplotype consisting of alleles from linked STR loci, each allele expressing the numbers of short tandem repeats per locus. The mutation model is stepwise, thus new variation is introduced by mutations that either decrease or increase the number of repeats in a locus. The mutation rate is $0.0007/\text{locus}/10 \text{ years}$. Our aim is to observe how different simulation scenarios affect the genetic diversity and whether any of the models produce diversity measures similar to those perceived in present-day Finns.

The population is simulated through its entire history (Figs 1 and 2). The simulation starts at 11,000 BP with a small initial pioneer population of 250 females and 250 males that is divided into two subpopulations: Saami and Other Finland. This is run for 200 steps, or 2000 years, to create a more natural, highly interrelated small founder population. We employ two different scenarios for

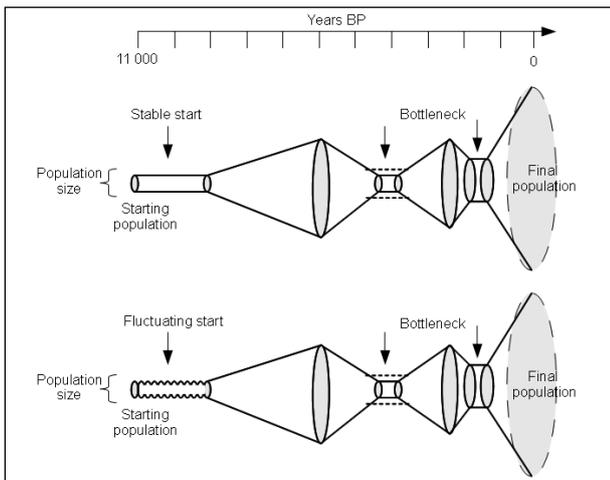


Figure 2. The general demographic model. The width of the cone represents the relative population size (not to scale) at the time before present depicted on the x axis.

this initial phase: a stable starting population is compared with a fluctuating start i.e. serial founder effects. In the stable starting population models, the population size remains constant at 500 individuals (Saami 250, Other Finland 250), whereas in the serial founder effect models, the population fluctuates between 240-630 (Saami 120-315, Other Finland 120-315 respectively). Both subpopulations have the same size change rate, with population minima reached ca. every 200 years.

The total population size is set to grow exponentially from 9000 BP (simulation step 200) onward until the next bottleneck occurs. Immediately after 7000 BP (simulation step 400), and at a size of 15,000 individuals, the Other Finland subpopulation is divided into two: the South-western (SW) subpopulation and the North-eastern (NE) subpopulation. The respective sizes for these subpopulations are 5,000 and 10,000. Rather than a drastic population event taking place during a single simulation step, this split event is necessary to later enable internal migration. In simulation scenarios with internal gender-specific migration, the SW and NE subpopulations are allowed to exchange individuals.

We include two migration waves from neighbouring populations, the TCW migration wave at around 6000 BP (simulation steps 499, 500 and 501) and the CW migration wave at around 5200 BP (simulation steps 579, 580 and 581). Total

population is set to reach a maximum of 25,000 shortly before 5750 BP (simulation step 525), the Stone Age population peak. The exponential growth slows down before this and the population size remains approximately the same for a brief period after this time point. After 5750 BP the population begins to gradually decline over 1600 years towards the Stone Age bottleneck. After the bottleneck the population recovers, followed by a less severe second bottleneck at 1500-1300 BP. During the last 1300 years the population size is set to grow to 1,000,000 individuals, the final population size.

2.1 The genetic diversity in the background populations

A few migration waves have been alleged to have entered Finland from neighbouring populations (see e.g. Carpelan 1999a, Lavento 2001). In addition to these, there has probably been minor, more or less constant, gene flow into the country. To allow for these phenomena, we formed three background populations: Archaic European, Archaic Scandinavian and Background Saami. From these, individuals migrate into the Finnish subpopulations with rates described below. We have selected actual ancient mtDNA haplotypes to form these two archaic background populations (Haak et al. 2005; Rudbeck et al. 2005; Melchior et al. 2007, 2008; Bramanti et al. 2009; Malmström et al. 2009; Skoglund et al. 2012). Since there is no information of ancient Saami DNA, both the Saami subpopulation and the Background Saami population were initialized with an approximation of present day Saami haplotype composition.

The Archaic European background population includes 50,000 individuals, the Archaic Scandinavian 25,000 individuals and the Background Saami 5,000 individuals. The sizes here are suggestive as we merely wish to evaluate if gene flow from other populations would have dramatic effects on the variation of Finnish subpopulations. Two sizes of migration waves were tested (see Table 1, scenarios E1-L2): temperate and small. In the temperate migration waves, the Typical Comb Ware migration wave replaces ca. 8% of the north-eastern subpopulation with Archaic European and the Corded Ware migration wave replaces ca. 3% of the south-western subpopulation with Archaic European. With smaller migration waves the

Scenario	Population size first 2000 years	Bottleneck size at 4100-3800 BP	Internal migration between subpopulations	Migration waves (TCW and CW)	Constant gene flow
A1	stable	1000	no	-	-
A2	stable	200	no	-	-
B1	stable	1000	yes	-	-
B2	stable	200	yes	-	-
C1	fluctuating	1000	no	-	-
C2	fluctuating	200	no	-	-
D1	fluctuating	1000	yes	-	-
D2	fluctuating	200	yes	-	-
E1	stable	1000	no	small	temperate
E2	stable	200	no	small	temperate
F1	stable	1000	yes	small	temperate
F2	stable	200	yes	small	temperate
G1	fluctuating	1000	no	small	temperate
G2	fluctuating	200	no	small	temperate
H1	fluctuating	1000	yes	small	temperate
H2	fluctuating	200	yes	small	temperate
I1	stable	1000	no	temperate	small
I2	stable	200	no	temperate	small
J1	stable	1000	yes	temperate	small
J2	stable	200	yes	temperate	small
K1	fluctuating	1000	no	temperate	small
K2	fluctuating	200	no	temperate	small
L	fluctuating	1000	yes	temperate	small
L2	fluctuating	200	yes	temperate	small

Table 1. Top. A reference matrix of different simulation scenarios (A1-L2). In addition to different simulation starts and bottleneck sizes, the other parameters are internal and gender specific migration between subpopulations, migration waves from Typical Comb Ware (TCW) and Corded Ware (CW) and constant gene flow from neighbouring populations.

	Checkpoint	Time BP	Simulation steps
1	Starting point	11,000	0
2	Population starts to grow	9,000	200
3	Sampling before the population split	7,000	400
4	One step after the population split	6,990	401
5	Population peak	5,750	525
6	Before the first bottleneck	4,650	635
7	Immediately after the population minimum	4,090	691
8	Second population peak	2,000	900
9	In the middle of the second bottleneck	1,400	960
10	Present time, final generation	0	1099

Table 2. Bottom. Simulation checkpoints. In every checkpoint, a random sample of 1000 individuals is extracted from the population extant at the time. If total population size at this point is smaller than 1000, all individuals in the population are included in the sample.

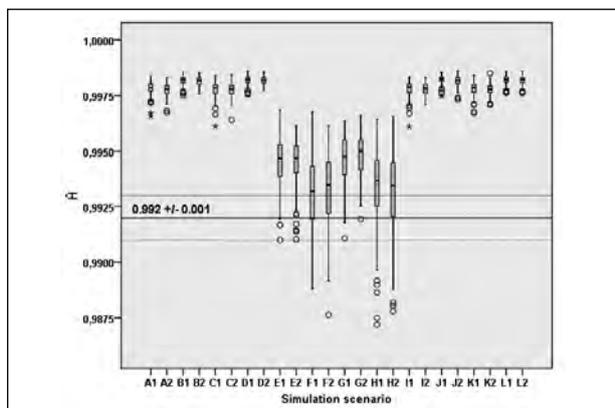


Figure 5. Y-STR haplotype diversity (\hat{H}) in final population (0 BP) in a sample of 907 individuals. The reference line represents the value observed in the present (Palo et al. 2009).

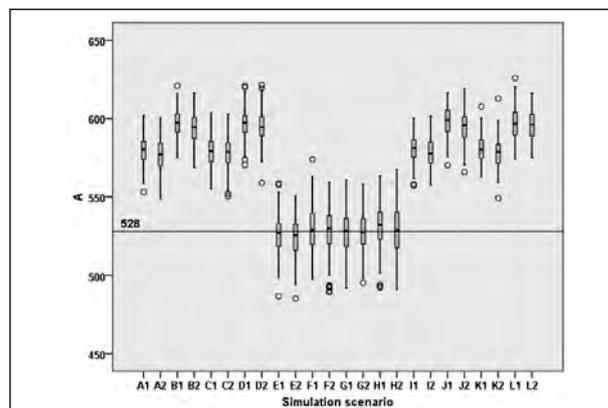


Figure 6. Number of Y-STR haplotypes (A) in final population (0 BP) in a sample of 907 individuals. The reference line represents the value observed in the present (Palo et al. 2009).

on the subpopulation level closer to the observed contemporary genetic diversity in Finland (Fig. 3).

Second, our results indicate that simulation models beginning with serial founder effects clearly reduce genetic diversity at the first checkpoint after the initial phase (Fig. 4).

Finally, the simulation scenarios with temperate constant gene flow produce Y chromosomal diversity measures similar to those observed in present-day Finnish population (Figs 5 and 6).

4. Discussion

The ultimate aim of these simulations is to determine which population scenarios are most likely to be true by assessing whether any of the models produce diversity measures similar to those perceived in present-day Finns. Previously we simulated Finland as one large population without substructure. In this refined model we divided the population into subpopulations, which, in addition to being more realistic, also enables gender-specific internal migration. Population subdivision also allows genetic drift to operate more realistically. In comparison with our previous study, the gender-specific migration brings the simulated mitochondrial genetic diversity closer to the observed contemporary genetic diversity in Finland. Y chromosomes in these simulations mutate faster, which obscures the effect of migration.

Our results indicate that simulation models beginning with serial founder effects clearly reduce genetic diversity at the first checkpoint after the initial phase. In later phases, however, other factors seem to obscure this effect. Therefore we cannot distinguish in this study whether models beginning with or without serial founder effects fit observed current genetic diversity better.

In our previous study (Sundell et al. 2010) we used higher gene flow rates and came to the conclusion that lower gene flow rates should be explored. The rates used there apparently forced the simulated populations' diversity to change towards the background populations'. The simulation scenarios with temperate constant gene flow produce Y chromosomal diversity measures similar to those observed in present-day Finnish population. Indeed, all the models with temperate constant gene flow result in diversity measures compatible with Finnish genetic diversity, indicating that the background gene flow is actually a more important factor than initial bottleneck size, at least with the bottleneck sizes we used.

3.1 Subsistence

The deglaciation of the region took several thousand years, during which there were major changes in the subsistence conditions of pioneer settlers. Population size fluctuation during the pioneering stage caused by serial founder effects can be argued to more realistically model the conditions

in the region during this phase of settlement. It is highly plausible that some small groups migrating into the country have not persisted. The reproduction rate has most likely been extremely low at times, barely sufficient for inhabitation to continue. In the Boreal climate zone livelihood preconditions are limited and small changes in climatic conditions can already be significant in determining the success of different subsistence models.

3.2 Limitations of simulations

Simulations are a useful tool in studying population processes unreachable by other means. They may help us to find the demographic model capturing the important features that shaped genetic diversity. However, simulations have their limitations. Simulation models can never be an exact replication of the complex reality, but instead, are more useful as evidence that a certain model may depict reality better than another. However, one must also note that the similarities between the modelled and observed diversities do not directly prove causality. Therefore, simulations can be considered to perform best when refuting some hypotheses. Furthermore, different models created by different simulation settings may replicate the observed patterns of diversity equally well, thus leaving us with several possibilities. In addition, with an increase in parameters, simulation can become very complex, i.e. modelling with too many variable parameters may obscure otherwise justified conclusions.

Nevertheless, simulations are beneficial in evaluating the credibility of different scenarios. One of the principal benefits of a simulation model is that you can begin with a simple approximation of a process and gradually refine the model as your understanding of the process improves. This allows achieving good approximations of very complex models. Additionally, simulations allow randomness and are an error-tolerant approach enabling repeatability. As a result, models with several interrelating parameters will always benefit from using simulation. Despite the weaknesses entailed in simulations, the importance of acquiring credible models by simulating is evident. As statistician George E. P. Box put it when exploring the relationships between interdependent variables: “Essentially, all models are wrong, but some are

useful” (Box and Draper 1987).

3.3 Future prospects

Incorporation of autosomal loci is one of our future goals. Adding them would bridge the gap between the effects of maternal and paternal inheritance. Furthermore, studying population admixture in more detail would also be highly interesting. Genetic consequences of population admixture, given that the source populations are somewhat different, can be observed in allele frequencies e.g. mtDNA and Y chromosome distributions and marker-to-marker linkage disequilibrium (LD) patterns. Interestingly, new research on natural and sexual selection in a historical Finnish population has been published very recently (Courtiol et al. 2012); it was shown that even in historical times, males have had higher variance in reproductive success than females. This could be added to our mating model to evaluate how much it would affect the Y chromosomal diversity.

Acknowledgements

The authors would like to thank the sample donors, as well as Prof. Antti Sajantila for the Finnish mitochondrial and Y-STR data. We also thank two anonymous referees for their comments on the manuscript. This work was supported by the Academy of Finland. This paper is part of a multidisciplinary research project, Argeopop (<http://www.helsinki.fi/bioscience/arageopop/>), researching Finnish population prehistory (Sundell 2008, 2009; Pesonen & Tallavaara 2008; Sundell et al. 2010, Sundell & Onkamo 2010, Heger 2011; Kammonen in preparation).

References

- Argeopop. 2012. “Argeopop - University of Helsinki.” Last modified June 27, 2012. <http://www.helsinki.fi/bioscience/arageopop/>.
- Box, G. E. P., and N. R. Draper. 1987. *Empirical Model-Building and Response Surfaces*. New York: Wiley.
- Bramanti, B., M. G. Thomas, W. Haak, M. Unterlaender, P. Jores, K. Tambets, I. Antanaitis-Jacobs, M. N. Haidle, R. Jankauskas, R., C.-J. Kind, F. Lueth, T. Terberger, J. Hiller, S. Matsumura, P. Forster, and J. Burger, J. 2009.

- “Genetic Discontinuity Between Local Hunter-Gatherers and Central Europe’s First Farmers.” *Science* 326: 137-40.
- Carpelan, C. 1999a. “Käännekohtia Suomen esihistoriassa aikavälillä 5100-1000 eKr. Pohjan poluilla. Suomalaisten juuret nykytutkimuksen mukaan.” In *Bidrag till kännedom av Finlands natur och folk*, edited by P. Fogelberg, 249-80. Helsinki: Societas Scientarum Fennica.
- Carpelan, C. 1999b. “On the Postglacial Colonisation of Eastern Fennoscandia.” In *Dig it all. Papers dedicated to Ari Siiriäinen*, edited by M. Huurre, 151-71. Helsinki: The Finnish Antiquarian Society.
- Cavalli-Sforza, L.L., P. Menozzi, and A. Piazza. 1994. *The History and Geography of Human Genes*. Princeton: Princeton University Press.
- Courtial, A., J. E. Pettay, M. Jokela, A. Rotkirch, V. Lummaa. 2012. “Natural and sexual selection in a monogamous historical human population.” *Proceedings of the National Academy of Sciences* 109 (21): 8044-49.
- De la Chapelle, A., and F. A. Wright. 1998. “Linkage disequilibrium mapping in isolated populations: the example of Finland revisited.” *Proceeding of the National Acadademy of Scince USA* 95: 12416-23.
- De la Chapelle, A. 1999. “Geenimutaatiot Suomen väestön juurten tutkimuksissa. Pohjan poluilla. Suomalaisten juuret nykytutkimuksen mukaan.” In *Bidrag till kännedom av Finlands natur och folk* 153, edited by P. Fogelberg, 307-10. Helsinki: Societas Scientarum Fennica.
- Edgren, T. 1999. “Käännekohtia Suomen kivikaudessa. Pohjan poluilla. Suomalaisten juuret nykytutkimuksen mukaan.” n *Bidrag till kännedom av Finlands natur och folk* 153, edited by P. Fogelberg, 281-93. Helsinki: Societas Scientarum Fennica.
- Edgren, T. 2007. “On the non-megalithic mortuary practices in Finland.” In *Innovation and Continuity – Non-Megalithic Mortuary Practices in the Baltic. New Methods and Research into the Development of Stone Age Society*, edited by L. Larsson, F. Lüth and T. Terberger, 501-20. Mainz: Bericht der Römisch-Germanischen Kommission 88.
- Haak, W., P. Forster, B. Bramanti, S. Matsumura, G. Brandt, M. Tänzler, R. Villems, C. Renfrew, D. Gronenborn, K. W. Alt, and J. Burger. 2005. “Ancient DNA from the First European farmers in 7500-Year-Old Neolithic Sites.” *Science* 310: 1016-8.
- Halinen, P. 1999. “Burial Practices and the Structure of Societies during the Stone Age in Finland.” In *Dig it all. Papers dedicated to Ari Siiriäinen*, edited by M. Huurre, 173-9. Helsinki: The Finnish Antiquarian Society.
- Halinen, P. 2005. “Prehistoric Hunters of Northernmost Lapland. Settlement patterns and subsistence strategies.” PhD diss. University of Helsinki.
- Hannelius, U., Salmela, E., Lappalainen, T., Guillot, G., Lindgren, C. M., von Döbeln, U., Lahermo, P., and Kere, J. 2008. “Population substructure in Finland and Sweden revealed by the use of spatial coordinates and a small number of unlinked autosomal SNPs.” *BioMedCentral Genetics* 9(54): 1-12.
- Hedman, M., Brandstätter, A., Pimenoff, V., Sistonen, P., Palo, J. U., Parson, W., and Sajantila, A. 2007. “Finnish mitochondrial DNA HVS-I and HVS-II population data.” *Forensic Science International* 172: 171-8.
- Heger, M. 2011. “Population Genetic Simulations of the Savonian Expansion in Finland.” MSc thesis, University of Helsinki.
- Huurre, M. 1990. *9000 vuotta Suomen esihistoriaa*. Keuruu: Otava.
- Huurre, M. 2001. *Kivikauden Suomi*. Keuruu: Otava.
- Jakkula, E., Rehnström, K., Varilo, T., Pietiläinen, O.P.H., Paunio, T., Pedersen, N. L., deFaire, U., Järvelin, M-R., Saharinen, J., Freimer, N., Ripatti, S., Purcell, S., Collins, A., Daly, M.J., Palotie, A. and Peltonen, L. 2008. “The Genome-wide patterns of Variation Expose significant Substructure in a Founder Population.” *The American Journal of Human Genetics* 83: 787-94.
- Kammonen J., Sundell T., Moltchanova E., Pesonen P., Oinonen M., Heger M., Haimila M., and Onkamo P. 2012. “Bayesian spatial analysis of archaeological finds and radiocarbon datings: An example from Finland 4000-3500 cal BC.” In *Revive the Past. Computer Applications and Quantitative Methods in Archaeology (CAA). Proceedings of the 39th International Conference, Beijing*, edited by M. Zhou, I. Romanowska, Z. Wu, P. Xu, and P. Verhagen, 318-25. Amsterdam: Pallas Publications.

- Kammonen, J. *In preparation*. "Using simuPOP to Model Finnish Demography Relative to Large Neighbouring Populations." MSc Thesis, University of Helsinki.
- Kere, J. 2001. "Human population genetics: Lessons from Finland." *Annual Review of Genomics and Human Genetics* 2: 103-28.
- Kimura, M. 1980. "A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences." *Journal of Molecular Evolution* 16: 111-20.
- Kittles, R. A., Perola, M. & Peltonen, L. 1998. "Dual origins of Finns revealed by Y chromosome haplotype variation." *The American Journal of Human Genetics* 62: 1171-9.
- Kittles, R. A., Bergen, A. W., Urbanek, M., Virkkunen, M., Linnoila, M., Goldman, D. & Long, J. C. 1999. "Autosomal, Mitochondrial and Y Chromosome DNA Variation in Finland: Evidence for a Male-Specific Bottleneck." *American Journal of Physical Anthropology* 108: 381-99.
- Lahermo, P. 1998. "Mitochondrial and Y Chromosomal Variations in the Finno-Ugric-Speaking Peoples." PhD diss. University of Turku.
- Lahermo, P., Sajantila, A., Sistonen, P., Lukka, M., Aula, P., Peltonen, L., Savontaus, M-L. 1996. "The Genetic Relationship between the Finns and the Finnish Sami (Lapps): Analysis of Nuclear DNA and mtDNA." *The American Journal of Human Genetics* 58: 1309-22.
- Lahermo, P., Savontaus, M-L., Sistonen, P., de Knijff, P., Aula, P., Sajantila, A. 1998. "Y chromosomal polymorphisms reveal founding lineages in the Finns and the Sami." *European Journal of Human Genetics* 7(4): 447-58.
- Lao, O., Lu, T.T., Nothnagel, M., Junge, O., Freitag-Wolf, S., Caliebe, A., Balascakova, M., Bertranpetit, J., Bindoff, L.A., Comas, D., Holmlund, G., Kouvatsi, A., Macek, M., Mollet, I., Parson, W., Palo, J., Ploski, R., Sajantila, A., Tagliabracci, A., Gether, U., Werge, T., Rivadeneira, F., Hofman, A., Uitterlinden, A.G., Gieger, C., Wichmann, H.E., Ruther, A., Schreiber, S., Becker, C., Nurnberg, P., Nelson, M.R., Krawczak, M., and Kayser, M. 2008. Correlation between Genetic and Geographic Structure in Europe. *Current Biology* 18 (16): 1241-1248.
- Lappalainen, T., Koivumäki, S., Salmela, E., Huoponen, K., Sistonen, P., Savontaus, M-L., and Lahermo, P. 2006. "Regional differences among Finns: A Y-chromosomal perspective." *Gene* 376: 207-15.
- Lappalainen, T., Laitinen, V., Salmela, E., Andersen, P., Huoponen, K., Savontaus, M-L., and Lahermo, P. 2008: "Migration Waves to the Baltic Sea Region." *Annals of Human Genetics* 72: 337-48.
- Lappalainen, T. 2009. "Human Genetic Variation in the Baltic Sea Region: Features of Population History and Natural Selection." PhD diss. University of Helsinki.
- Lavento, M. 1997. "Geoarchaeological Observations on the Early Metal Period dwelling sites in the Ancient Lake Saimaa area." In *Slavjane i Finno-ugri. Arheologija, istorija, kultura*, 38-46. St. Petersburg: Universitätsverlag.
- Lavento M. 1998. "Sisämaan vanhemman metallikauden väestö tutkimusongelmana." *Muinaistutkija* 4: 46-55.
- Lavento, M. 2001. "Textile Ceramics in Finland and on the Karelian Isthmus. Nine Variations and Fugue on a Theme of C. F. Meinander." PhD diss. University of Helsinki.
- Malmström, H., Gilbert, M.T.P., Thomas, M.G., Brandström, M., Storå, J., Molnar, P., Andersen, P.K., Bendixen, C., Holmlund, G., Götherström, A., Willerslev, E. 2009. "Ancient DNA Reveals Lack of Continuity between Neolithic Hunter-Gatherers and contemporary Scandinavians." *Current Biology* 19: 1758-62.
- McEvoy, B., G. Montgomery, A. McRae, S. Ripatti, M. Perola, T. Spector, L. Cherkas, K. Ahmadi, D. Boomsma, G. Willemsen, J. Hottenga, N. Pedersen, P. Magnusson, K. Kyvik, K. Christensen, J. Kaprio, K. Heikkilä, A. Palotie, E. Widen, J. Muilu, A.-C. Syvänen, U. Liljedahl, O. Hardiman, S. Cronin, L. Peltonen, N. Martin, and P. Visscher. 2009. "Geographical structure and differential natural selection among North European populations." *Genome Research* 19: 804-14.
- Meinander, C. F. 1984. "Kivikautemme väestöhistoria." In *Bidrag till kännedom av Finlands natur och folk*, edited by S.-E. Åström, 21-48. Helsinki: Societas Scientiarum Fennica.
- Melchior, L., Gilbert, M.T.P., Kivisild, T., Lynnerup, N., and Dissing, J. 2007. "Rare mtDNA Haplogroups and Genetic Differences in Rich and Poor Danish Iron-Age

- Villages.” *American Journal of Physical Anthropology* 135: 206-15.
- Melchior, L., Kivisild, T., Lynnerup, N., and Dissing, J. 2008. “Evidence of Authentic DNA from Danish Viking Age Skeletons Untouched by Humans for 1,000 Years.” *PLoSone*, 3(5): e2214.
- Mökkönen, T. 2002. “Chronological Variation in the Locations of Hunter-Gatherer Occupation Sites vis-à-vis the Environment.” In *Huts and Houses. Stone Age and Early Metal Age Buildings in Finland*, edited by H. Ranta, 53-64. Helsinki: Finnish National Board of Antiquities.
- Mökkönen, T. 2011. “Studies on Stone Age housepits in Fennoscandia (4000-2000 CAL BC). Changes in ground plan, site location, and degree of sedentism.” PhD diss. University of Helsinki.
- Nei, M. 1987. *Molecular evolutionary genetics*. New York: Columbia University Press.
- Nevanlinna, H. R. 1972. “The Finnish population structure. A genetic and genealogical study.” *Hereditas* 71: 195-236.
- Norio, R. 2000. *Suomi-neidon geenit*. Helsinki: Otava.
- Norio, R. 2003a. “Finnish Disease Heritage I: characteristics, cause, background.” *Human Genetics* 112 (5-6): 441-56.
- Norio, R. 2003b. “Finnish Disease Heritage II: population prehistory and genetic roots of Finns.” *Human Genetics* 112 (5-6): 457-69.
- Norio, R. 2003c. “Finnish Disease Heritage III: the individual diseases.” *Human Genetics* 112 (5-6): 470-526.
- Norio, R. 2004. *Suomalaisten mutkalliset geenijuuret*. In *Ennen, muinoin. Miten menneisyttämme tutkitaan?* edited by R. Grünthal, 174-87. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Nuñez, M. 1987. “A Model for the Early Settlement of Finland.” *Fennoscandia archaeologica* IV: 3-18.
- Nuñez, M. and Okkonen, J. 1999. “Environmental Background for the Rise and Fall of Villages and Megastructures in North Ostrobothnia 4000-2000 calBC.” In *Dig it all. Papers dedicated to Ari Siiriäinen*, edited by M. Huurre, 173-9. Helsinki: The Finnish Antiquarian Society and The Archaeological Society of Finland.
- Palo, J. U., Ulmanen, I., Lukka, M., Ellonen, P., and Sajantila, A. 2009. “Genetic markers and population history: Finland revisited.” *European Journal of Human Genetics* 17, 1336-46.
- Peng B., and Kimmel M. 2005. “SimuPOP: a forward-time population genetics simulation environment.” *Bioinformatics* 21: 3686-7.
- Peltonen, L., Palotie, A., and Lange, K. 2000. “Use of Population Isolates for Mapping Complex Traits.” *Nature Reviews Genetics* 1: 182-90.
- Pesonen, P. 2002. “Semisubterranean Houses In Finland – A Review.” In *Huts and Houses. Stone Age and Early Metal Age Buildings in Finland*, edited by H. Ranta, 53-64. Helsinki: Finnish National Board of Antiquities.
- Pesonen, P. 2005. “Sarvingin salaisuus – Enon Rahakankaan varhaismesoliittinen ajoitus.” *Muinaistutkija* 2: 2-13.
- Pesonen, P., and Tallavaara, M. 2008. *Suomen esihistoriallinen väestömäärä 8500 calBC – 1000 calAD*. Helsinki: University of Helsinki.
- Rankama, T., and Kankaanpää, J. 2008. “Eastern arrivals in post-glacial Lapland: The Sujala site 10000 cal BP.” *Antiquity* 82 (318): 884-899.
- Register of the Ancient Monuments. 2012. “Muinaisjäännösrekisteri.” Accessed 27 June 2012. http://kulttuuriymparisto.nba.fi/netsovellus/rekisteriportaali/read/asp/r_default.aspx.
- Rudbeck, L., Gilbert, M.T.P., Willerslev, E., Hansen, A.J., Lynnerup, N., Christensen, T., and Dissing, J. 2005. “MtDNA Analysis of Human Remains from an Early Danish Christian Cemetery.” *American Journal of Physical Anthropology* 128: 424-9.
- Saipio, J. 2008. “Hävisikö sisämaan asbestikeraaminen väestö pronssikaudella?” *Muinaistutkija* 2: 2-18.
- Sajantila, A., Lahermo, P., Anttinen, T., Lukka, M., Sistonen, P., Savontaus, M-L., Aula, P., Beckman, L., Tranebjaerg, L., Gedde-Dahl, T., Isser-Tarver, L., DiRienzo, A., and Pääbo, S. 1995. “Genes and Languages in Europe: An Analysis of Mitochondrial Lineages.”

Genome Research 5: 42-52.

Sajantila, A., Salem, A-H., Savolainen, P., Bauer, K., Gierig, C., and Pääbo, S. 1996. "Paternal and maternal DNA lineages reveal a bottleneck in the founding of the Finnish population." *Proceedings of the National Academy of Sciences USA* 93: 12035-9.

Salmela, E., Taskinen, O., Seppänen, J. K., Sistonen, P., Daly, M. J., Lahermo, P., Savontaus, M-L., and Kere, J. 2006. "Subpopulation difference scanning: a strategy for exclusion mapping of susceptibility genes." *Journal of Medical Genetics* 43(7): 590-7.

Salmela, E., Lappalainen, T., Fransson, I., Andersen, P. M., Dahlman-Wright, K., Fiebig, A., Sistonen, P., Savontaus, M-L., Schreiber, S., Kere, J., and Lahermo, P. 2008. "Genome-Wide Analysis of Single Nucleotide Polymorphisms Uncovers Population Structure in Northern Europe." *PLoS ONE* 3 (10): e3519.

Savontaus, M-L., and Lahermo, P. 1999. "Uralilainen muinaisuutemme väestögenetiikan valossa. Pohjan poluilla. Suomalaisten juuret nykytutkimuksen mukaan." In *Bidrag till kännedom av Finlands natur och folk* 153, 60-3. Helsinki: Societas Scientiarum Fennica.

Service, S., DeYoung, J., Karayiorgou, M., Roos J.L., Pretorius, H., Bedoya, G., Ospina, J., Ruiz-Linares, A., Macedo, A., Palha, J.A., Heutink, P., Aulchenko, Y., Oostra, B., van Duijn, C., Jarvelin, M.R., Varilo, T., Peddle, L., Rahman, P., Piras G., Monne, M., Murray, S., Galver, L., Peltonen, L., Sabatti, C., Collins, A., and Freimer, N. 2006. "Magnitude and distribution of linkage disequilibrium in population isolates and implications for genome-wide association studies." *Nature Genetics* 38: 556-60.

Siiriäinen, A. 1981. "On the cultural ecology of the Finnish Stone Age." *Suomen Museo* 87: 5-40.

Skoglund, P., Malmström, H., Raghavan, M., Storå, J. Hall, P., Willerslev, E., Gilbert, M.T.P., Götherström, A., and Jakobsson, M. 2012. "Origins and Genetic Legacy of Neolithic Farmers and Hunter-Gatherers in Europe." *Science* 336: 466-9.

Sundell, T. 2008. "Populaatiopullonkaulat kivikauden Suomessa – arkeologinen ja geneettinen tulkinta." MA thesis, University of Helsinki.

Sundell, T. 2009. "Kadonneet geenit: Suomen

kivikautisten populaatiopullonkaulojen tutkiminen geneettisin menetelmin." *Muinaistutkija* 4: 10-8.

Sundell T., Heger M., Kammonen J., and Onkamo P. 2010. "Modelling a Neolithic Population Bottleneck in Finland: A Genetic Simulation." *Fennoscandia Archaeologica* 27: 3-19.

Sundell, T., and P. Onkamo, 2010. "Argeopop: Keramiikasta geenivirtoihin?" *Muinaistutkija* 1: 3-8.

Surovell, T. A., and Brantingham, P. J. 2007. "A note on the use of temporal frequency distributions in studies of prehistoric demography." *Journal of Archaeological Science* 34: 1868-77.

Surovell, T. A., Finley, J. B., Smith, G. M., Brantingham, P. J., and Kelly, R. 2009. "Correcting temporal frequency distributions for taphonomic bias." *Journal of Archaeological Science* 36: 1715-24.

Takala, H. 2004. "The Ristola Site in Lahti and the Earliest Postglacial Settlement of South Finland." PhD Diss. University of Turku.

Tallavaara, M., and Seppä, H. 2012. "Did the mid-Holocene environmental changes cause the boom and bust of hunter-gatherer population size in eastern Fennoscandia?" *Holocene* 2: 215-225.

Tallavaara, M., Pesonen, P. & Oinonen, M. 2010. "Prehistoric population history in eastern Fennoscandia." *Journal of Archaeological Science* 37: 251-60.

Uimari, P., Kontkanen, O., Visscher, P. M., Pirskanen, M., Fuentes, R., and Salonen, J. T. 2005. "Genome-Wide Linkage Disequilibrium from 100,000 SNP's in the East Finland Founder Population." *Twin Research and Human Genetics* 8(3): 185-97.

Varilo, T., Laan, M., Hovatta, I., Wiebe, V., Terwilliger, J. D., and Peltonen, L. 2000. "Linkage disequilibrium in isolated populations: Finland and a young sub-population of Kuusamo." *European Journal of Human Genetics* 8: 604-12.

Varilo, T., Paunio, T., Parker, A., Perola, M., Meyer, J., Terwilliger, J. D., and Peltonen, L. 2003. "The interval of linkage disequilibrium (LD) detected with microsatellite and SNP markers in chromosomes of Finnish populations with different histories." *Human Molecular Genetics* 12: 51-9.

Vuorinen, J. 1982. *Piikivi ja Suomen kampakeraaminen piikauppa*. Helsinki: University of Helsinki.

SteppingIn - Modern Humans Moving into Europe - Implementation

Fulco Scherjon

Leiden University, The Netherlands

Abstract:

This research focuses on the possible routes taken by the earliest Modern Humans moving into Europe. Suggested routes include the Danube Corridor, the Gibraltar Passage, and the Sicilian Crossing. The aim of this research is to develop a simulation tool able to compare alternative scenarios. The simulator, conveniently called SteppingIn, represents groups of modern humans as autonomous agents. Simulations are run with varying model and individual parameters. Archaeological Checkpoints in Space and Time (CSTs) with known location and chronology serve to assess simulation results. Chosen routes are plotted on a map showing emerging colonization patterns through time. Such density plots can serve as a rudimentary predictive model on the likelihood of certain preferences. The simulator is implemented in the REcursive Porous Agent Simulation Toolkit (Repast Symphony 2.0), selected both for its geographical capabilities and the ease of Java integration. Output is presented in Google Earth for further analysis.

Keywords:

Modelling, Simulation, Modern Humans, Paleolithic Europe

1. Introduction

Modern humans, not only archaeologists, are intrigued by questions about our origins. One of those questions is simply put: “how did modern humans colonize Europe around 40.000 years ago?”. It is one of those key fundamental questions in paleoanthropology and paleolithic archaeology (Roe 2009, 5). It seems a question that is easily answered but upon closer inspection one that generates a myriad of other questions about the influence of paleoclimates, possible routes into Europe, seafaring capabilities, Neanderthals, extinction of multiple large animal species, the need for and use of language, the flexibility and effectiveness of toolsets, and about the working of the human mind. Unfortunately, archaeological data is so far insufficient to answer all aspects of that literal “how” question. It is unknown when modern humans first touched European soil (however you define Europe) and which route was taken afterwards...

The aim of this research is to deliver a simulation tool able to compare alternative scenarios for moving into Europe. The simulator, conveniently called SteppingIn, can calculate with several model parameters and show their effects on the geographical movement of our species. Simulation

Corresponding Author: f.scherjon@umail.leidenuniv.nl / fulco.scherjon@xs4all.nl

techniques have previously been used to analyse important events, like SteppingOut with *Homo erectus* leaving Africa (Mithen and Reed (2002) but also Nikitas and Nikita (2005)), the arrival of modern humans into the Americas (Steele et al. 1998), or the settlement patterns in the spreading agricultural way of life around the Mediterranean (Falconer 2005). SteppingIn will represent modern humans as autonomous agents with individual preferences in coping with the different challenges offered by the European environment from roughly 50.000 to 30.000 before present (BP).

There are several possible and contested dispersal colonization patterns as interpreted from find distributions, site locations and dating results (Hoffecker 2009). Alternatives include following the Danube river system (Conard and Bolus 2003; Nigst 2006), colonization along the Mediterranean coast (Davidson et al. 2006), coming from the Russian plains passing north of the Black Sea (Anikovich et al. 2007, 225), or directly from Africa, crossing at Gibraltar (Van Andel and Davies 2003, 44) or Sicily (Villa 2001, 113). Most prehistoric coastlines have disappeared and with it any sites that could provide information on coastline hypotheses (Bailey, Weaver and Hublin 2008). The easiest crossing at Gibraltar is even with a lowered sea surface on the far Atlantic side and more than 100 km long (Close 2009, 42) and with a dangerous turbulence (Straus 2001, 92-

93). Within the model for SteppingIn crossing at Gibraltar is possible and occurs but experiments with SteppingIn will focus on the Danube corridor hypothesis (see below).

The main drive behind movement in the SteppingIn model is the availability of water. If possible, the modelled humans will try to head for or follow rivers (small, medium and large), lakes and coastlines. Water resources have proved essential in population dispersals in later periods (Bertuzzo et al. 2007; Davison et al. 2006) and the hypothesis in SteppingIn is that water was no less important in palaeolithic times. The model takes Aurignacian sites as proxies (indicators) for the presence of modern humans (Bailey, Weaver and Hublin 2009; Mellars 2006, 167; Nigst 2006, 297; Verpoorte 2005). In each simulation a set of agents with different preferences for their environment move through geographical space. SteppingIn will allow the user to define exactly what parameters are important and which ones are not. Then the

system will show what the effects of these settings are on the modelled system, in what routes are taken by the agents and how fast colonization occurs. Checkpoints with known location and chronology serve to compare simulation results with each other and with the archaeological reality. The chosen routes in the simulations are plotted on maps and those can serve as rudimentary predictive models indicating where archaeology can be expected.

Agent-based Modelling (ABM) focuses on the behaviour of an individual entity that acts according to certain rules. It enables the study of how system-level and individual-level patterns emerge without central controller and how the system evolves over time (Bonabeau 2002). Agents interact in rule-based ways. An agent perceives and acts upon her environment. Her behaviour depends partially on her own experience and she herself decides what actions are needed to satisfy her objectives. Key characteristics of the agents in ABM are individuality, interaction with the environment,

	Preference name	Abbreviation	Meaning
#0#	IS_CLOSER_1000METER	d1000m	Preference for points closer than 1000 meter.
#1#	IS_CLOSER_2000METER	d2000m	Preference for points closer than 2000 meter.
#2#	IS_CLOSER_3000METER	d3000m	Preference for points closer than 3000 meter.
#3#	IS_CLOSER_4000METER	d4000m	Preference for points closer than 4000 meter.
#4#	IS_CLOSER_5000METER	d5000m	Preference for points closer than 5000 meter.
#5#	IS_CLOSER_10000METER	d10000m	Preference for points closer than 10000 meter.
#6#	IS_ON_RIVER	river	Points that are located on a small or medium river system.
#7#	IS_ON_SAMERIVER	same	Points that are located on the same river as the current destination.
#8#	IS_ON_MAINRIVER	drain	Points that are located on a main drainage system (main river, estuary or coastal line).
#9#	IS_AWAY_FROM_ORIGIN	factory	Preference for points that are further away from the point of origin (for instance a factory or starting location) than the current destination.
#10#	IS_SAME_LONGITUDE	long	Points that are on the same longitude as the point of origin.
#11#	IS_SAME_LATITUDE	lat	Points that are on the same latitude as the point of origin.
#12#	IS_SAME_DIRECTION	samedir	Points that are in the same direction as the direction followed when heading for the current destination.
#13#	IS_ON_COAST	coast	Points that are on the coast.
#14#	IS_NOT_IN_TERRITORIUM	no territory	Not implemented yet, but intended as points that are not occupied. All preference values are effectively 0.
#R#		random	Values for all preferences above are randomly generated at the beginning of each simulation.
##		500	All preferences have a value of 500.

Table 1. Implemented Preferences and their meaning as used in SteppingIn.

agents that are autonomous, agents that are flexible in that they can learn and adapt, and a local view for each agent of the environment (Macal and North 2005). Due to these properties agents form a natural specification of the concepts in many social problems (Bankes 2002, 7199). ABM has been extensively used in ecology, biology, and the social sciences. In archaeology ABM has been used to investigate a number of research questions. Examples are the MAGICAL project with agents harvesting resources (Lake 2000), ENKIMDU simulating Mesopotamian settlement systems (Christiansen and Altaweel 2005) and the Prehistoric Patagonia model where hunter-gatherers become farmers (Barceló, del Castillo and Mameli, 2008, with in chapter 3 an overview of ABM in archaeology).

The REcursive Porous Agent Simulation Toolkit (Repast Simphony 2.0 beta version) is used as the development environment for building SteppingIn (Macal and North 2009). Repast was selected because it is open source software and has a large and active user community. Programming in Repast can be done in Java and many examples are available. The SteppingIn modelling system is implemented in a bottom-up, pattern based approach. In this approach, a pattern in the real world is observed and within the model variables and processes are included to enable (but not force) that pattern and others to emerge in the simulations (Grimm et al. 2005, 987). In the real world one can see hunter gatherers explore their territories navigating river systems, in SteppingIn both rivers and the possibility to follow these are built in. This technique is bottom-up in that at the individual level agents are given certain characteristics. The emerging system level properties and patterns are analysed. Although the design and implementation of SteppingIn do not strictly follow the standard ABM protocol of Overview, Design concepts and Details (ODD) most elements have been implemented: purpose of the model, process overview, design concepts, scheduling, state variables and input and output (Grimm et al. 2006).

2. The Model

In the spatiotemporal model of dispersal through the landscape the modern humans in SteppingIn will determine their next destination according to an attraction function. This function

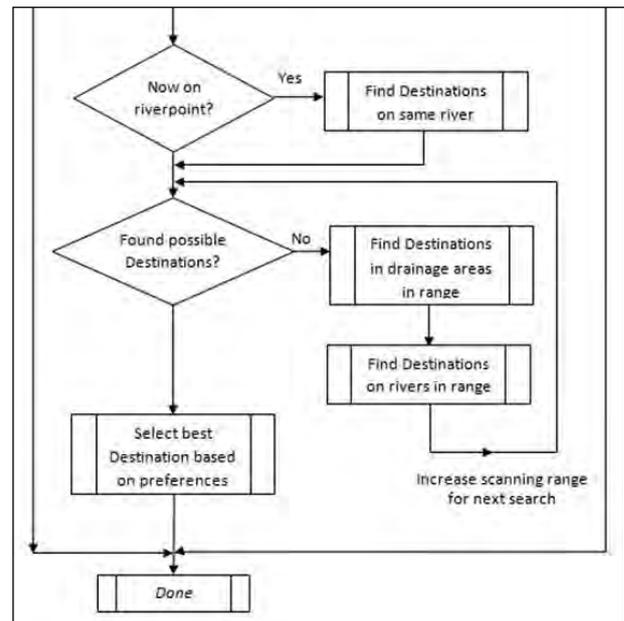


Figure 1. Decision tree to determine the next destination.

selects from all possible destination points the most attractive one as the next destination for the group. The attraction function is implemented using genetic algorithm techniques. In particular each group has an array of real numbers indicating the importance of each attraction characteristic to that group: their preferences (see Table 1 for an overview of implemented preferences). Each point has indicators for any attraction characteristic if it is implemented in that area. For instance, an attraction point located on a river has IS_ON_RIVER as true and thus for the groups with preference for rivers a higher attraction value than for other groups. The preference value of a point is the sum of all preferences that apply to that point. The selection of destinations is according to the 'ideal free distribution' model, where "colonists entering the region [...] occupy the highest-ranked patch first" (Bird and Connell 2006, 155). The modelled humans thus simply head for the most attractive point. The scanning range or searching radius that is used to locate the possible destinations can be configured. The larger the selected range the faster the colonization process. A value of 0.2 degrees or ~24 kilometres is chosen as default, but any value can be selected by the user.

The decision making process of an agent (Fig. 1) is as follows: if the current position is on a river, the group will try to follow that river scanning the

environment for points from this river. Up to four times the normal scanning range is used for this scanning process, making sure that no river points are missed. If then no possible new destinations are found, the environment is scanned for destinations in drainage areas and on other river systems. This is done in steps, increasing the scanning range each time until valid points are found. Then preference values are calculated. The destination with the highest preference value for this group is selected

as the next destination. All other destinations are flagged as *visited*, indicating that this group has seen these points. River points that have been flagged cannot be visited again, the larger drainage points can be visited twice.

The movement process itself is straightforward: the agent moves a specific distance along the calculated angle towards the next destination. Each turn each agent can move up to 35 steps of each

Locality	Coordinates	14C Age BP	OxCal 4.1 (95.4%)		Fairbanks0107 (1 σ)	
Arbreda	42° 9' 39.60"N 2° 44' 49.20"E	38307 ±552	-41772	-39993	-41813	-40799
Isturitz	43° 22' 0.12"N 1° 11' 46.00"W	37180 ±420	-40705	-39430	-40740	-39994
Willendorf	48° 19' 22.80"N 15° 23' 56.40"E	36937	-40186	-39586	-40304	-39956
Oase	45° 2' 60.00"N 21° 49' 0.01"E	34950 +990/-890	-40072	-35861	-39274	-37380
Renne	47° 35' 28.03"N 3° 45' 53.96"E	34950	-38660	-37505	-38528	-38180
Fumane	45° 37' 47.64"N 10° 55' 54.12"E	34947	-38657	-37502	-38521	-38177
Clochette	47° 07' 49.45"N 5° 34' 13.77"E	34605	-38241	-37032	-38197	-37845
Kent's Cavern	50° 28' 12.00"N 3° 31' 48.00"W	34050	-37314	-36660	-37622	-37310
Cova Gran	41° 55' 11.78"N 0° 49' 25.14"E	33061	-36531	-35205	-36666	-36360
Vogelherd	48° 33' 22.00"N 10° 11' 25.80"E	32986	-36451	-35060	-36587	-36283
Mladeč	49° 42' 0.00"N 17° 1' 0.12"E	31150 ±210	-34394	-33122	-34815	-34311
Muierii	45° 35' 39.12"N 25° 17' 36.60"E	29940 ±166	-33071	-32182	-33603	-33197
Kostenki	51° 23' 24.92"N 39° 2' 5.02"E	29882	-32861	-32536	-33479	-33219
Paviland	51° 34' 0.12"N 4° 16' 59.88"W	29133 ±137	-32540	-31345	-32797	-32403
Cioclovina	45° 41' 0.31"N 23° 4' 59.95"E	28755	-31537	-30959	-32347	-32063

OxCal: <https://c14.arch.ox.ac.uk/oxcal/OxCal.html>, after login. Result as un-modelled BC/AD year. Cairbanks: <http://radiocarbon.ldeo.columbia.edu/cgi-bin/radcarbcal>. From the results 1950 (BP) is subtracted and then the values are converted into a range so the values can be compared to the OxCal results. Links for OxCal and Fairbanks verified 24/4/2011.

Table 2. Overview of selected checkpoints with calibrated calendar ages, ordered on radiocarbon age. Years BC calculated using OxCal and Fairbanks online calibration.

100 meter, with a maximum of five destinations per turn. A moving group will thus cover an distance of around 1-5 kilometres each turn. Demographic functions (growing, reproduction, merging and dying) have been implemented in SteppingIn but have been disabled in the simulation experiments limiting the number of variables to allow efficient analysis of the results (following Grimm (1999)).

There are two other types of agents that may perform actions each turn: the Checkpoints in Space and Time (CSTs) and the Modern Human Factories. The CSTs are the control points to verify the model against archaeological data when the simulation has finished. They monitor the system by keeping a record of all groups visiting their area. Each turn they scan for groups of modern humans within scanning range (configurable but implemented as 0.2 degrees or around 24 km). Any group within range is assumed to have visited this checkpoint and is stored in the database. It is possible that a group of modern humans passes the area of a checkpoint without staying in that area. These groups are assumed not to have visited the CST. See Table 2 for an overview of implemented archaeological checkpoints and their calculated date of first arrival. Points are included based on their undisputed association with Aurignacian archaeology. When points are located geographically together (like the Aurignacian sites in the Swabian Jura) only one checkpoint is created. The other agent type is Modern Human Factory which, when included in the simulation, will produce new groups during the simulation. These will then start to explore the environment. Production of new groups is a stochastic process, mimicking a continuous migration from these points. Preferences for newly created groups are either the Factory defaults or randomized values.

Climate is not directly modelled in SteppingIn. Indirectly there are however two climate-related features that are implemented. NoGo Areas are geographical locations where no attraction points are present. Effectively this prevents the agents to go into these regions. Examples are the higher regions of the Alps, the Pyrenees and the Caucasus. During the period of interest (Oxygen Isotope Stage 3) there were also glaciers preventing access to certain areas. The NoGo Area feature is used to implement these areas but also to exclude regions where AMH were

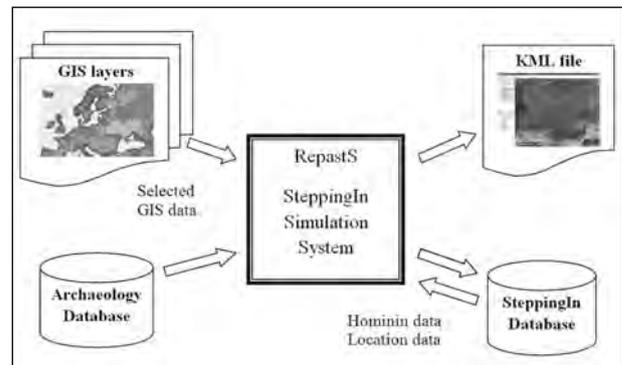


Figure 2. Data modules in the SteppingIn Simulation System.

not present, like the North African coast between Egypt and Tunisia (Close (2009), although this view is highly disputed). Some NoGo areas are implemented as *borders*. Agents crossing these borders are removed from the simulation, assumed to leave the geographical area of interest. Borders are located along the eastern edge of Europe and along the northern edge towards Scandinavia. Another climate related feature is the latitude preference for agents: one of the possible factors that define movement is preference for a specific latitude, normally the latitude of origin. This assumes that latitude influences local climate significantly (Van Andel and Davies 2003, 254; Kelly 2003, 52; but see Gamble (1995, 283)).

3. The Implemented System

The design and implementation of SteppingIn is highly modular. Both the data flow as well as the Java implementation of the programming logic. The most important data modules of the SteppingIn modelling system are shown in Figure 2. The controlling element in the SteppingIn simulation system is the Repast framework. The central object in a Repast simulation is the `Context` class, which stores agents and maintains the relationships between agents (called 'projections'). The `ContextBuilder` stores the information on the agents in the `Context`. Repast uses configuration files (`model.score` and `scenario.xml`) to specify the roles of the classes in a simulation model, to identify the `Contextbuilder` class and to define the default context (Collier and North 2010). Setting files and other configuration parameters for the Eclipse development and runtime environment include `context.xml`, `parameters.xml`, `scenario.xml`,

user_Path.xml and the SteppingIn Model.launch batch file (North et al. 2005).

The Repast environment offers most elements of a simulation system (Law and Kelton 2000, 9): system state variables, a simulation clock, an event list, initialization and timing routines, event and error routines, and different output libraries. The developer fills in these stubs and then adds the objects of interest, each in her own context and with attributes and programmed logic. The main routine must also be built and the environment for the agents constructed (type and geographical data). The following source files implement the Java classes for SteppingIn:

- SteppingIn.java – main class, initialization, scheduling
- Constants.java – global constants
- Destination.java – information per destination
- ModernHuman.java – group of modern humans (agent)
- ModernHumanContext.java – placeholder for ModernHuman class
- HumanFactory.java – the factories that can produce new groups of modern humans
- HumanFactoryContext.java – placeholder for HumanFactory class
- Checkpoint.java – the Checkpoint in Space and Time
- CheckpointContext.java – placeholder for Checkpoint class
- Geographical classes:
 - MapContext.java – placeholder for all geographical contexts
 - RiverPoint.java – the small and medium river points of interest
 - RiverPointContext.java – placeholder for RiverPoint class
 - DrainagePoint.java – the main drainage area and coast line points of interest
- DrainagePointContext.java – placeholder for DrainagePoint class
- NoGoArea.java – no go areas, POIs are removed from these areas
- NoGoAreaContext.java – placeholder for NoGoArea class

Geography in SteppingIn is implemented as Points of Interest (POIs) as opposed to an evenly spaced grid with properties for each cell. These POIs are not regularly spaced but are the nodes taken from the polygons in the original GIS layers. When converting GIS layers to fields with POIs geographical information and accuracy is lost. The scanning function for the agents is designed to partly compensate these losses. The geographical information about the environment is extracted from the ADC Worldmap data. This data is in MapInfo format and stored as points, lines and polygons. Selection of the SteppingIn POIs involves selecting only those layers and areas of interest, creating points from the polygons using Vertical Mapper, removing duplicates, assigning attributes, inserting new object locations and converting all data into shape files using the Universal Translator Tool. Currently, three types of POIs are implemented in SteppingIn:

- River points – nodes from small and medium size rivers;
- Drainage points – nodes from drainage area's like estuaries, large river systems and lakes;
- Coastal points – these are actually Drainage points with an extra `isCoastal` attribute.
- The GIS data input files for the SteppingIn system are the following:
 - EURASDNA_ptp_point.shp – the main drainage points and major rivers;
 - EURASDNLO1_ptp_point.shp – the small and medium river points;
 - EURCheckpoints.shp – the Checkpoints in Space and Time;
 - EURHumanFactories_region.shp – the modern human factories;

- EURHumans_point.shp – starting locations of modern human groups;
- EURNoGo_region.shp – no go areas, overlays from where all POIs are removed.

The results of each run are written into kml output files at specific time steps. In these files for all groups of modern humans the paths they followed are stored. Such files can be displayed using Google Earth (GE). In GE each route can be shown individually or in combination with any other route(s) at any chosen zoom level. Visually the most impressive feature of GE is the *tour* possibility. When you select an individual route it is possible to move along this path from beginning to end, at a selectable height and speed.

Analysis of ABM simulation data can be done on micro- and macrolevel. On micro level one validates the movement of the individual groups (Fig. 3), on macrolevel emerging system patterns become visible. The micro-level analysis is mostly done when building the model, to verify the correct implementation of certain preferences. Macro-level patterns become visible during and after the simulation run. Executing a run could take from a few minutes to days (depending on the number of groups, how many time steps, the number of geographical elements and the capacity of the computer). Therefore results and statistics are produced after each run in log files. Information include the execution time in milliseconds, the number of groups and the size of each group (total number of modern humans), the number of destinations visited and total length of the travelled path, the preferences before and after each run (when the learning capability is included), and a list of all the visits to all the CSTs. A separate Delphi program has been written to compile and analyse the results from multiple runs. Optional Debug information is logged at the desired level in a debug file.

4. Simulations Results

In the SteppingIn model the agents move according to individual preferences. To analyse the effect of those preferences scenarios are used. A scenario is defined as all the settings, configuration parameters and input data needed to run a single simulation. The simulation within a scenario will



Figure 3. Close-up of the Danube river valley with rainfall runoff channels in the satellite picture of Google Earth overlain by the chosen routes. Three checkpoints are visible with Cioclovina and Mueirii in the mountainous area visited by many agents, and Oase less so.

be executed several times to review the stochastic nature of the model. Each simulation is time based, meaning the simulation will run for a set time frame that is configurable but per default set at 20.000 turns, as by that time the larger patterns have become visible and recognizable. A scenario has been set up to simulate modern humans coming into Europe from the Black Sea, the Singleton Danube scenario, which serves to test the usefulness of the SteppingIn simulation system in the context of the colonization of Europe.

All groups start from the same single geographical location in the Danube estuary. Each modern human group in this scenario is named Dan#?#, with the question mark represents values from {0-14, [], R}. In the Singleton Danube scenario some features are turned off to be able to focus on the effects of the preferences only: territories, factories, landscape learning, merging, dying and generating offspring are all disabled. For the Singleton Danube scenario the following geographical configuration is used:

- RiverPointContext: 865102 River points
- DrainagePointContext: 595425 Drainage points
- Checkpoint Context: 15 Checkpoints
- NoGoAreaContext: 9 NoGo areas
- drainage points: 413764 removed
- river points: 251101 removed
- ModernHumanContext: 17 agents

The Singleton scenarios get their names from to the fact that all modern human groups start the simulation from the same geographical location. The agents are specialized, meaning that they have only one preference set to maximum and all others set to zero (see Table 1). Exceptions are the #R# group, with random values for all preferences, and the ## group, with a set value of 500 for each preference. These two groups could be designated as being generalists. Also #14# can be called a generalist, because her preference (to avoid territories) is not implemented in this scenario and thus this group selects each time a new destination completely at random, with no specific preference at all. Having all values of zero is different from all values 500 as most points satisfy multiple preferences and the final attraction value is the sum of all satisfied preference values.

Eight simulations were run in the Singleton Danube scenario. The results are presented in the following two tables. In Table 3 for each run the number of visits per checkpoint and the names of the visiting groups are given. Table 4 presents the chronology of these runs. For each group is indicated if and when checkpoints are visited. As the checkpoint columns are given in chronological order this table is used to compare the simulation results with the archaeological data. These simulation results will be discussed in the next section.

5. Discussion

The SteppingIn application not only implements a model of the early modern humans who arrived in Europe, it also allows parameterisation of that model. This effectively implements a range of models. For instance, variation of the scanning

#	CST	-1-	-2-	-3-	-4-	-5-	-6-	-7-	-8-
9	Mladec	#13#, #5#	#2#, #4#	#6#	#8#		#13#		#11#, #13#
10	Fumane	#11#, #7#	#5#	#6#	#13#	#8#	#9#		# 1 4 # , #11#, #2#
6	Renne	#11#, #7#, #R#					#2#	#R#	#R#
5	Cova Gran	##, #R#				#8#		#R#	#R#
15	Oase	#13#, #1#	#10#	#2#, #11#, #8#, #6#	#8#	#7#	#9#	#R#	#11#, #10#, #4#, #7#
11	Willendorf	##, #1#	#4#	#6#		#8#	#2#	#5#, #R#	# 1 0 # , #13#, #R#
10	Isturitz	##, #R#		#6#	##, #0#	#8#	#2#, #9#	#R#	#R#
7	Vogelherd	##	#5#			#8#	#2#	#R#	#14#, #2#
5	Kents Cave	##, #R#			##		#2#	#R#	
31	Muierii	#11#, #7#, #13#, #5#, #6#, #1#	#3#, #11#, #4#, #7#, #6#	#2#, #11#, #10#	#8#, #0#, #1#	#6#	#10#, #4#, #5#, #1#	#3#, #11#, #1#, #R#	#2#, #11#, #10#, #8#, #7#
27	Cioclovina	#8#, #7#, #13#, #5#, #1#	#11#, #6#	#3#, #2#, #11#, #10#, #8#, #6#	#0#	#14#, #4#, #6#, #1#	#11#, #4#, #8#	#2#, #5#, #R#	# 1 1 # , #10#, #4#
5	Kostenki			#14#		#14#		#10#, #13#	#5#
5	Paviland	##, #R#			##		#2#	#R#	
10	Clochette	#11#, #7#	#5#		#13#	#8#	#2#		#14#, #2#, #11#, #R#
3	Arbreda	#R#				#8#		#R#	

Table 3. Visiting agents per checkpoint and per simulation run. The first column is the total number of times a checkpoint is visited in all eight runs combined.

range used by the agents to detect favourable points of interest effectively implements a different model for each chosen value. Sensitivity analysis of the output parameters (in this case visited CSTs and the geographical spread throughout Europe) to the chosen input values can be part of a validation process of a model (Pitt, Box and Knowlton 2003, 114-117). For SteppingIn such a rigorous analysis has not yet been attempted and instead a modest statistical analysis of the results is presented here. A 'valid' model would have no internal inconsistencies and have no flaws (errors), but 'verification' of an open model like SteppingIn is impossible (Malleon et al. 2010; Oreskes, Shrader-Frechette, and Belitz 1994, 642).

In SteppingIn the groups move through the landscape using their individual preferences to select the next destination. Circumventing large mountain chains like the Alps, the Pyrenees or the Caucasus poses no problem for the simulated humans and islands like Cyprus, Sardinia or the United Kingdom are visited in most simulation runs. Larger water bodies are also crossed, for instance from Sicily to Tunisia, among the Greek Islands or over to Ireland.

And Gibraltar is regularly used to colonize Morocco, but also several times the other way around, visiting Spain from the south. Interestingly the middle of Germany remains sparse with few traces of humans. Inspection of these sparse areas reveals no special circumstances with water resources abundant.

Analysing the chosen geographical routes one of the most eye-catching emerging patterns is the general tendency to head north. Many preferences have groups end up against the glacier ice covering northern Europe. An explanation for this appears upon closer inspection. For instance the character of the Carpathian Mountains in Romania creates a south-north oriented valley system near the Danube River (see Figure 3 which is also a good illustration of why certain sites are more often visited than others). Through these valleys water runs downhill. The Carpathians have more or less the same orientation as the Alps and the other major mountains ranges, the Pyrenees and the Caucasus. They are caused by the same phenomenon, namely Africa tectonically moving into the Eurasian plate. This pushes the mountains up and creates an east-west orientation of the ranges. Together they create many north-

	Arbre.	Isturitz	Willen.	Oase	Renne	Fuma	Cloch.	Kents	C. Gran	Vogel.	Mladec	Muierii	Kost.	Pavila.	Ciocio.	#	C	S
	38,3ka	37,2ka	36,9ka	35,0ka	35,0ka	34,9ka	34,6ka	34,1ka	33,1ka	33,0ka	31,2ka	29,9ka	29,9ka	29,1ka	28,8ka			
#14#						4256	8237			11313			4281		919	5	4	3
#13#			6387	17130		12155	13501				6089	2027	9825		18559	8	14	3
#12#																0	0	0
#11#				2172	15457	10736	11870				4829	305			540	7	6	2
#10#			18858	1411								1259	4678		8100	5	5	3
#9#		2107		563		19568										3	2	2
#8#	7905	9140	1016	517		4736	7000		8113	18160	5664	372			757	11	23	5
#7#				4395	8475	6835	10039					1143			1277	6	6	2
#6#		12862	9532	7865		10354					9064	1019			1018	7	2	2
#5#			1925			2945	19251			19845	7067	2181	3260		2240	8	14	4
#4#			13031	8062							14799	1859			2980	5	3	2
#3#												688			2127	2	1	2
#2#		19878	4480	6474	7871	8209	8717	11604		5753	3305	2240		15492	1984	12	25	6
#1#			10619	7696								571			604	4	1	2
#0#		19814										2879			3450	3	1	2
#R#	11328	9748	719	358	2044		18712	3843	10370	1091		139		4109	292	12	24	3
##		11511	3483					15992	15742	4381				15556		6	8	2

Table 4. The chronology in SteppingIn. Given is the first simulation turn any one group with a specific preference arrives at any CST. The column '#’ contains a count of these visits, the ‘C’ column the number of visits that are in chronological order and the ‘S’ column has the maximum series of visits in chronological order.

south oriented river systems. Following the river is one of the characteristics of the modelled humans in SteppingIn, so when they follow these river systems some head south towards the Mediterranean but eventually most head north. Other causes for this orientation are the watering systems from (earlier) glaciers systems and (further north) the physical traces of the movement of those glaciers in the landscape.

When looking at the chosen routes for the individual preferences it becomes clear that every so often the stochastic circumstances allow large distances in Europe to be covered (colonized) in a short time (for preferences Dan#R#, Dan#0#, Dan#2#, Dan#5# or Dan#9# one or more groups head far west and reach Morocco, Spain or England within the given time frame). These results support the observation of Nikitas and Nikita (2005, 602-603) that *'simulation models [...] estimate the probability of the earliest distribution of hominins throughout Eurasia'* but cannot *'give final answers to what exactly happened'*. What exactly had happened *could* be one of the possible outcomes of the model.

In SteppingIn, chronological data is compared to the compiled data from the visits to the CSTs to assess the validity of the implemented model (Young 2002, 157). The first visit of each agent type to any CST is collected for all eight simulations and presented in Table 4. For example, the earliest moment in any simulation that a group with preference #13# (prefers coastal routes) reaches Willendorf is after 6387 turns. To compare this timing information with the archaeological data, the uncalibrated ¹⁴C dates as given in Table 2 are used. The columns are ordered chronologically, with the oldest (Arbreda) first. Three statistical variables are used: the total number of visits, the total number of visits that are in chronological order, and the longest series of visits that are in chronological order (the last three columns in Table 4). An example: #14# visits five CSTs, Fumane (F) at timestep 4256, Chlochette (Ch) at 8237, Vogelherd (V) 11313, Kostenki (K) 4281 and Cioclovina (Ci) at 919. The archaeologically estimated date of F is chronologically before Ch, V and K, and Ch is dated before V giving a total of four chronologically ordered visits. Of these, F, Ch and V form a consecutive series of three, the value in the last column.

Using these three variables, three preferences seem best to match the archaeological order: #8# (preference for main drainage systems), #2# (preference for new destinations found closer than 2000m) and #R# (initialized with random values for all preferences). Of these, #8# and #2# have series of 5 and 6 ordered visits respectively, where #R# has a maximum length of 3 ordered visits. The large number of chronologically ordered visits for #R# (24) is caused by visits to some of the oldest sites early on (Willendorf and Oase). #R# matches the archaeological order less than #8# and #2# based on the length of chronological sequences, especially since there are quite a few others with series of 3 or even 4. Interestingly, the randomly generated values for the preferences of three groups in #R# that visited most CSTs included large values for drain, longitude, d2000(!) and river. The exact influence of these specific combinations of preferences for #R# is not intuitive and needs further research. Also, the uncertainties in ¹⁴C dates are such that great care must be taken when ordering these dates, especially when dates near the maximum range are involved (Bocquet-Appel and Demars 2000).

During the development of the SteppingIn model it appeared that the scanning range which the agents use to search for new destinations has major influence on the distance (and thus the number of turns) it takes to move through an area. In other words, with a scanning range that is doubled Europe is colonized twice as fast, although not as thoroughly. This can be explained by the fact that when more destination points in an area are visited, the path taken will be longer and traversal will take more turns. It would be interesting to compare this pattern to the theory about moving distance, foraging range and logistic radius (Binford 1982, 10; Grove 2009, 222). Binford has shown that the logistic radius is largely determined by the contribution of hunting to the diet of the group and the dependence of logistic mobility. Logistic radius is not included in SteppingIn, and Palaeolithic information on diet and logistic mobility is not abundant. Foraging radius is defined by habitat quality, and that is a statistic that can be included in a future version of the model.

6. Conclusions

SteppingIn was built to create a framework in which real geographical information on a large scale

is used to model dispersals that happened in the distant past. One of the more important dispersals in archaeology is the colonization of Europe by modern humans. SteppingIn uses this major event in human history as a test case to model migration movements. The SteppingIn framework uses agent-based modelling as a technique and Repast Symphony as an implementation tool. Agents move from one point of interest to the next, with water resources converted into points of interest. From the archaeology it is known that 60kya there were no modern humans in Europe and that around 30kya modern humans were basically everywhere. SteppingIn models this colonization.

The implemented Danube corridor test case shows that the implemented framework is a powerful modelling system, able to simulate real world events. It is also a flexible system to which extensions like colonisation modes, additional preferences, or demographic features are easily added. Caution must be taken when adding too much additional functionality in that the results must still represent the intended reality and that the resulting system does not become too complex to analyse. In SteppingIn it is therefore possible to turn specific functionality *off*. Because part of the output is in kml format the analytic power of Google Earth is included in the SteppingIn framework. This enables close observation of patterns, both at the micro level (to validate individual group movement) as at the macro level (to identify global colonization patterns).

From the discussion of the results it appears that simple behaviours like following rivers can indeed generate population distributions and dispersals patterns similar to those observed in the archaeological record. Especially the results for preferences DAN#2#, which implements searching for new destinations in the scanning range of 2.000 metres, and DAN#8#, heading for major drainage points, appear to match the archaeological data. But the stochastic nature of the implementation plays an important part in route selection, as the results for DAN#R# illustrate. Also an interesting result is the fact that many groups head north due to the specific geographical structure of Europe. Combined these results suggest that following the rivers might not have been the only motivation of modern humans moving into Europe.

References

- Anikovich, M.V., A.A. Sinitsyn, J.F. Hoffecker, V.T. Holliday, V.V. Popov, S.N. Lisitsyn, S.L. Forman, G.M. Levkovskaya, G.A. Pospelova, I.E. Kuz'mina, N.D. Burova, P. Goldberg, R.I. Macphail, B. Giaccio, and N.D. Praslov. 2007. "Early Upper Paleolithic in Eastern Europe and Implications for the Dispersal of Modern Humans." *Science* 315(5809):223-226.
- Bailey, G., J.S. Carrión, D.A. Fa, C. Finlayson, G. Finlayson, and J. Rodríguez-Vidal. 2008. "The coastal shelf of the Mediterranean and beyond: Corridor and refugium for human populations in the Pleistocene." *Quaternary Science Reviews* 27: 2095-2099.
- Bailey, S.E., T.D. Weaver, and J.-J. Hublin. 2009. "Who made the Aurignacian and other early Upper Paleolithic industries?" *Journal of Human Evolution* 57: 11-26.
- Bankes, S.C. 2002. "Agent-based modelling: A revolution?" *Proceedings of the National Academy of Sciences of the United States of America* 99(3): 7199-7200.
- Barceló, J.A., F. del Castillo, and L. Mameli. 2008. "The Computer Simulation of Social Dynamics and Historical Evolution - The case of "Prehistoric" Patagonia." Paper presented at the *II Jornadas sobre Simulación Social y Análisis de Sociedades Artificiales Barcelona, 20-21 November 2008*.
- Bertuzzo, E., A. Maritan, M. Gatto, I. Rodriguez-Iturbe, and A. Rinaldo. 2007. "River networks and ecological corridors: Reactive transport on fractals, migration fronts, hydrochory." *Water Resources Research* 43: 1-12.
- Binford, L. R. 1982. "The Archaeology of Place." *Journal of Anthropological Archaeology* 1: 5-31.
- Bird, D.W., and J.F. O'Connell. 2006. "Behavioral Ecology and Archaeology." *Journal of Archaeological Research* 14(2): 143-188.
- Bocquet-Appel, J.-P., and P.Y. Demars. 2000. "Neanderthal contraction and modern human colonization in Europe." *Antiquity* 74: 544-552.
- Bonabeau, E. 2010. "Agent-based modeling: Methods and techniques for simulating human systems." *Proceedings of the National Academy of Sciences of the United States of America* 99(3):7280-7287.

- Christiansen, J.H., and M.R. Altaweel. 2005. "Understanding Ancient Societies: A New Approach Using Agent-Based Holistic Modeling." *Structure and Dynamics: eJournal of Anthropological and Related Sciences* 1(2): 1-39.
- Close, A.E. 2009. "The Middle-Upper Palaeolithic hiatus of insular north Africa." In *The Mediterranean from 50 000 to 25 000 BP: Turning points and new directions* edited by Marta Camps and Carolyn Szmidi, 35-50. Oxford: Oxbow Books.
- Collier, N., and M. North. 2010. "Repast Java getting started." Accessed April 2011. <http://repast.sourceforge.net/docs/RepastJavaGettingStarted.pdf>.
- Conard, N.J., and M. Bolus. 2003. "Radiocarbon dating the appearance of modern humans and timing of cultural innovations in Europe: new results and new challenges." *Journal of Human Evolution* 44: 331-71.
- Davison, K., P. Dolukhanov, G. R. Sarson, and A. Shukurov. 2006. "The role of waterways in the spread of the Neolithic." *Journal of Archaeological Science* 33: 641-652.
- Falconer, S.E. 2005. *A First Generation Model of Agricultural Landuse*. Tempe: Arizona State University.
- Gamble, C.S. 1995. "The earliest occupation of Europe: the environmental background." In *The Earliest Occupation of Europe: Proceedings of the European Science Foundation workshop at Tautavel (France), 1993*, edited by W. Roebroeks and T. van Kolfschoten, 279-296. Leiden: *Analecta Praehistorica Leidensia* 27.
- Grimm, V. 1999. "Ten years of individual-based modelling in ecology: what have we learned and what could we learn in the future?" *Ecological Modelling* 115: 129-148.
- Grimm, V., E. Revilla, U. Berger, F. Jeltsch, W. M. Mooij, S.F. Railsback, H.-H. Thulke, J. Weiner, T. Wiegand, and D.L. DeAngelis. 2005. "Pattern-oriented modeling of agent-based complex systems: lessons from ecology." *Science* 310(5750): 987-991.
- Grimm, V., U. Berger, F. Bastiansen, S. Eliassen, V. Ginot, J. Giske, J. Goss-Custard, T. Grand, S.K. Heinz, G. Huse, A. Huth, J.U. Jepsen, C. Jørgensen, W.M. Mooij, B. Müller, G. Pe'er, C. Piou, S.F. Railsback, A.M. Robbins, M.M. Robbins, E. Rossmanith, N. Rüger, E. Strand, S. Souissi, R.A. Stillman, R. Vabø, U. Visser, D.L. DeAngelis. 2006. "A standard protocol for describing individual-based and agent-based models." *Ecological Modelling* 198: 115-296.
- Grove, M. 2009. "Hunter-gatherer movement patterns: Causes and constraints." *Journal of Anthropological Archaeology* 28(2): 222-233.
- Hoffecker, J.F. 2009. "The spread of modern humans in Europe." *Proceedings of the National Academy of Sciences of the United States of America* 106(38): 16040-16045.
- Lake, M.W. 2000. "MAGICAL computer simulation of Mesolithic foraging." In *Dynamics in Human and Primate Societies: Agent-Based Modelling of Social and Spatial Processes*, edited by T.A. Kohler and G.J. Gumerman, 107-143. New York: Oxford University Press.
- Law, A.M., and W.D. Kelton. 2000. *Simulation Modeling and Analysis, third edition*. Boston: The McGraw-Hill Companies, inc.
- Macal, C.M., and M.J. North. 2005. "Tutorial on Agent-based Modeling and Simulation." In *Proceedings of the 2005 Winter Simulation conference*, edited by M.E. Kuhl, N.M. Steiger, F.B. Armstrong and J.A. Joines, 2-15. Orlando: Winter Simulation Conference.
- Macal, C.M., and M.J. North. 2009. "Agent-based Modeling and Simulation." In *Proceedings of the 2009 Winter Simulation Conference*, edited by M.D. Rossetti, R.R. Hill, B. Johansson, A. Dunkin and R.G. Ingalls, 86-98. Austin: Winter Simulation Conference.
- Malleson, N., A. Evens, A. Heppenstall, and L. See. 2010. "Evaluating an Agent-Based Model of Burglary." *Working Paper* 10/1. School of Geography, University of Leeds, LS2 9JT. Accessed 4 February 2012. http://www.geog.leeds.ac.uk/fileadmin/documents/research/csap/10_1.pdf.
- Mellars, P. 2006. "Archeology and the Dispersal of Modern Humans in Europe: Deconstructing the 'Aurignacian'." *Evolutionary Anthropology* 15: 167-182.
- Mithen, S.J., and M. Reed. 2002. "Stepping Out: A computer simulation of hominid dispersal from Africa." *Journal of Human Evolution* 43: 433-462.
- Nigst, P.R. 2006. "The first modern humans in the Middle Danube Area? New Evidence from Willendorf II (Eastern

SteppingIn - Modern Humans Moving into Europe - Implementation
Fulco Scherjon

- Austria)." In *When Neanderthals and Modern Humans Met*, edited by N. J. Conard, 269-304. Tuebingen: Kerns Verlag.
- Nikitas, P., and E. Nikita. 2005. "A study of hominin dispersal out of Africa using computer simulations." *Journal of Human Evolution* 49: 602-617.
- North, M.J., T.R. Howe, N.T. Collier, and J.R. Vos. 2005. "The repast symphony development environment." In *Agent 2005 Conference on Generative Social Processes, Models, and Mechanisms*, Argonne National Laboratory, Argonne, IL, USA, edited by C.M. Macal, M.J. North and D.L. Sallach, 159-166. Chicago, IL, USA: Argonne National Laboratory.
- Oreskes, N., K. Shrader-Frechette, and K. Belitz. 1994. "Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences." *Science* 263(5147): 641-6.
- Pitt, W.C., P.W. Box, and F.F. Knowlton. 2003. "An individual-based model of canid populations: modelling territoriality and social structure." *Ecological Modelling* 166: 109-121.
- Roe, D.A. 2009. Foreword in *The Mediterranean from 50 000 to 25 000 BP: Turning points and new directions*, edited by Marta Camps and Carolyn Szmids, v-x. Oxford: Oxbow Books.
- Straus, L.G. 2001. "Africa and Iberia in the Pleistocene." *Quaternary International* 75: 91-102.
- Steele, J., J. Adams, and T. Slukin. 1998. "Modelling Paleoindian dispersals." *World Archaeology* 30: 286-305.
- Van Andel, T.H., and W. Davies. 2003. *Neanderthals and Modern Humans in the European Landscape during the Last Glaciation: Archaeological Results of the Stage 3 Project*. Cambridge: Cambridge University Press.
- Verpoorte, A. 2005. "The first modern humans in Europe? A closer look at the dating evidence from the Swabian Jura (Germany)." *Antiquity* 79: 269-279.
- Villa, P. 2001. "Early Italy and the colonization of Western Europe." *Quaternary International* 75: 113-130.
- Young, D. A. 2002. "A New Space-Time computer Simulation Method for Human Migration." *American Anthropologist* 104(1): 138-158.

Geospatial Virtual Heritage: A Gesture-Based 3D GIS to Engage the Public with Ancient Maya Archaeology

Heather Richards-Rissetto

Bruno Kessler Foundation (FBK), Trento, Italy and University of New Mexico, Albuquerque, NM, USA

Jim Robertsson

Umeå University, Umeå, Sweden

Jennifer von Schwerin

German Archaeological Institute, Bonn, Germany

Giorgio Agugiaro, Fabio Remondino

Bruno Kessler Foundation, Trento, Italy

Gabrio Girardi

Graphitech Foundation, Trento, Italy

Abstract:

This paper presents our research to develop a gesture-based 3D GIS system to engage the public in cultural heritage. It compares two types of interaction—device-based vs. natural interaction—and summarizes the beta-testing results of a 3D GIS tool for archaeology, called QueryArch3D, in which participants used device-based interaction (i.e. mouse and keyboard). It follows with a description of the gesture-based system—that we developed in response to these beta-tests. The system uses QueryArch3D and Microsoft’s Kinect to enable people use body movements (in lieu of keyboard or mouse) to navigate a virtual reality landscape, query 3D objects, and call up photos, videos, and text. The case study is the ancient Maya city of Copan—today an UNESCO World Heritage site in Honduras. The low cost and portable system is ideal for museum exhibitions even in developing countries and can be adapted for archaeological content throughout the world.

Keywords:

3D GIS, Kinect, Virtual Heritage, Public Engagement, Ancient Maya

1. Introduction¹

Cultural heritage plays an important role in understanding and shaping our past, present, and future and influences identity, community, and political and social processes. For these reasons, among others, it is important to engage and educate the public about the past. Cultural heritage includes both the tangible (e.g. artefacts, architecture, and landscapes) and intangible (e.g. cosmology, folklore, and oral histories) expressions of human culture (Vecco 2010). There are many ways to engage the public in cultural heritage, including on-site visits to monuments and archaeological sites, videos, documentaries, books/articles, songs, games, and re-enactments as well as websites, blogs, and even virtual worlds. In the last two decades,

Corresponding author: hmrichards@fbk.eu

virtual heritage—the application of computer visualisation, virtual reality (VR), and information and communication technologies (ICT) to cultural heritage—is increasingly commonplace (Forte and Siliotti 1997).

Virtual heritage ranges from virtual displays at traditional museums to web-based applications (including online museums) to mobile devices (Ioannides et al. 2010). These different contexts include diverse media such as images, movies, maps, and 3D models and use a variety of visualisation and interaction tools to engage and educate the public (Beusing and Posluschny 2009, Pescarin et al. 2012; Pujol and Lorente 2012; Virtual Museum Transnational Network 2011). In order to reach broader audiences and to foster respect for cultural diversity, archaeologists, cultural resource

managers, and museum curators increasingly use digital tools (Dawson, Levy, and Lyons 2011; de Byl and Khan 2011; Fischer 2012). In 2009, the MayaArch3D Project began to develop a prototype 3D GIS tool—called QueryArch3D—that integrates Geographic Information Systems (GIS), multi-resolution 3D models, and VR for archaeological research in the ancient Americas (Agugiaro et al. 2011). In 2011, project members began to collaborate with the HUMlab, Umeå University in Sweden (HUMlab 2012) to develop a gesture-based system using the QueryArch3D tool and Microsoft’s Kinect that could be used to create a virtual heritage exhibit to engage the public in ancient Maya archaeology. The system is low-cost and portable and can easily be set-up in classrooms, community centres, or traditional museums. The exhibit uses two types of interaction technologies and two levels of immersion to appeal to different audiences and learning styles.

2. Interactivity and Immersion

The earliest applications of virtual heritage were often non-interactive with users simply watching a video or looking at static text or photos (Styliani et al. 2009); however, the use of interactive technologies to promote cultural heritage has been increasing for the past two decades. Interactive technologies “demand” users to observe, perceive, and manipulate objects and/or their actions through their own decisions and subsequent movements, offering a “more natural and instinctive way of learning” than non-interactive technologies (Carrozzino and Bergamasco 2010, 452).

Interaction technologies are often classified as either device-based or natural. The majority of virtual heritage applications employ device-based interaction systems that typically use a mouse, joystick, keyboard, or touch-screen to interact with content. In contrast, natural interaction systems use gesture-based or speech-based interaction systems, manipulating content with body movements or speech. Using gestures seems to create a sense of embodiment and spatial awareness that provides users with a better sense of space and consequently a better sense of place. Generally speaking, gesture-based navigation offers a different perspective or frame of reference (from device-based navigation) to explore cultural heritage.

Related to interaction technology is the level of immersion, defined here as the degree to which users feel part of a simulated environment. Three levels of immersion have been defined by Carrozzino and Bergamasco (2010): Non-Immersive, Low Immersion, and High Immersion. *Non-immersive* applications communicate cultural heritage content using, for example, a desktop/laptop computer or mobile devices (e.g. smart phones or touchpads). *Low immersion* systems include hand-immersive workbenches that use haptic hardware (such as gloves or glasses) and stereo display technologies to enable users to “see” and “touch” virtual 3D objects. They also include Augmented Reality systems that use head-mounted or handheld devices to modify and enhance real-world environments by overlaying digital information (such as videos, 3D graphics, etc.) on the real world (Noh, Sunar, and Pan 2009). Such low immersion systems, however, do not permit users to employ a full-range of body movements to control interaction with content, and they do not fully immerse users in their surroundings. *High immersion* systems such as a CAVE (cave automatic virtual environment), however, combine stereoscopic projection (on four to six surfaces), 3D computer graphics, and 3D sound to create a sense of full-body presence in a virtual environment (Acevedo et al. 2001). One example is the collaborative system for Teleimmersive Archaeology developed by the University of California, Merced, and the University of California, Berkeley that allows users, as 3D avatars, to interact in the same cyberspace with models of artefacts, monuments and sites (Forte and Kurillo 2010).

2.1 Virtual heritage in archaeology: some advantages

While virtual heritage content is not limited to the past, this paper focuses on archaeological applications. The use of computer visualisation, virtual reality (VR), and information and communication technologies (ICT) is ideally-suited to engaging the public in archaeology. Some reasons include:

- Web-based applications can reach broader and more diverse audiences than books or even films;
- Multi-media exhibits (museum, web-based, and others) comprising digital photos, maps, and



Figure 1. The Etruscan Tombs-Panoramic Virtual Tours Exhibit illustrating use of diverse 3D technologies in a single exhibit; (left) using stereo glasses to view reality-based 3D models of artefacts; (middle) photo panoramas of modern landscape; (right) interactive, reality-based 3D reconstruction of Etruscan tomb (Photos courtesy of 3DOM, FBK).

videos as well as 3D visualisations allow people to virtually “visit” archaeological sites—something rarely possible for most people;

- Museum exhibitions have space constraints and can display only a limited number of objects—virtual exhibits can store and display many more objects;
- Archaeological materials such as ceramics, bone tools, and textiles are often too fragile or delicate to put on public display, while digital replicas are not;
- Archaeological sites do not appear as they did in the past—overlaying 2D drawings and 3D reconstructions onto reality-based images (of sites as they look today) helps people understand how sites looked in the past;
- Digital tools can be used to turn on and off temporal layers to provide a sense of change through time;
- Remote-sensing technologies such as photogrammetry and laser scanning can be used to capture reality-based 3D data to create 3D digital objects that the public can virtually explore (Remondino 2011);

VR can simulate ancient cities and their landscapes to provide context for narratives creating a sense of space and place for ancient cultures (Barcelo, Forte, and Sanders 2000; Forte and Siliotti 1997; Frischer and Hild 2008).

In sum, virtual heritage has the potential to reach broad audiences, use multimedia to communicate archaeological information to address different learning styles, and facilitate interactivity allowing users to acquire and create knowledge as they take the lead in learning (Pantano and Tavernise 2009, Roussou 2008).

2.2 Virtual heritage in archaeology: device-based interaction vs. natural interaction

The use of 3D technologies and visualisations for cultural heritage education is rapidly increasing and the ways in which people are interacting with virtual heritage are changing as multimedia is integrated with new, diverse technologies. One example is Colonial Williamsburg, a living museum in Virginia (USA), which uses a range of digital techniques and tools to engage the public in cultural heritage: podcasts, 3D visualisations, videos, online exhibits, onsite games with PDAs, etc. (Fischer 2012). Another example is the Etruscan Project of FBK Trento, Italy where reality-based 3D data are

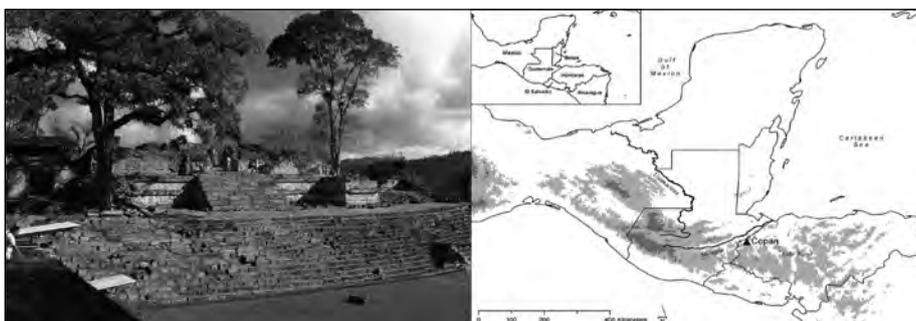


Figure 2. Temple 22 in main civic-ceremonial complex at Copan, Honduras (left). Map showing location of Copan, Honduras (right).

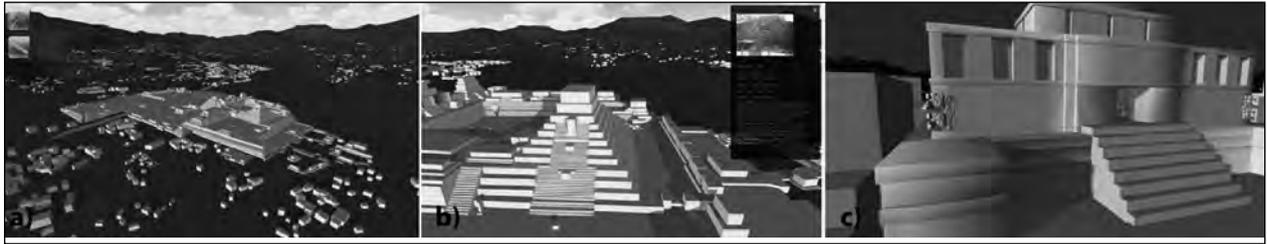


Figure 3. Views of *QueryArch3D* tool. a) aerial view with LoD1 models b), Queries on geometric models) and c) LoD3 with interior walls/rooms and some simplified reality-based 3D elements.

used as multimedia material to exhibit heritage that is not fully visible and accessible to the public (Remondino et al. 2011). Users can move through panoramas of the present-day landscape, observe 360 degree views at set points, and enter and explore several frescoed tombs and use an Augmented Reality platform to interact with digitized artefacts located in other museums (Fig. 1).

All of the above-mentioned examples offer device-based interaction—the norm in virtual heritage. This is because the technology necessary for systems that use natural interaction (gesture or speech-based systems) was expensive and required expertise in computer programming. For the past few years, however, since the release of Microsoft's Kinect (in November 2010) and other low-cost multi-purpose sensors (e.g. Axu Xtion) that allow touchless remote control to interact with a computer (Lange et al. 2001, Suma et al. 2011) applications that use natural interaction are increasing. Archaeologists are exploring their potential for virtual heritage applications (Abate et al. 2012, Giles 2010, Villaroman, Rowe, and Swan 2011). One example is the *Aquae Patavinae* Project, which uses a range of interaction technologies including a web camera for gesture-based interaction in a virtual landscape of *Aquae Patavinae*, Italy (Aquae Patavinae 2012). Both device-based and natural interaction technologies offer important opportunities to engage the public in virtual heritage.

3. MayaArch3D Project: The Case Study at Copan, Honduras

The *MayaArch3D* Project is exploring both device-based and natural systems for education and research on the ancient Maya city of Copan, Honduras (*MayaArch3D* 2012, von Schwerin et al. 2011). Copan, located on the southern periphery of the Maya civilization (600 B.C. - A.D. 1521),

was an important centre for commercial and cultural exchange (Fig. 2). Copan is one of the most thoroughly excavated Maya sites, and its long history of excavation and research provides voluminous data that form the basis for many interpretations about the ancient Maya (Andrews and Fash 2005). In 1980, Copan became a UNESCO World Heritage Site and is one of the top tourism destinations in Honduras for international visitors. The more than 150,000 visitors each year to the archaeological park can sense the city's past grandeur; however, because much of the ancient city does not fall within park boundaries, it is difficult for visitor to imagine Copan's size and complexity, and for cultural heritage managers it is difficult to convey this information through traditional means. To convey a sense of the city's spaces to people who never visit the site, is even more difficult. A promising avenue to provide both onsite and virtual visitors a sense of the scale and size of the city as well as engage visitors in the archaeology of the ancient Maya are interactive 3D visualisations.

To explore the possibilities of 3D digital tools and GIS for archaeological and art historical research on ancient Maya architecture and landscapes, the *MayaArch3D* project began in 2009 to bring together art historians and archaeologists with computer scientists and remote sensing specialists to develop a prototype 3D WebGIS tool called *QueryArch3D* (Agugiaro et al. 2011, von Schwerin et al. 2011). *QueryArch3D* is a visualisation and query tool that links multi-resolution 3D models to archaeological data that are accessed while navigating in a VR environment. The prototype tool contains a virtual landscape of Copan that covers 24 km² and contains 3D schematic models of over 3,000 ancient structures. It also contains a 3D Studio Max model of an 8th century temple, and reality-based models of sculptures and stelae linked to archaeological data (Fig. 3). The prototype integrates three technologies:

1) Geographic Information Systems (GIS), 2) three-dimensional digital models, and 3) virtual reality.

3.1 The QueryArch3D tool: benefits of 3D GIS to archaeology

By connecting 3D models in a VR environment with a GIS (i.e. spatial database), 3D GIS—of which the QueryArch3D tool is one—offer a number of benefits to archaeologists and also museum visitors. The advantage of such a system is that users can move through a virtual landscape and query and interact with 3D models of objects and their associated archaeological data. This new capability allows users to actively and more intuitively engage with archaeological information, and thereby generate new knowledge. To go briefly into more detail:

This system has the ability to provide the user with access to different types of 3D models of individual objects or whole archaeological sites via an interactive, virtual environment. The advantages include:

- access to virtual artefacts when original artefacts cannot be transported or handled due to fear of damage;
- a sense of mass and space;
- analysis of objects from multiple perspectives (e.g. bird's eye, exterior vs. interior view, anterior vs. posterior view);
- interaction with reality-based 3D models—derived from technologies such as laser scanning and photogrammetry—that illustrate what can be seen today of an archaeological site or individual artefacts (Manferdini and Remondino 2012);
- interaction with 3D models generated using computer graphics (based on historical and archaeological data) that provide alternative reconstructions of how individual buildings, architectural complexes, and entire cities may have looked in the past (and at different time periods), which is useful for archaeological investigations and to educate the public about archaeological methods and theories (Sanders 2008);

- comparisons of data and hypothetical reconstructions;

- metric and quantitative analysis.

Because this system links these 3D models with a GIS, users also have the capability to overlay and link datasets, and therefore a range of new ways to interact with archaeological data. While archaeologists can use the database back-end feature of the tool to bring together buildings, sculpture, and artefacts into an integrated landscape to reveal relationships, patterns, and trends that are not evident when using traditional (non-spatial) databases (Conolly and Lake 2006, Wheatley and Gillings 2002), a more simple interface is offered to the public so that they can interact with the data connected to the 3D models in the following ways:

- search using pre-set queries such as “show me all excavated temples”;
- turn layers of information on/off, for example, vegetation layers to illustrate different paleo-environmental interpretations;
- use time sliders to visualise changes through time, for example, the construction campaigns of different rulers (Zerneke et al. 2006);
- click on 3D models to access text, 2D media such as photographs and maps, or videos (Huk, Steinke and Floto 2010).

Overall, the most important benefit of a 3D GIS for museum visitors is that the QueryArch3D tool can situate 3D models of objects or buildings in their larger spatial contexts within a virtual environment, thereby extending the scale of analysis and experience. This offers a sense of embodiment in architecture or landscape (sense of place) (Forte and Bonini 2010), and increases awareness of spatial relationships between objects and associated data (Frischer and Hild 2008), and most importantly, helps the public to better understand archaeological information.

3.2 Beta-testing results of QueryArch3D Tool

Firstly, the project wanted to understand and evaluate the advantages and limitations of virtual

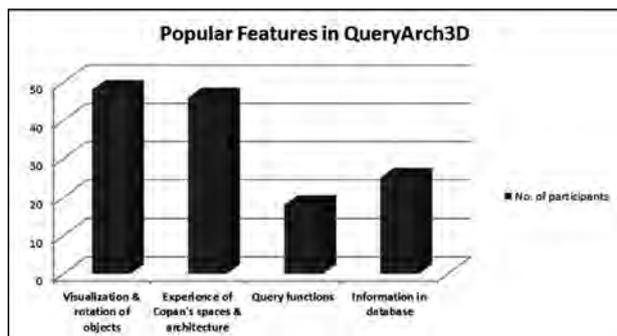


Figure 4. Most popular features in QueryArch3D tool.

heritage in terms of engagement and education (Champion, Bishop, and Dave, 2012, Flatman, Chidester and Gadsby 2012). For this reason, in September-October 2011, the MayaArch3D Project conducted beta-testing of the QueryArch3D tool at five universities—University of New Mexico (USA), University of Merced California (USA), California State University Stanislaus (USA), University of Bonn (Germany), and Umeå University (Sweden). Most participants were anthropology and art history students (graduates and undergraduates) between the ages of 15-35, who had little experience with 3D models and were interested in the tool from an educational standpoint. Five professors of anthropology and archaeology also participated in the beta-tests. Participants were provided with background information about the MayaArch3D Project and the QueryArch3D tool and received instructions on how to open and navigate the tool. They spent approximately 30 minutes exploring the tool and then filled out a survey. The survey included 14 questions about user demographics (i.e. age, gender, work area), computer specifications (e.g. platform (MAC or PC), browser, RAM), most and least popular features, and most important features for immediate development. The survey also asked participants: What did you learn from the tool? And, how was this different from your traditional classroom studies?

Figure 4 shows that the most popular features were the “visualization and rotation of objects” and the “experience of Copan’s spaces and architecture”. However participants would have preferred to see more textured and less blocky models. Along these lines, they also wanted a transparency feature to indicate the degree of certainty in the 3D models (i.e. procedural vs. 3D Studio Max vs. reality-based). With respect to learning, participants cited a better

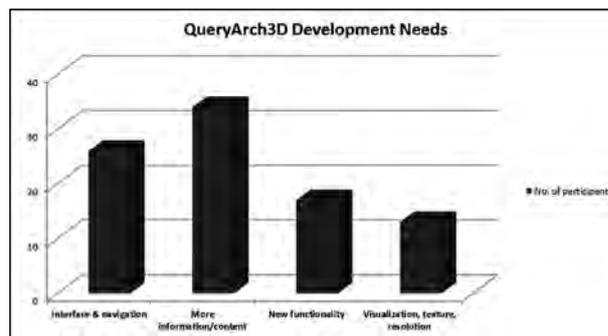


Figure 5. Development Needs in QueryArch3D tool

understanding of the size and scale of Copan and of the heights and spatial relationships of buildings than from 2D photos and textbooks. They also mentioned that while the VR gave them the sense that Copan was a large city, and not simply a ritual centre, the lack of people and objects made the city seem unrealistically vacant. The participants liked being able to access information in the database via the 3D models; however, they mentioned that more data are needed in the database before they could adequately assess the value of this feature for educational purposes. The Query Functions were the least popular. These are the functions at LOD 1 where users can select, for example, all stelae that were erected by ruler 16 and they will be highlight in the VR. This may be because these were designed with researchers in mind.

Fig. 5 illustrates that, according to beta-testers, the most important feature needing immediate development is “interface and navigation”. In response to this question, several participants suggested that the addition of avatars and/or more realistic movement within the VR would enhance user experience and engagement. Building on this feedback, we reviewed the state-of-the-field for virtual heritage employing maps, GIS, and 3D zs, and investigated device-based and gesture-based virtual heritage. Based upon our findings, we began a collaborative project with the HUMlab at Umeå University (Sweden) to link the QueryArch3D tool to Microsoft’s Kinect in order to offer an alternative way for users to navigate the QueryArch3D tool.

4. QueryArch3D-Kinect: System Components, Development and Set Up

The gesture-based 3D GIS system is low-cost and portable requiring four items:

Keyboard Command	Onscreen movement	Kinect gestures
W	Move Forward	Right arm forward
S	Move Backward	Left arm forward
Q	Left Turn	Rotate body left
E	Right Turn	Rotate body right
Y	Look Up	Lean backwards
H	Look Down	Lean Forwards
T	Start walkthrough or enter LOD4	Right arm down
O	Open information (access database)	Right arm up
I	Close information	Left arm up
Keyboard Command	Queries/Interaction with high-resolution 3D objects	Kinect gestures
Z	Zoom in	Right arm forward
X	Zoom out	Left arm forward
←	Rotate object left	Left arm out
→	Rotate object right	Right arm out
↑	Rotate object up	Lean backwards
↓	Rotate object down	Lean forwards
O	Open information	Right arm up
I	Close information	Left arm up

Table 1. Keyboard commands and Kinect gestures for the gesture-based 3D GIS.

- QueryArch3D tool- free and available online using a free Unity 3D plug-in or offline,
- Kinect,
- standard desktop/laptop,
- display device.

QueryArch3D uses a DBMS for data modelling and storage, and 3D visualisation to link multi-resolution 3D models to archaeological data that are accessed while navigating in a VR environment. The DBMS consists of PostgreSQL with the PostGIS extension to reduce data heterogeneity and allow storage of both non-spatial and spatial data. For the interactive navigation and 3D visualisation, the tool uses Unity3D, a game engine development tool for the creation of 3D interactive contents accessed off-line and on-line using a free web player plugin. Finally, a PHP interface links Unity 3D and PostgreSQL allowing the data retrieval from the database and the (online) visualisation.

The initial development of QueryArch3D programmed movements with mouse navigation. To allow gesture-based control of QueryArch3D,

we selected Microsoft's Kinect—a motion sensing input device composed of different sensors (RGB camera, infrared depth sensor, motorized tilt function, and a microphone array) that enables users to control and interact with video games (or PCs) without the need to touch a game controller. Originally designed as motion sensing input device for playing videos games with Microsoft's XBOX 360, developers and researchers are now exploring alternative applications for the Kinect beyond its original entertainment purpose (Giles 2010). There are currently four options to use Kinect with the QueryArch3D tool: (1) employ the Flexible Action and Articulated Skeleton Toolkit (FAAST), (2) use the ZDK Unity3D bindings in the Zigfu Development Kit (works with OpenNI/NITE and the Microsoft Kinect SDK, but it is not free), (3) develop a custom interface in the .NET framework using Kinect for Windows SDK, or (4) develop a custom interface in the OpenNI framework. For our initials tests we have selected the FAAST option because it is open source and quickly and easily customizable. The mouse navigation commands, in Unity 3D, were reprogrammed to keyboard commands and then the keyboard commands were correlated to body posture and specific gestures (Table 1). At this

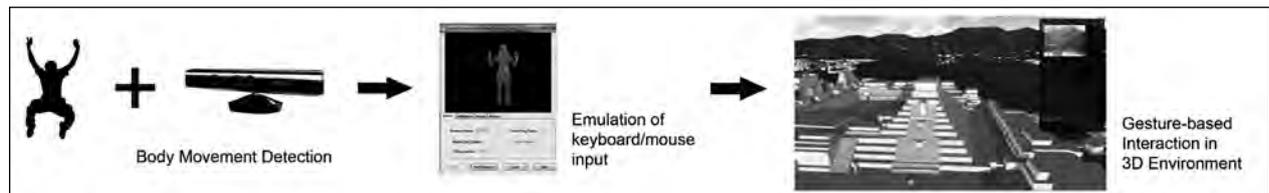


Figure 6. The system set-up, with the Kinect recording the human movements and FFAST emulating the keyboard input to control the QueryArch3D tool.



Figure 7. Set-up of the gesture-based platform to control the 3D environment.

point, these commands can be emulated using the Flexible Action and Articulated Skeleton Toolkit (FAAST) allowing for gesture-based control of the QueryArch3D tool (Suma et al. 2011).

The system set up is simple and straightforward (Figs 6 and 7):

- Download FFAST onto a PC;
- Connect Kinect to the PC via a USB adapter;
- Open the QueryArch3D tool on www.MayaArch3D.unm.edu (for first use the Unity 3D plug-in must be installed on the PC);
- In FFAST, open the configuration file (.cfg) that indicates how the keyboard commands are translated to body posture and gestures;
- FFAST performs a skeleton calibration (specific to the user);
- Begin gesture-based navigation of QueryArch3D tool.

4.1 The interactive display(s) set-up: interaction and immersion

For a public display/exhibit, five options based on different interaction and immersion technologies were identified:

Non-immersive/Device-based - This option uses device-based interaction technologies, i.e., visitors use a mouse and keyboard to control the QueryArch3D tool. The display device consists of a large monitor or a projector and screen/wall.

Non-immersive/Natural-based - This option uses natural-based interaction technologies, i.e. visitors use gestures (via the Kinect and FFAST) to control the QueryArch3D tool. The display device consists of a project and screen/wall.

Non-immersive/Device and Natural-based - This option uses both device-based and natural-based technologies, i.e. two stations are set up. For example, a desktop with monitor for the device-based interaction and a large screen/wall with projector for gesture-based interaction. This option better serves the needs of diverse visitors and their different learning styles.

Low-immersion/Device-based - This option uses a 3D stereo display with device-based interaction with the mouse and keyboard requiring stereoscopic driver software and glasses. The display device consists of a screen or projector and one or three screens. This option is more expensive, but allows visitors to see stereoscopic 3D objects, providing a more realistic sense of space for users as they navigate and explore Copan's architecture.



Figure 8. Using gestures to navigate and query information in the QueryArch3D tool

Low-immersive/Natural-based- The set up for this option is the same as option #4; however, the system uses the Kinect and FFAST instead of a mouse and keyboard. Using gesture-based interaction gives visitors a sense of using their own body to walk through the ancient city of Copan, offering a sense of embodiment that may be lacking with device-based interaction.

4.2 Practical factors to consider for the gesture-based 3D GIS exhibit

While the system set-up is straightforward, there are several factors to be considered for when using the gesture-based system for a public exhibit:

While the range limits of the Kinect as a hands-free controller are 1.2–3.6 meters, our tests indicate that the performance of the gesture-based system increases the closer the user is to the Kinect. However, the small field of view (FOV) of the Kinect might not entirely capture a user who stands too close to the sensor. Thus, when setting up a public exhibit, markers notifying users of appropriate distance to stand from the display device should be made. The ideal distance that a visitor stands from the display depends on user height. For example, while children need to stand closer to the display, taller adults must stand further away to enable the Kinect to perform the skeletal calibration and gesture recognition. To accommodate and overcome these problems, we programmed and tested FFAST for two skeleton modes: (1) Full Body and (2) Upper Body. Our tests indicate that a solution to poor system performance due to distance can be overcome by using the Upper Body mode;

FFAST requires that different body movements be programmed to certain distances or degrees to trigger specific keyboard commands. This factor corresponds directly to the angle of the Kinect and thus, the optimal height placement of the Kinect is at the torso level of the user. The latest version

of FFAST (v.1.0) includes a manual pitch (up/down angle) function for tracking users in order to account for people of different heights. Additionally, the actions in FFAST are programmed to be triggered at specific distances, for example, an information table can be programmed to be opened when a user raises his/her arm 15 inches above the head; however, if a child or person of small stature is using the system, the action will not be triggered if he/she cannot raise his/her arms high enough above the head;

The system's performance increases when the user directly faces the Kinect; vice versa, when the Kinect is positioned off-centre, maybe due to room-size constraints, the system's performance decreases. To avoid this problem, markers should also be used indicating where users should stand. In other words, a "boundary box" (perhaps tape on the floor) identifying the ideal distance and placement of visitors should be marked to facilitate maximum performance of the gesture-based system. Along these lines, markers indicating the ideal distance/ placement of children compared to adults are also important. We suggest a set-up test that allows visitors to identify the optimal distance and placement from the display;

According to our tests, the system's performance increases when user gestures are relatively slow and well-defined. We suggest that the set-up test also guides visitors through each of the gestures required to navigate the system, providing them with feedback on the success of their movements. In other words, the set-up test should inform them on the range of movement and speed required to optimally use the system;

The upper or entire skeleton identification and gesture recognition are fully automated procedures; therefore, the illumination conditions of the environment are crucial for correct gesture identification and command transmission. For example, our tests indicate that the system's



Figure 9. East Court as it looks today (left) and 3D digital reconstruction of East Court in QueryArch3D showing now collapsed buildings (right).

performance decreases with larger amounts of ambient light (i.e. natural light) and it also decreases in the presence of highly reflective objects such as glass. It is suggested that the display be housed in a relatively dark space with few objects.

In summary, the gesture-free display should be mounted in a separate room with floor markers telling users of different heights where to stand, and include a set-up text for each user or training session or “game” or “test” in which they briefly learn the proper gestures.

4.3 Exhibit content: sample experiences

An important advantage of virtual heritage is the ability to include many types of media such as photos (including panoramas), drawings, videos, maps, and 3D visualisations—the QueryArch3D tool not only does this in a single, navigable VR, but it also allows users to perform attribute queries by clicking on 3D virtual objects (i.e. stelae, buildings, and sculpture) and call up photos, maps, drawings, and textual descriptions that are stored in a spatial database. Moreover, because the tool is a 3D GIS it also allows users to perform spatial queries; for example, they can select all structures belonging to a specific sub-community within the ancient city of Copan and these structures will be highlighted in the VR (Fig. 8). Coupled with the ability to simultaneously query 3D models, the integration of 3D GIS and controller-free interaction technologies offer a unique and innovative way to engage and educate the public about the ancient Maya (see video at <http://vimeo.com/50673261>).

Another advantage of virtual heritage is that it is easier and less expensive to devise different ways to communicate cultural heritage content to the public. We offer four options that offer different levels of interactivity:

Non-interactive Narration: Videos are played that present a storyline, or narrative, that was created from the VR environment in QueryArch3D tool; visitors are not free to explore the VR environment.

Interactive Narration: self-guided virtual tours where a storyline, or narrative, in chronological or sequential order unfolds as visitors “walk” along pre-set paths; visitors are free to explore the pre-set path at their own pace.

Hot-spots: visitors select highlighted locations from overview map to explore; at each of these locations a specific theme(s), e.g. cosmology, are described through interaction (e.g. rotation of 3D virtual objects such as a stela to “read” hieroglyphs).

Unguided exploration—visitors move around freely and click on any object to retrieve data from database; this option allows visitors to move through the city at their own pace, inspect specific features, or read hieroglyphic texts; new knowledge creation is likely.

An exhibit—based on the QueryArch3D tool and the case study of Copan—could be used to engage and educate the public about the ancient Maya, and archaeology more generally. Some topics include:

- Explain archaeological methods and theories using the 3D digital models of buildings with linked excavation photos, drawings, etc. to tell the history of archaeological research at Copan. For example, excavations could be selected and highlighted based on year excavated to illustrate broader changes in the discipline of archaeology from focus on civic-ceremonial centres to settlement patterns;
- Use the reality-based and 3D Studio Max models to illustrate ancient Maya construction

techniques and materials;

- Visualise buildings that have been completely destroyed in a landscape context (Fig. 9);
- Explain ancient Maya indigenous concepts such as cosmology, astronomy, time (Maya calendars), social organization, or divine kinship. For example, excavations in the city's main civic-ceremonial complex (Principal Group) have uncovered layers of architecture with sculpture, imagery, and hieroglyphs showing that Copan had a dynasty of sixteen kings that ruled over five centuries (A.D. 427-820) (Andrews and Fash 2005). Visitors could a guided virtual tour that traces the city's dynastic history using the temples and other monuments that city's various kings commissioned.

5. Conclusions and Future Direction

This paper has explored the potential of interactive technologies to facilitate teaching and learning by linking two technologies, 3D GIS and gesture-based interaction, for the purposes of developing a new system to engage the public in ancient Maya archaeology. Beyond the practical requirements of setting up a public gesture-based 3D GIS exhibit, it is important to establish learning objectives and take into account diverse learning styles, educational background, and informatics skills (Hsu 2011). By offering different interaction (device vs. natural) and immersion technologies combined with multimedia and different levels of interactivity (ranging from a passive experience such as watching videos to a highly active experience using the gesture-based system for unguided exploration through the ancient city of Copan) exhibition curators can engage a range of visitors in the cultural heritage of the ancient Maya. The gesture-based system is suited to today's youth—visual learners who are intrigued by, and proficient in using interactive and dynamic technologies. The system is also well-suited to individuals lacking experience with computers (specifically using mouse and keyboard) or have physical constraints because it can be programmed in multiple ways to have body movements correspond to a range of different gestures (upper body vs. full-body interaction). In the future, the project will continue to evaluate the advantages and disadvantages of

using a touchless control system to navigate a 3D GIS for archaeological outreach, keeping an eye on technological developments and exploring the range of benefits for cultural heritage.

Acknowledgements

This project was partly supported by two Digital Humanities Start-Up Grants from the National Endowment for the Humanities (#HD50583, #HD5097910), National Science Foundation (#1064648), Alexander von Humboldt Foundation (Germany) and HUMlab, Umeå University (Sweden). Thanks to Maurizio Forte of the University of Merced for assistance with beta-testing and for advice about the direction of the project, and to Keith Prufer of the University of New Mexico, Nikolai Grube of the University of Bonn, and Ellen Bell of the University of California Stanislaus for assistance with beta-testing. We also thank the Honduran Institute of Anthropology and History (IAHA) for the collaboration and permission to work at Copan.

References

- Abate, D., G. Furini, S. Migliori, and S. Pierattini. 2012. "Virtual Hands Free Interaction with 3D Objects and Environments." Paper presented at the *Computer Applications and Quantitative Methods in Archaeology Conference, Southampton, UK, March 26-29*.
- Acevedo, D., E. Vote, D. Laidlaw, and M. S. Joukowsky. 2001. "ARCHAVE: A Virtual Environment for Archaeological Research." *CAA 2000 Proceedings: Computing Archaeology for Understanding the Past*, edited by T.Z.V. Stancic, 313-316. Oxford: Archeopress.
- Agugiaro, G., F. Remondino, G. Girardi, J. von Schwerin, H. Richards-Rissetto, and R. De Amicis. 2011. "A Web-based Interactive Tool for Multi-Resolution 3D Models of a Maya Archaeological Site." In *Proceedings of the 4th ISPRS International Workshop 3D-ARCH 2011: "3D Virtual Reconstruction and Visualization of Complex Architectures" Trento, Italy, 2-4 March 2011*, edited by F. Remondino and S. El-Hakim, 23-30. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences (on CD-ROM).
- Andrews, W.E., and W.L. Fash. 2005. "Issues in Copán Archaeology." In *Copán: The History of an Ancient*

- Kingdom, edited by E. Wyllys Andrews and W.L. Fash, 395-425. Santa Fe: School of American Research Advanced Seminar Series.
- Aquae P. 2012. "La Visita Virtuale." Accessed June 23, 2012. http://www.aquaeptavinae.lettere.unipd.it/portale/?page_id=2174.
- Barcelo, J.A., M. Forte, and D. H. Sanders. 2000. *Virtual reality in archaeology*. Oxford: Archaeopress.
- Beusing, R., and A. Posluschny. 2009. "The NetConnect-Project – New Access to Ancient Sites." Paper presented at *Computer Applications to Archaeology, Williamsburg, Virginia, USA. March 22-26, 2009*.
- Carrozzino, M., and M Bergamasco. 2010. "Beyond virtual museums: Experiencing Immersive Virtual Reality in Real Museums." *Journal of Cultural Heritage* 11: 452-458.
- Champion, E., I. Bishop, and B. Dave. 2012. "The Palenque project: evaluating interaction in an online virtual archaeology site." *Virtual Reality* 16: 121–139.
- Conolly, J., and M. Lake. 2006. *Geographical Information Systems in archaeology*. Cambridge: Cambridge University Press.
- Dawson, P., R. Levy, and N. Lyons. 2011. "Breaking the fourthwall: 3D virtual worlds as tools for knowledge repatriation in archaeology." *Journal of Social Archaeology* 11(3): 387-402.
- de Byl, P., and M. Khan. 2011. "Preserving our Past with Toys of the Future." *Changing Demands, Changing Directions. Proceedings of ascilite Hobart 2011*, edited by G. Williams, P. Statham, N. Brown, and B. Cleland, 718-728. Hobart: University of Tasmania. .
- Flatman, J., R.C. Chidester, and D.A. Gadsby. 2012. "What Public Engagement in Archaeology Really Means." In *Archaeology in Society: Its Relevance in the Modern World*, edited by Marcy Rockman, Marcy and Joe Flatman, 65-76. London: Springer.
- Fischer, L. E. 2012. "Virtually Visiting Williamsburg: The Use of 3D Modelling to Enrich the Museum Guest Experience." Paper presented at the Computer Applications and Quantitative Methods in Archaeology Conference, Southampton, United Kingdom, March 26-9.
- Forte, M., and A. Siliotti. 1997. *Virtual Archaeology*. London: Thames and Hudson.
- Forte, M., and E. Bonini. 2010. "Embodiment and Enaction: A Theoretical Overview for Cybercommunities." In *Heritage in the Digital era*, edited by M. Ioannides, A. Alonzo, A. Georgopoulos, L. Kalisperis, A. Brown, and D. Pitzalis, 45-56. Essex: Multi-Science Publishing.
- Forte, M., and G. Kurillo. 2010. "Cyberarchaeology: Experimenting with Teleimmersive Archaeology." Paper presented at the 16th International Conference on Virtual Systems and Multimedia, Oct 20-23, 2010, Seoul, South Korea.
- Frischer, B., and A. Dakouri-Hild. 2008. *Beyond Illustration: 2D and 3D Digital Technologies as tools for Discovery in Archaeology*. Oxford: Archaeopress.
- Giles, J. 2010. "Inside the race to hack the Kinect." *The New Scientist* 208 (2789): 22-23.
- Hsu, Hui-mei J. 2011. "The Potential of Kinect in Education." *International Journal of Information and Education Technology* 1 (5): 334-338.
- Huk, T., M. Steinke, and C. Floto. 2010. "The educational value of visual cues and 3D-representational format in a computer animation under restricted and realistic conditions." *Instructional Science* 5: 455-469.
- HUMLab. 2012. "HUMLab Homepage." Accessed June 11, 2012. <http://www.humlabs.umu.se>.
- Ioannides, A., A. Addison, A. Georgopoulos, L. Kalisperis, A. Brown, and D. Pitzalis. 2010. *Heritage in the Digital Era*. Essex: Multi-Science Publishing.
- Lange, B., A. Rizzo, C. Chang, E. Suma, and M. Bolas. 2011. "Markerless Full Body Tracking: Depth-Sensing Technology within Virtual Environments." Paper presented at the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC) 2011.
- Manferdini, A., and F. Remondino. 2012. "A review of reality-based 3D model generation, segmentation and web-based visualization methods." *International Journal of Heritage in the Digital Era* 1(1): 103-124.
- MayaArch3D Project. 2012. "MayaArch3D Project Homepage." Accessed 11 June 2012.

<http://mayaarch3D.unm.edu>.

Noh, Z., M. Shahrizal Sunar, and Z. Pan. 2009. "A Review on Augmented Reality for Virtual Heritage Systems." In *Learning by Playing. Game-based Education System Design and Development, 4th International Conference on E-learning, Edutainment 2009 LNCS 5670*, edited by M. Chang, R. K. Kinshuk, G. Chen, and M. Hirose, 50–61. Berlin and Heidelberg: Springer-Verlag.

Pantano, E., and A. Tavernise. 2009. "Learning Cultural Heritage Through Information and Communication Technologies: A Case Study." In *Learning Culture and Languages through ICTS: Methods for Enhanced Instruction*, edited by Maiga Change and Chen-Wo Kuo, 103- 119. London: IGI Global.

Pescarin, S., W. Hupperetz, A. Pagano, C. Ray, and M. Wallergård. 2012. "Evaluating Virtual Museums: archeovirtual test-case." Paper presented at the *Computer Applications and Quantitative Methods in Archaeology Conference, Southampton, United Kingdom, March 26-9*.

Pujol, L., and A. Lorente. 2012. "The Virtual Museum." Paper presented at the *Computer Applications and Quantitative Methods in Archaeology Conference, Southampton, United Kingdom, March 26-29*.

Remondino, F. 2011. "Heritage Recording and 3D Modeling with Photogrammetry and 3D Scanning." *Remote Sensing Journal* 3: 1104-1138.

Remondino, F, Alessandro Rizzi, Belen Jimenez, Giorgio Aguiaro, Giorgio Baratti, Giorgio and Raffaele De Amicis. 2011. "The Etruscans in 3D: from space to underground." *Geoinformatics* 6:283-290.

Roussou, M. 2008. "The Components of Engagement in Virtual Heritage Environments." *New Heritage: New media and cultural heritage*, edited by Yehuda Kalay, Thomas Kvan and Janice Affleck, 225-241. London and New York: Routledge.

Sanders, D.H. 2008. "Interactive 3D Computer Graphics for Teaching, Research, and Publication in Cultural

Heritage." In *Proceedings of the International Symposium on Information and Communication Technologies in Cultural Heritage, Oct. 16-18, University of Ioannina, Greece*, edited by T. Mikropoulos, N. Papachristos, 143-158. Ioannina: Graphic Arts.

Styliani, S., L. Fotis, K. Kostas, and P. Petros. 2009. "Virtual museums, a survey and some issues for consideration." *Journal of Cultural Heritage* 10: 520-528.

Suma, E., B. Lange, A. Rizzo, D. Krum, and M. Bolas. 2011. "FAAST: The Flexible Action and Articulated Skeleton Toolkit." *Proceedings of IEEE Virtual Reality*, 247-248. Singapore.

Vecco, M. 2010. "A definition of cultural heritage: From the tangible to the intangible." *Journal of Cultural Heritage* 11 (3): 321-324.

Villaroman, N., D. Rowe, and B. Swan. 2011. "Teaching Natural User Interaction Using OpenNI and the Microsoft Kinect Sensor." In *Proceedings of the 2011 SIGITE conference on Information Technology Education*, edited by B. Goda, E. Sobiesk, and R. Connolly, 227-232. New York: ACM.

Virtual Museum Transnational Network. 2011. "What is a Virtual Museum?" Accessed May 30. <http://www.v-must.net/virtual-museums/what-virtual-museum>.

von Schwerin, J., H. Richards-Rissetto, F. Remondino, G. Aguiaro, M. Forte, and R. Maqueda. 2011. "The MayaArch3D Project; Digital Technologies for Research in Maya Archaeology." Final Performance Report and White Paper for NEH Digital Humanities Level II Start-Up Grant, 2011. <https://securegrants.neh.gov/PublicQuery/main.aspx?f=1&gn=HD-50975-10>.

Wheatley, D.W., and M. Gillings. 2002. *Spatial technology and archaeology: a guide to the archaeological applications of GIS*. London: Taylor & Francis.

Zerneke, J.L., M.K. Buckland, and K. Carl. 2006. "Temporally Dynamic Maps: The Electronic Cultural Atlas Initiative Experience." *Human IT* 8 (3): 83–94.

Integration of 3D Laser Scanning and Virtual Reconstructions as Research and Educational Tools for Representing the Past. The Case Study of Roman Baths of Edeta

Vito Porcelli, Fernando Cotino Villa, Josep Blasco i Senabre
Global Geomática SL, Grupo Global, Valencia, Spain

Vicent Escrivá Torres
Museu Arqueològic de Lliria (MALL), Spain

Julian Esteban Chaparría
Conselleria de Cultura, Generalitat Valenciana, Spain

Abstract:

In the last few years 3D laser scanning techniques have been used more frequently as a means for recording archaeological evidence. By contrast, 3D reconstructions, more often than not, serve purely educational purposes, and their efficiency for scientific analysis is disregarded. A combination of the two approaches supports the likely transition of the virtual reproductions; extending them to the interest of varied categories of users. This paper puts an emphasis on the process of reconstructing the Roman Baths of Edeta, Lliria. It shows the capabilities of the 3D Laser scanning approach of using accurate digital data derived from the real evidence and the restoration of the different locations and shapes of the feature during the past. Additionally, 3D virtual reconstruction is used to further analyse the site and, eventually, be used to extend the general public's knowledge of its cultural heritage through highly realistic images.

Keywords:

3D Laser Scanning, 3D Modelling, Edeta, Point Cloud, Roman Baths, Lliria

1. Introduction

Computer Graphics has completely converted the traditional archaeological recording system. Currently, "Digital Archaeology" would be the best term to refer to it (Earl et al. 2012). Whether the information concerns spatial data or physical remains, drawings or pictures of the sites, they are archived and processed as digital information.

Thanks to the scholars involved in the process of preservation and conservation of the site, the archaeological data derived from the Roman Baths of *Edeta* have been recorded following this digital methodology. The work has culminated with the introduction of realistic virtual restoration of the edifices, principally focusing on educational purposes.

The hypothetical restoration has been generated taking advantage of accurate spatial data
Corresponding author: porcelli.v@gmail.com

captured through 3D Laser Scanning. The use of this technology as well as other significant digital information of the site, has likely turned realistic 3D virtual models of Roman buildings into useful data. In fact, the visualisations might be appreciated not just by amateurs of history or inexpert users, but also by scientists involved in the analysis of the archaeological evidence.

Virtual development of raw geometry, traditionally recorded using photogrammetry, CAD drawings or even 3D Laser Scanning, has never been performed without criticism by experts (Lock 2003). However, visual representations as drawings or paintings of ancient buildings have often been produced to support the conservation of cultural heritage. Examples of them might be seen in several old publications, even though most of them were produced for public dissemination. The 3D virtual model of the Roman Baths of *Edeta*, together with all the recent virtual reconstructions of archaeological sites, should be considered as a transition from



Figure 1. Photograph. Current view of the Roman Baths of Edeta. The cella of the temple (on the left) and the Basilica Thermarum are shown. Photograph coming from the archive of the project.

those 2D sheets to a modern approach for recording, measuring and visualising information in a 3D virtual world.

2. Edeta, the Iberian and Roman Village

Llíria is a small village of 20.000 inhabitants lying on the river Turia's plain, approximately 25km northwest of Valencia (Spain).

On the nearby hill, next to the village Tossal de Sant Miquel, the Iberian settlement (Pre-Roman) continued to develop and expand from the VIII century BC. until the II BC. In the course of the I century BC. the new Iberian village was founded on the river Turia's plain and took the name of *Edeta*, already conveying Roman influences. Afterwards, in the Augustan Age, it was converted into a *Municipium* under the Latin Rights (*Ius Latinum*). During the rest of the Roman Empire, the village was expanded until it covered more than 20.000m². This extension included the central area of the oracular sanctuary and the baths (Escrivá Torres and Vidal Ferrús 1995:238). In the late phase of the empire this important area began to be neglected. The ground was reoccupied during the Visigothic Age, whereupon a monastery was built and consequently abandoned from the second half of the VII century AD onwards.

The first excavation campaigns began in the last century, during the 70s, discovering the sanctuary with the lateral *Aedicula* and part of the *Hospitium*. Subsequently, the excavations restarted in the 90s



Figure 2. Photograph. Current northern view of the Hypocaustum of the Caldarium. Photograph coming from the archive of the project.

and in 1994 the male and female baths were found. The male baths were extensively excavated, with the full structure revealed during the actual excavation campaign. The next step for work at this site, regards the minor female baths, which will be excavated and analysed in the coming years (Escrivá Torres and Vidal Ferrús 1995, 238-239).

The male Roman baths consist of stairs off of the public street for access into the *Basilica Thermarum*. This building was composed of two floors with a portico surrounding a central atrium and a small *Taberna* on the northern side (Fig. 1). On the eastern side a central door allowed access into the other rooms, in the following sequence: *Apodyterium* – changing room, *Frigidarium* – cold pool, *Districtarium* or *Unctorium* – cold room, *Laconicum* – dry sweating room and finally, the *Caldarium* – hot room. This latter room was arranged with a labrum – basin containing cold water, in the apse and opposite to it an *Alveus* – long warmed pool, next to high windows positioned at the shortest northern wall. Between the *Apodyterium* and the *Caldarium*, on the northern side, there was the *Praefurnium*, the central structure for the heating system of the rooms. The level of conservation is excellent, above all, zones of the hypocausts, in particular the one beneath the *Caldarium* (Fig. 2).

In the last few years, thanks to a desire to improve and preserve the facilities in the urban area of Llíria, a project of restoration and consolidation of the Roman buildings has been carried out. As a result, the area has been thoroughly analysed, recorded and is currently being converted into



Figure 3. Image. View in SCENE of the point cloud of the significant finds recorded with 3D Laser Scanning. The shown sphere on the tripod is a common point used for matching the others captured point clouds.

a museum. The attractive room dedicated to the visitors will include videos, panels and an interactive virtual 3D model of the Roman Baths, along with other devices designed to test augmented reality of the archaeological remains.

3. Data Acquisition

The data used for the project has been acquired during the last years utilising different methods which reflect the new technologies and methodologies in archaeology, computer science and topography.

The 3D reconstruction of the Roman Baths has certainly been the main purpose of the entire project. However, given other subsequent aims, such as the future production of rendered images and virtual animations, the model could not be composed of a large amount of polygons, rather it needed to be particularly light to remain easy to handle in the future.

To make a distinction between the two 3D objects, the one generated from 3D Laser Scanning and the other modelled with low resolution polygons in 3dsMAX, this paper introduces the term of “3D mesh”. This term is a standard one, generally applied to both types of object, signifying a simple 3D cube as well as a 3D model generated by means of point clouds and often, consisting of millions of

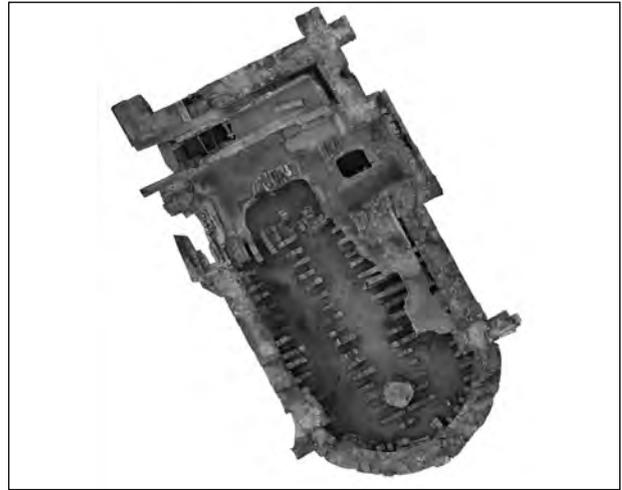


Figure 4. Orthophoto. Plan of the Caldarium produced with photogrammetry. Data coming from the archive of the project.

faces. In this case, to avoid misunderstandings, the term “3D mesh” will be cited exclusively for the high resolution meshes from 3D Laser Scanning.

With regard to the 3D Laser Scanning of the site, the point clouds have been captured using the 3D Laser Scanner FARO FOCUS 3D. It is compact and light, capturing up to 976.000 points/sec and reaching a distance up to 120m. Though these values are dependent on the features of the objects captured, the area and the atmospheric conditions. The point clouds were concentrated on the zone of the male Roman Bath and the external Temple, along with some adjacent buildings. In addition, a significant amount of archaeological finds in 3D has been digitally stored, recognising amongst them: capitals, portions of columns, cornices and stones from the collapsed naves of the buildings (Fig. 3).

The 3D meshes are extremely useful for producing a 3D virtual reconstruction. They precisely represent geometric information of the archaeological features and often have helped with the complete restoration of small objects. Many studies have focused on these kinds of projects and have ended with high resolution results, therefore large optimisation of the meshes is not needed (Patay Horváth 2011). These meshes might also represent large sites and used as base-plan for restoring hypothetic upper shapes. Nevertheless, in order to easily handle them, particularly in this case, they have to be optimised. This process still takes much more

time than starting the model from scratch, as, for instance, using simple 2D extruded polygons. And, furthermore a significant reduction of the resolution implies, in any case, loss of spatial information. Hence, this paper will describe alternative uses of them, explained in detail in the following paragraphs.

The 3D Laser Scanning has been contemporary with the 3D reconstruction of the site. Hence, the data collected has not been entirely used to elaborate the solids in 3D. In association, general photos, photogrammetry and CAD plans recorded during previous excavation campaigns have supported the 3D reconstruction process, occupying a secondary position. The CAD plans were drawn from the digital orthophotos. The original close-range photos were acquired and ortho-rectified using common points identified with coloured pins placed on the terrain. Apart from their use for mapping and matching the photos, the pins were placed in order to record their real spatial position in the global coordinate system using the Total Station. In this case, the *Universal Transverse Mercator* coordinate system. (Fig. 4). Therefore, in the few zones where 3D Laser Scanning was not yet used, the 3D objects have been modelled from the polylines manually drawn in CAD system from the orthophotos. Those produced in the described manner still conserve signs of human error, completely removed by means of the digital process performed by the 3D Laser Scanner. Nevertheless, the high resolution orthophotos remain useful to identify characteristics of the field or to quickly measure distances on the plan in CAD or in a scaled printed version.

As detailed, the visual imagery including orthophotos and standard photos, supply a visual aid in the 3D reconstruction. For this reason, several other pictures have been captured in the archaeological site to support the project. They allow volumetric analysis of the building, recognising some existent features on the ground and on the walls. They are also helpful to elaborate textures subsequently applied on the 3D model.

Finally, to support the 3D reconstructions, other archives have been obtained. Such as,

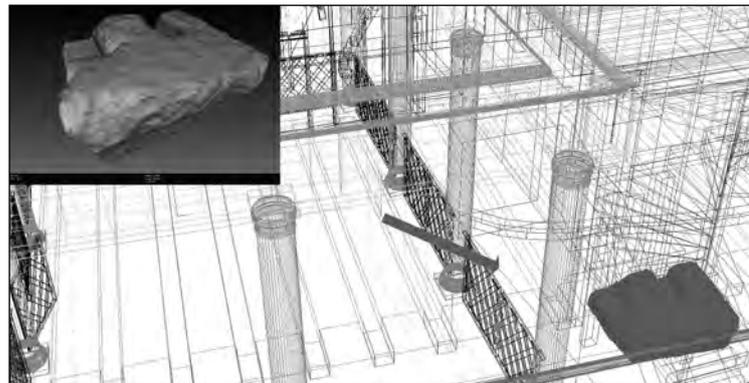


Figure 5. Image. View of the reconstructed Basilica Thermanum in wireframe. Hypothetic relocation of the 3D meshes of a cornice and its textured view in MeshLab.

photographs coming from other paralleled archaeological sites, In particular, Pompeii, Perga and Herculaneum. Furthermore, academic publications concerning the Roman architecture have been consulted, with the intent of knowing and understanding it as much as possible. Among them, the analysis and research conducted by Peter Connolly, Stefano Giuntoli and Maurizio Martinelli (Connolly 1994; Giuntoli and Martinelli 1989).

4. Methodology of the Digital Reconstructions

4.1 Spatial interpretation

The 3D reconstruction of the Roman Baths of Liria has been mainly achieved thanks to the experts involved in the project. They helped understanding the archaeological remains preserved in the area and respecting the limits set by the interpretation of the ancient architecture.

The 3D scans of the archaeological remains have been relocated in the virtual reconstruction, to make hypothesis of their likely ancient position and to support rebuilding the collapsed structures. For instance, the size of some stones discovered in the *Hypocaustum* of the *Caldarium*, useful to reconstruct the barrel vault. The ancient cornices supporting the beams of the first floor in the *Basilica Thermanum*, retrieving both the dimensions of the beams and the distance among them (one *cubitus*), suggesting a *bipedalis* floor (Fig. 5). Additionally, the height of its columns, composing and matching together their 3D meshes with their bases and

capitals. Another example is the restoration of the frame surrounding the monumental gate accessing the baths.

The recreated shapes of the perimeter walls are undoubtedly correct, as they have been extruded from the plans derived from 3D scans. However, the height of the restored walls mainly depends on the kind of stones employed in the past, on the Roman techniques applied, which kept transforming during the centuries and also, on similar and contemporary buildings.

One of the main issues have obviously concerned the height of the structures, whether they had to follow, the example of the Forum Baths in Pompeii still well conserved *insitu* and of the same age. Its barrel vault of the *Caldarium*, clearly shows width and height of the building having equal measures. In fact, the barrel vault is $\frac{1}{2}$ of the height of the building (Fig. 6). The first hypothesis made for the *Caldarium* of the Baths of Edeta is shown in figure 7 and followed this rule. Alternatively, restoring its barrel vault in accordance with the classic architectural proportions transmitted by Vitruvius, suggesting a proportion of $\frac{1}{3}$ (Vitruvius Pollio 1761: V). Finally, this latter proportion has been adopted for all the principal buildings, given that the contemporary Roman baths in Perga (Turkey) shows the same typology of stones of *Edeta* and respect Vitruvius' proportion rules. These have been useful for restoring the temple nearby too (Vitruvius Pollio 1761: IV).

The comparisons with surviving details conserved in others sites or with the measurements cited by Vitruvius have been essential. They have regarded several 3D objects in the scene, such as, the arrangement of the beams on the first floor in the *Basilica Thermarum*, the dimension and the shape of the storage shelves in the *Apodyterium* or for example, other details as the *labrum* in the *Caldarium* (Fig. 8), or the wooden benches located inside of it.

The existence of some others elements instead, is confirmed looking at the remains conserved *insitu*, such as the parapets among the columns or the boiler system of the warmed pool. However, when the archaeological information have been not enough to reconstruct these elements or when



Figure 6. Photograph. The internal view of the *Caldarium* of the Forum Baths in Pompeii. “The *Caldarium* of the Forum Baths”. Pompeii Interactive. Accessed November 20, 2011. Available from. <http://www.pompeii.co.uk/CDROM/FBATH/FAMES/F9-11.HTM>.



Figure 7. Rendered Image. Southern view of the first reconstructive hypothesis of the *Caldarium* in 3dsMAX, showing the passage from wireframe to textured 3D model. In this reconstruction the height of the barrel vault matches the height of the lateral walls.



Figure 8. Photograph. View of a *Labrum* conserved in Herculaneum. “Ercolano Scavi Archeologici”. Università degli Studi di Napoli. Accessed November 19, 2011. Available from. http://www.ercolano.unina.it/fotoErcolano/tdf_27.jpg.

others descriptions coming from ancient epigraphic data have been difficult to be found, the 3D solids have been mainly modelled following the drawings and the illustrations included in the Connelly's (1994) and Giuntoli's (1989) publications.

4.2 The software

The virtual reconstruction of the Roman Baths of Lliria, has been carried out by means of different software, specifically dedicated to 3D data management. Every task has generally its appropriate software. Currently, most of the Software-houses are expanding their adaptability, with the objective of using only one application to cover different fields of the Computer Graphics and to compete against their business rivals. For instance, using only one platform to manage data from 3D Laser Scanning, modelling in low polygons and modify raster 2D images at the same time.

However, for this project, every step has been processed with specific software, which better and quicker could process the digital information. With concern to the 3D Laser Scanning, the management and the matching of the point clouds have been performed in SCENE, software released by FARO. Whilst, MeshLab has been used for generating and optimising the 3D meshes as well as geo-referencing them. To produce the 3D virtual reconstruction, 3ds MAX 2009 has been the main software employed, with occasional utilisation of DraftSight for CAD data and REALWORKS by TRIMBLE, especially, prior the modelling process, to retrieve information from the point clouds. Eventually, the open-source software GIMP has been utilised to manage the 2D raster images.

4.3 3D modelling

The capability of 3D Laser Scanning to support reconstructing the buildings is practically shown in the modelling process. Recovering spatial information from the 3D meshes and comparing them with the 3D reconstruction, to make sure the hypothesis respects approximately the original evidence.

In order to record the entire site, 43 point clouds have been captured, setting 4mm as an average distance among the points. Small polystyrene

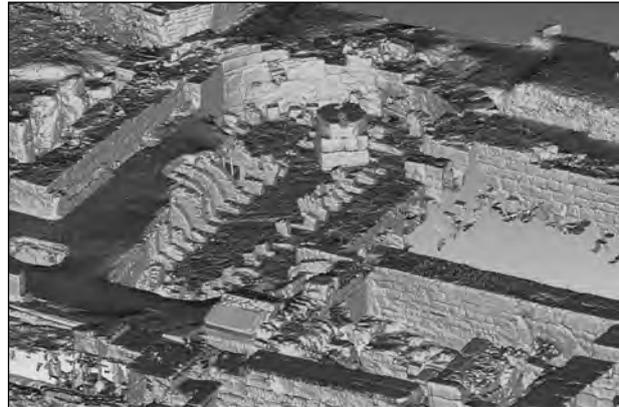


Figure 9. Image. View of a general point cloud showing the Caldarium, the Praefurnium and part of the Laconicum (on the right). The shown point cloud contained approximately 53,500,000 of points.

spheres have been used to mark the common points within the point clouds. After performing this task in the program SCENE, the point clouds have been imported into MeshLab., The point clouds have been cleaned from unnecessary noise, then optimised and joined, creating a general point cloud representing the archaeological site of Lliria (Fig. 9). The general point cloud has been georeferenced using the UTM coordinates of three known points recorded during the scanning process. They have been converted into a 3D mesh, digitally stored and ready for further utilisations.

In addition to the 3D Laser Scanning of the site, as above written, the project has also concerned the creation of a digital database of specific remains. Hence, about 660 significant finds have been three-dimensionally captured and transformed into 3D meshes, approaching in the same described manner, though turning upside down the objects to capture their back-faces (Fig. 3).

The future purpose of the virtual reconstruction implies the elaboration of a 3D model, with a polygon count sustainable for the rendering engine (Mental Ray), with an attempt to not exceed 1,000,000 polygons. The utilisation of the high resolution point clouds could have increased this value. On the other hand, in the case the point clouds were fairly reduced to respect the parameters, difficultly, they could still show the original shape in some detail, producing loss of data. In any case, many perimeter walls might be represented with squared shapes



Figure 10. Image. Point cloud of the Basilica Thermarum. Drawing of the polylines in CAD.

composed of few vertexes, enhanced by textures and easily renderable.

The quickest and easiest way applied to take advantage of the 3D meshes and at the same time to minimise the polygon count have principally been two. Firstly, by means of the projection of 2D orthophotos raster images of the point clouds in REALWORKS, in order to better recognise the contour lines of plans and subsequently, importing them into a CAD system to draw polylines (Alshawabkeh 2005, 44).

Secondly, and especially for others details challenging to be projected, directly importing the point cloud as DXF format in DraftSight (Fig. 10). The polylines have been subsequently extruded in 3ds MAX to recreate the main volumes. The usefulness of the point cloud is represented with its points extended three-dimensionally and showing already the location of the features in the 3D space (Alvaro 2009, 7).

There were a large number of small 3D meshes which had an extremely high spatial resolution, composing approximately 350,000 polygons each. This indicates that an approach similar to the 3D scans of the general area could be more effective. A good example would be extruding the most appropriate lines that match the profiles of the 3D meshes of the cornices or the capitals of the columns and rotating or extruding them by 360 degrees (Kampel and Sablatnig 2003) (Fig. 11).

A final comparison of the 3D mesh of the entire site with the final reconstructed 3D model has

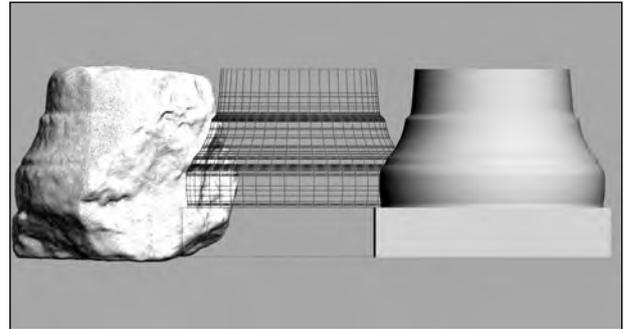


Figure 11. Rendered Image. Passage from high to low polygon 3D model in 3dsMAX. On the left the 3D mesh of the base of a column located in the Basilica Thermarum recorded with 3D Laser Scanning. On the centre and on the right, the wireframe and the textured restored 3D model.



Figure 12. Rendered Image. Southern view of the reconstructed Caldarium. Visual comparison of the structures visualised as wireframe and textured 3D models associated with the 3D mesh of the site.

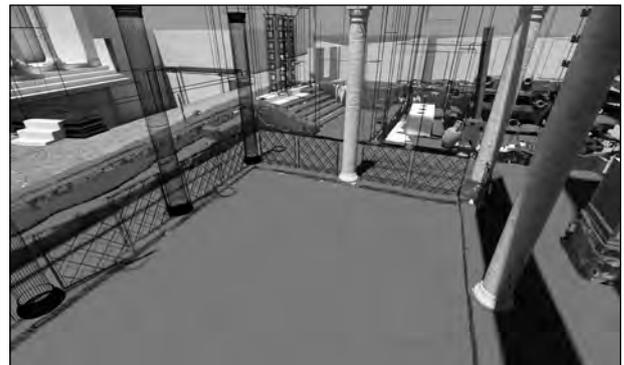


Figure 13. Rendered Image. Northern view of the reconstructed Basilica Thermarum. Visual comparison of the structures visualised as wireframe and textured 3D models associated with the 3D mesh of the site.



Figure 14. Photograph. Current view of the remaining walls conserved in the Basilica Thermarum. Detail of the paintings. Photograph coming from the archive of the project.

been performed, considering the possible existing discrepancies. Since the solids representing the 3D mesh have been intensely minimised and given the employment of orthophotos to create some of them too. This has been carried out superimposing the 3D data in 3ds MAX, observing carefully the differences among the surfaces and adjusting the 3D solids when required (Figs 12 and 13).

The aesthetic enhancements of the Roman baths and the holy area, especially the furniture and some decorations, have been modelled with minor scientific techniques. Nevertheless, always attempting to respect the known Roman measurements and looking for information through the cited publications upon Roman architectural styles.

The photographs have helped with numerous aspects of the virtual reconstructions. In the modelling process, they have been necessary as aid to understand better the volumes and the features still preserved in the site. The 3D meshes, sometimes store useless details of solids, which can be appreciated only acquiring real pictures for the specific purpose. An example is the *Hypocaustum* of the *Caldarium* and the *Laconicum*, where interpreting the 3D scans of the complex system of conducts and arcs has been extremely complicated.

Such as described, the pictures of others archaeological sites have supported for replicating objects partially disappeared in Lliria (Fig. 8).

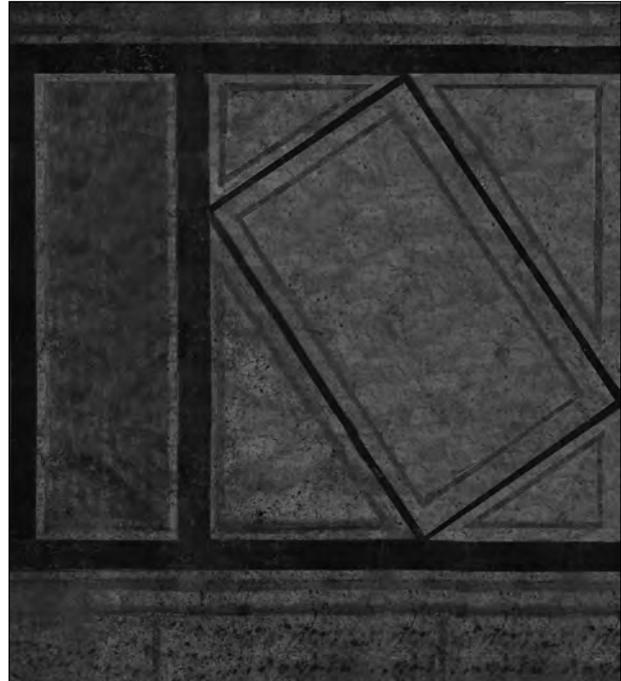


Figure 15. Image. Texture applied on the 3D model of the Basilica Thermarum. Hypothesis restored from the archaeological evidence.

Eventually, some high quality photos have been utilised to recover information about colours and patterns of surfaces, subsequently creating textures applied on the virtual Roman buildings (Fig. 14).

4.4 The mapping process

Normally, textures occupy an important role for generating 3D virtual models. In the video-game industry, they often replace the presence of solids in the scene to accelerate the rendering process.

In this reconstruction, the aim has mainly been the scientific elaboration of 3D buildings, highlighting the likely architecture more than its likely ancient colours. As a result, the materials and their included textures have been secondary in this simulation, even because, they mostly depend on the type of final rendering. Generally, they are also strictly related to the software used for producing the virtual images.

Therefore, the materials have been procedurals for most of the structures, which are prearranged by the software in terms of textures and physical

simulations. They are also directly adapted for the rendering engine and others virtual atmospheric conditions, such as: lights, reflections and refractions. Furthermore, some of the maps have been elaborated with the intent of reconstructing colours still preserved *insitu*; for instance, those of the columns or the virtual restoration of the paintings still conserved on the walls of the *Basilica Thermarum* (Figs 14 and 15).

5. Visualisations of the Virtual Roman Baths

The rendered images have not been produced as specific final products to be shown in the museum yet, since the project is still on-going. The final purpose is a public display using a transparent touch screen device. This will allow a virtual navigation in the archaeological site as currently is, together with the possibility of generating 3D buildings on the top of it, by means of the virtual reconstruction modelled and described in this paper. These elements will be supported with textual explanations of the area visited.

These will be located in the visitors' hall, in front of high window glasses displaying the real archaeological remains. The opening date of the museum is scheduled for March 2013.

The rendering process has been mainly arranged for testing the materials applied on the 3D model as well as testing the geometry. Several images have been calculated throughout the modelling step in Mental Ray, setting up illumination, exposure and other effects according to the type of images desired (Fig. 16, 17, 18).

The virtual reconstruction is dynamic, just like our interpretation of the past, thus it will always be ready for updating with new ideas or new information discovered. These could consist of virtual images associated to brochures of the museum, videos, and future real time rendering for extensive interactivity.

6. Analysing the Results

In this project, 3D Laser Scanning easily shows its importance as a tool for recording spatial archaeological information. Its efficiency has

been tested in the whole reconstruction process. Recovering measurements for restoring the ancient structures, making assumptions about the original location of the remains and finally, comparing the general data to make sure the reconstruction fits the real archaeological evidence.



Figure 16. Rendered Image. Virtual Reconstruction produced with Mental Ray in 3ds MAX. Southern internal view of the Caldarium.

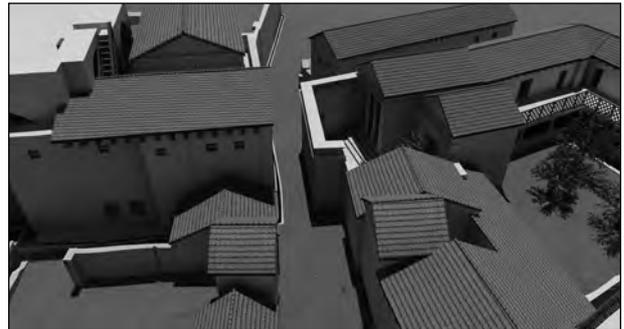


Figure 17. Rendered Image. Virtual Reconstruction produced with Mental Ray in 3ds MAX. Southern external view of the male Roman Baths and of the holy area.



Figure 18. Rendered Image. Virtual Reconstruction produced with Mental Ray in 3ds MAX. North-eastern external view of the male Roman Baths and of the holy area.

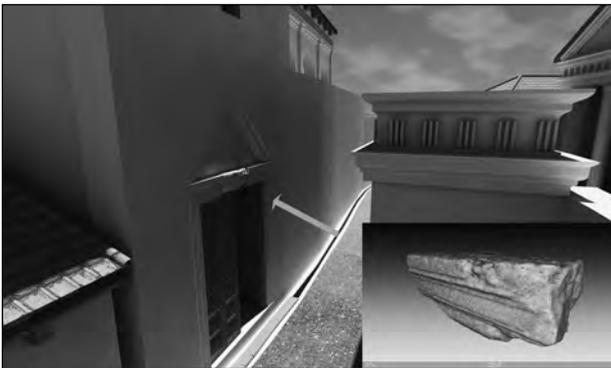


Figure 19. Rendered Image. Virtual Reconstruction produced with Mental Ray in 3ds MAX. Entrance to the Roman Baths. Hypothetic relocation of an element of the ancient frame, captured with 3D Laser Scanning. On the right, its textured view in MeshLab.

The point clouds are extremely accurate, depending on the 3D Laser Scanner and its settings. However, this precision implies negative aspects as well. Dust is quite often captured by the scanning equipment, causing anomalies in the architecture of certain objects. The archaeological site was not an excellent working area as excavation was still under way, causing dirt and dust to be scattered into the air, and on surrounding surfaces. Furthermore, atmospheric conditions, such as wind and rain, could be held responsible for disturbing the process. All these conditions can bring the production of noise in the point clouds, meaning more time in the laboratory creating more refined 3D meshes.

Traditional Total Stations can become tedious when there are numerous points which need to be recorded. 3D laser scanners can obtain these points with a fraction of the time and effort. The only issue with the use of a scanner in this project concerns the co-ordinate system of the general point cloud, as the scanner utilised does not provide an integrated Total Station. This problem was overcome by transforming the point clouds or 3D meshes using three points previously geo-referenced with the Trimble VX Spatial Station. This enables the ability to georeference 3D points with a minimum standard deviation of 10mm at $\leq 150\text{m}$.

The software currently in use permits an intense optimisation of the shapes but requires time, and can result in a loss of a small amount of data. This loss of data has encouraged the use of polylines that match the sections and plans of 3D



Figure 20. Rendered Image. Virtual Reconstruction produced with Mental Ray in 3ds MAX. The Basilica Thermarum and the Taberna. Hypothetic relocation of the base of a column, captured with 3D Laser Scanning. On the top, its textured view in MeshLab.

meshes. Millions of points can then be controlled and reduced to only a handful of data if needed, minimising the surfaces which are then present in a 3D model. This can reduce rendering time, size of files and increase performance on interactive software equipment.

The use of 3D laser scanning in this project has been important in the virtual reconstruction of the *hypocaustum*. The shapes of the remaining arches and their location have been utilised for restoring those that have collapsed. Virtual reconstructions of the arches are then included over the tops of the marks shown in the orthophoto, thus complementing and completing the laser scanned model (Fig. 12).

An interesting discovery in terms of visualisation has been the restoration of the small archaeological remains, such as the Tuscan order capitals of the columns located in the *Basilica Thermarum*. This operation has been hard to achieve with a small recorded capital showing a poorly conserved Corinthian order, due to the intricate pattern of the decorated acanthus leaves. The restoration of the surviving cornices has been treated in the same manner, by extruding the sections of their longitudinal curves previously recovered from the 3D meshes.

All the digital information has been correctly georeferenced, positioning the real coordinate point 30s 706458.96 4389550.93 100 UTM at the 0 0 0 UCS in 3ds MAX and processing on a one to one scale. Both 3D meshes and the 3D reconstructions

are comparable to each other. This technique has been successful in identifying zones or details left out during the reconstruction.

The parts in the virtual reconstruction that represent missing data are based on the epigraphic data and contemporary structures. This data was acquired by architects and archaeologists with specific skill sets. An accurate model was then reconstructed with principal ideas focussed upon an interactive virtual model which will be available to the general public. An example is shown in Figures 19 and 20, where the 3D meshes and the 3D objects have been superimposed onto original data in rendered images. This will assist in testing the hypothetical locations of the finds and also to arrange a likely way of visiting the museum.

The discussions among experts to achieve a correct restoration have significantly slowed down the modelling process, but this is essential to create an accurate reconstruction and to set appropriate measurements. Originally, the average height of the male Roman Bath was estimated to be 7.5m, but eventually it was decided a more accurate measurement of 12.5m. This is due to several reasons. These reasons include rendered images with the introduced daylight system, and most importantly, measurements from similar foundations in Perga.

The daylight system has been arranged at the pre-set time 12:00pm, while a Physical Sky has been used as a shader, with CIE Clear Sky as a virtual sky model to assist the calculation of the brightness in the scene. The closed perimeter of the *Basilica Thermarum* with only one door accessing the outside *Palestra* demonstrates that the diffusion of light inside of it had to be significantly low. This is due to the fact that the opened arches of the first floor had to be located at about 5.50m, above the upper pediment sustained by the columns. The next high walls of the *Apodyterium* on the east and the northern rooms located on the first floor of the Basilica had to further reduce the contribution of light in the ground floor. An example is shown in Figure 20.

According to the rendered images the global illumination of the other structures had to be affected, especially at floor level. In the *Apodytherium*, a semicircular window with a diameter of 7.70m has

been opened on the southern wall (Fig. 18), matching the size of the barrel vault and extending the shape of a previous hypothesis. Following a similar idea, trapezoidal windows have been adapted to the rendered environment of the *Laconicum* and the apse of the *Caldarium* (Fig. 16). However, even if the illumination system has somewhat influenced shapes and sizes of some details, it has occupied a secondary position in the reconstruction process.

7. Conclusions and Outlook

Recently, the utilisation of 3D Laser scanners for archaeological survey has increased, converting these devices into fundamental tools in the field. The digital storage of accurate 3D spatial information allows infinite further scientific analysis and application. 3D virtual reconstructions and simulations of the past shall still be developed, as they veil immeasurable potentialities for visually supporting the researchers and educating the public. The capability of 3D Laser scanning in archaeology and its ability to add more scientific appearances to 3D reconstructed ancient environments has been exhaustively tested throughout this project.

With regards to the Roman Buildings of *Edeta*, the area is still under research, implying difficulties for a correct virtual restoration of the structures, both in terms of new finds discovered and new assumptions made. Furthermore, there are still some areas which need to be excavated in the future, such as the female baths. This might bring new ideas and transform the proportions of the actual 3D model as well. As a consequence, the virtual reconstruction is a dynamic object, always ready to be reviewed and improved with new features.

To conclude, the “Comunitat Valenciana” is rich in traditions, history and art. The promotion of the knowledge of the village of Lliria and the dissemination of its worldwide cultural heritage by means of new graphic technologies is truly appreciable. The restoration of archaeological remain using virtual technologies is still one of the most effective ways for the public to understand the past. Digital reconstructions elaborated by the scientific community with an analytical approach, elevates the quality of the virtual immersion into the past. Researchers are allowed to make infinite and invasiveness modifications of the virtual structures.

Acknowledgments

This project has been financed and supported by the “Generalitat Valenciana – Conselleria de Turisme, Cultura i Esport”. We are grateful to all members of the Department of Architecture working on the project, especially to Elisa Moliner Cantos, Iván García Miñana and Patricia Calvo Cortés.

Fieldwork and laboratory analysis have been conducted by the “Museu Arqueològic de Lliria – MALL”, particularly by the archaeologists Carmen Martínez Camps and Xavier Vidal Ferrús. A special thanks to Carles Jordi Grau Giménez, who has collaborated with the interpretation of the plans of the Roman Bath.

The Graphic Department of Global Geomática has provided and elaborated the digital data. The photogrammetry and the 3D Laser Scanning have been produced thanks to the assistance of Rafael Tortosa García, Óscar Aparicio González, Luis Gimeno Martínez, David Martínez Imedio, Andrea Vilaplana Sellés and Josep Gimeno Marzal.

Finally, we acknowledge Tom Brughmans, Nicholas Martin Burrell and Elizabeth Richley for their collaboration in the editing process.

References

Alshawabkeh, Y. 2005. “Using Terrestrial Laser Scanning for the 3D Reconstruction of Petra /Jordan.” In *Photogrammetric Week '05*, edited by D. Fritsch, 39-47. Heidelberg: Wichmann Verlag. Accessed October 23, 2011. <http://www.eis.hu.edu.jo/Deanshipfiles/conf10633151.pdf>.

Alvaro, C., G. Liberotti, and D. Nepi. 2009. “Architectural analysis and 3D reconstruction: a case study of Leopoli – Cencelle in Italy.” In *CAA2009 proceedings*, Williamsburg, Virginia, USA, 2009, edited by B. Frischer, J. Webb Crawford and D. Koller, 1-9. Oxford: Archaeopress. Accessed October 23, 2011. <http://www.caa2009.org/>

[articles/Alvaro_Liberotti_Nepi.pdf](#).

Connolly, P. 1994. *Pompeii: The Roman World*. Oxford: Oxford University Press.

Earl, G., V. Porcelli, P. Constantinos, G. Beale, M. Harrison, P. Hembo, and S. Keay. *In press*. “Formal and informal analysis of rendered space: The Basilica Portuense.” In *Computational Approaches to Archaeology*, edited by A. Bevan and M. Lake. London: University College London Press.

Escrivá Torres, V., and X. Vidal Ferrús. 1995. “La partida de Mura (Lliria, Valencia). Un conjunto monumental de época Flavia.” In *Saguntum: Papeles del Laboratorio de Arqueología de Valencia (P.L.A.V.)* 29, 231-239. Valencia: Universitat de Valencia.

Giuntoli, S., and M. Martinelli. 1989. *Arte e Historia de Pompeya*. Florence: Bonechi.

Kampel, M., and R. Sablatnig. 2011. “Profile based Pottery Reconstruction.” In *IEEE/CVPR Workshop on Applications of Computer Vision in Archaeology. Madison (USA), 2003*. Accessed October 23, 2011. <http://www.caa.tuwien.ac.at/cvl/people/sab/papers/ACVA03.pdf>.

Lock, G. 2003. *Using computers in archaeology: towards virtual pasts*. London: Routledge.

Patay-Horváth, A. 2011. “The complete Virtual 3D Reconstruction of the East Pediment of the Temple of Zeus at Olympia.” In *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences XXXVIII*. Accessed February 16, 2012. <http://www.isprs.org/proceedings/XXXVIII/5-W16/pdf/patay.pdf>.

Vitruvius Pollio, M. 1761. “De Architectura”. In *Compendio de los diez libros de arquitectura de Vitrubio*, edited by C. Perrault. Translated by Joseph Castañeda. Paris: Real Academia de las Ciencias de París. Accessed September 13, 2011. http://www.cehopu.cedex.es/img/bibliotecaD/1761_C_Perrault_Los_diez_libros_de_arqu_de_Vitruvio.

Reconstructing the Baths of Caracalla

Taylor A. Oetelaar

Univeristy of Calgary, Canada

Abstract:

The Baths of Caracalla are a large public bathing complex or thermae built between 212 and 216 CE and the building itself covers an area of approximately 2.4 hectares. The initial purpose of this reconstruction was to recreate the caldarium – a room with hot baths – for a thermofluid analysis to determine its thermal environment. However, this quickly morphed into an interesting, yet difficult, reconstruction of the entire bathing complex for Rome Reborn. The reconstruction of the Baths of Caracalla required careful attention to many different facets, from room dimensions to mosaic patterns. Even though much of the structure still exists, the published, publically available data were surprisingly sparse. Given the complexity of the Baths, the data gaps and the need for an engineering-based geometry, the best option was to use two software suites – one for the geometry and the second for the application of textures.

Keywords:

Digital Reconstruction, Baths Of Caracalla, Parametric Modeling

1. Introduction

The digital reconstruction of the Baths of Caracalla began because I required an accurate representation of the air volume inside the *caldarium* or room with hot baths to create a mesh for a computational investigation of its thermal environment. This, however, quickly evolved into an interesting, yet difficult, reconstruction of the entire bathing complex for the Institute for Advanced Technology in the Humanities (IATH) project, *Rome Reborn*; the goal of which is to rebuild the city of Rome digitally as it existed in the fourth century CE. The baths' reconstruction requires careful attention to many different facets, from room dimensions to mosaic patterns. For those unfamiliar with the site, the Baths of Caracalla is a large 2.4 hectare complex in the middle of an 8.9 hectare walled garden in southeast Rome built between 212 and 216 CE. It included exercise rooms, saunas, a swimming pool, as well as three rooms with varying temperatures of bath water. Even though much of the structure still exists, the publically available data is surprisingly sparse. Given the complexity of the Baths, I decided to use a parametric modeling suite which necessitated the use of a second program for the application of textures. This paper provides a brief methodological overview of this reconstruction, the archaeological data as well as on-site observations from 2008. In so doing, the paper will make apparent the major benefits and stumbling blocks of parametrically modeling the Baths of Caracalla.

Corresponding author: taylor.a.oetelaar@gmail.com

2. Software

Because of its parametric capabilities, I chose to use the engineering computer-aided design (CAD) program, Dassault Systèmes' SolidWorks, for the geometrical portion of the reconstruction. In parametric design, as the name suggests, the researcher creates features by drawing shapes and defining modifiable parameters to hold the shapes in place. This allows for easy manipulation of pieces of the design if new information becomes available. However, the program is based on mathematical formulae, which limits the creativity of the researcher. Since parametric programs are based on mathematical formulae, forming intricate curves, such as a Corinthian capital, can be time-consuming and frustrating if not impossible. This is why parametric programs are almost always engineering-based. Therefore, for the more artistic finishing details, such as texture application, I used Autodesk's 3ds MAX.

3. Geometry

The catalyst for this reconstruction was Janet Delaine's exhaustive study (1997), which included the floor plan (Fig 1). As one can see, the building is very symmetric about an 'axis' of four central rooms. In an appendix she lists measurements of wall lengths, door widths and heights, window widths and heights, window sill heights and niche widths and heights of most rooms. However, the data are incomplete. In some cases, DeLaine could

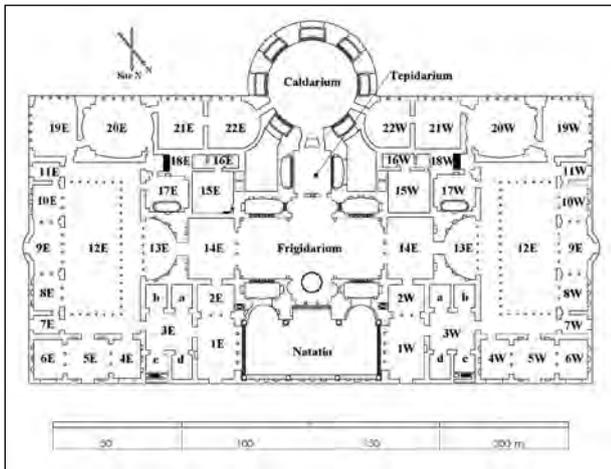


Figure 1. Floor plan of the Baths of Caracalla.

not give a dimension of a particular feature, either due to issues of access or preservation. For example, the N window of Room 1E is missing but DeLaine listed the N window dimensions of Room 1W so, based on the symmetry, I used the same dimensions for both windows. Even excluding the unattainable dimensions, her catalogue, though extensive, is not inclusive. Her most noticeable omissions are the placement of the doors, windows, and niches in the walls. Fortunately, for these, I measured DeLaine's scale drawings and came up with approximate dimensions.

The first step in this process was the room-by-room extrusion of the walls from sketches of their floor plan. However, to make the process easier, I modified the rooms from quadrilateral to rectangular – meaning all corners met at ninety degrees – by averaging opposing wall lengths. In most cases, the difference was two or three centimetres on a wall measuring up to twenty metres, a relatively insignificant amount. Parallel walls also allowed a simple extrusion to create the vaults instead of using a complicated, somewhat unpredictable loft. I completed the walls by adding the niche sills and lintels. This slide shows the reconstruction after the extrusion of the walls. Subsequently I added the door lintels, the window sills, and window lintels.

Rooms 15E/W and 16E/W required special attention because the data are particularly sparse due to access limitations. Rooms 15E/W are open-aired and 16E/W are at least two-storeys, possibly three, but, based on the comparatively



Figure 2. Picture of the east wall of 15W showing the outline of stairs (Photo credit: Joy Oetelaar).

small doorways, both were most likely reserved for servants and maintenance people. Before visiting the site in 2008, I had based my reconstruction largely on fragments contained within DeLaine's scale drawings. However, on site, I noticed the remains of a hanging staircase on the east wall of 15W (Fig 2) as well as a very narrow opening in the south wall of the opening with stairs leading upward which completely changed my opinion of these rooms. This is where one of the major benefits of parametric design becomes apparent. I was able to work on the interior of these two rooms without modifying the general superstructure or any of the adjoining rooms. I still think, however, that there is a missing aspect to the recreation of these rooms that requires further research. The reconstruction of these four rooms is tenuous and highly speculative.

The roofs were the next obstacle. While the vaulting was straightforward, the topsides were problematic, not because of a geometrical concern, but because the presence of rooftop terraces is ambiguous. Some reconstructions have gabled roofs on many of the major rooms while others (including DeLaine's) have all roofs as being flat. This reconstruction was somewhere in the middle. I observed remnants of staircases above 11W and from 18W to the roof of the portico above 12W while on site in 2008, which is corroborated by Brödner (1951). This, when combined with the lightwells in the staircase on the north wall of Room 3c, suggest that there were terraces on Rooms 4E/W-11E/W and on the top of the portico of Room 12E/W.

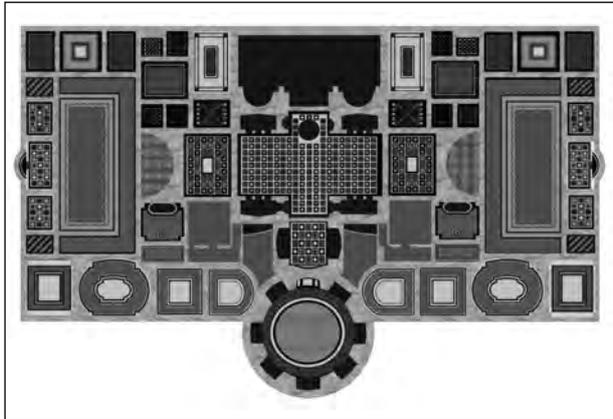


Figure 3. Reconstructed mosaic floor tiling of the Baths of Caracalla.

4. Texturing

With the geometry complete, the next step in the reconstruction was texturing. This is where the major drawback of *SolidWorks* becomes apparent. As an engineering design program, any models look metallic and the program's rendering capability is limited. Therefore, as mentioned earlier, I chose *3ds MAX* to apply the textures. The transition process, however, was not easy as *3ds MAX* cannot open *SolidWorks* files directly and IGES – a CAD interchange file format – did not provide adequate results. I found the best alternative was to open the *SolidWorks* files directly in another program, McNeel's *Rhino 4*, and saving the model as a 3D Studio file. In essence, I required a tertiary program to ensure the best results. The problem then became the large mesh size and the limitations associated with high memory demands. The intricacy of the Baths of Caracalla turned into a drawback because creating the mesh of the entire building required more computer memory than I had available. Instead of importing the entire model, it was better to import groups – superstructure, frames, glazing, mosaic floors, subfloor, columns, terrace and tiled roofs, and pool water. This also made applying the texture simpler because, for example, applying the bronze texture to all the frames at once is easier than applying it to each frame individually. To create the mesh of the superstructure, though, I had to break the complex down further into three blocks – the two wings and the central rooms. Since their coordinate system origins were the same, the reassembly of the groups was a simple importation.

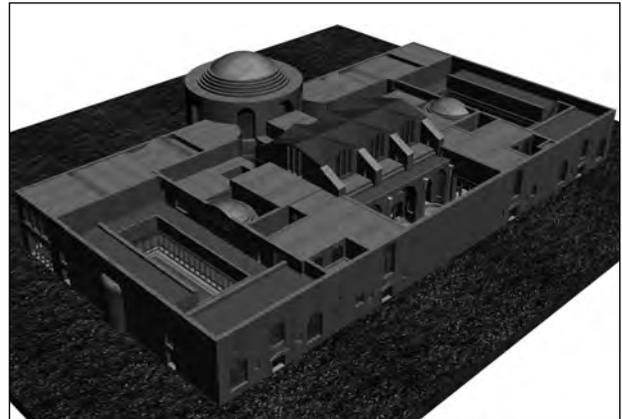


Figure 4. Reconstructed view of the Baths of Caracalla.

Because the decorative scheme is largely unknown and reconstructing it requires specialized expertise, the textures themselves were generic. However, I did attempt a reconstruction of the mosaic floors. I used drawings from Blouet (1828). I only had access to the microform version which was black and white so I extrapolated the colour scheme based on the fragments of mosaic still in situ, however, the assumption that the colour scheme remains similar throughout the baths is questionable as evident by the remnants. Room 3 and its subsidiaries (a-d) were all black with white detailing while Room 12 had white, tan, green, and maroon with some sea green accents. This interpretation used the colours of Room 12 to try to ensure some degree of continuity (Fig 3).

5. eDrawings

As shown previously, parametric design provides both advantages and drawbacks. One additional benefit of using *SolidWorks* specifically is Dassault Systèmes' eDrawings program. eDrawings is a free downloadable program which allows the user to view and manipulate files created in *SolidWorks*. One can rotate parts, pan, zoom in or out, and take select cross-section views. These allow the user to view a particular model from almost any vantage point which is particularly useful when disseminating information to a lay audience. However, one of the most valuable features for scholarly applications is the ability to measure anything on the model. With this, one can share a catalog of dimensions in the context of the overall building with one file. This eliminates some of the uncertainty associated with many tabular catalogs, including DeLaine's, as they

need more dimensions than they logistically have room for. With the ability to take any dimension, the onus then falls on the creator of the model to ensure accuracy. The main downside of eDrawings is that, since it was designed for engineering use, its rendering capabilities are limited so texturing is not advised. However, a monochrome model can efficiently show the geometry and, by using SolidWorks, eDrawings provides another way to easily and cost-effectively share raw data with both other scholars and the general public.

6. Conclusions

This digital reconstruction of the Baths of Caracalla (Fig 4) relied heavily on the work done by DeLaine. Her composite of dimensions and scale drawings provide its foundation. However, there were perplexing aspects and missing elements which necessitated assumptions and, because of these, the recreation is not static. The data surrounding

Rooms 15E/W and 16E/W were incomplete and problematic thereby making their recreation highly speculative. Furthermore, the decorative scheme as it stands now does not truly reflect the opulence of the Baths as they existed in antiquity. Hopefully, though, this model still gives the impression of immensity that an Imperial *thermae* possessed.

References

- Blouet, A. 1828. *Restauration des Thermes d'Antonin Caracalla, Rome*. Paris: Didot.
- Brödner, E. 1951. *Untersuchungen an den Caracallathemen*. Berlin: Walter de Gruyter & Co.
- DeLaine, J. 1997. "The Baths of Caracalla : A study in the design, construction, and economics of large-scale building projects in Imperial Rome." *Journal of Roman Archaeology Supplementary Series* 25.

Instruments and Methods for the Survey and Analysis of Amphitheatres

Martina Ballarin, Francesco Guerra

Università IUAV di Venezia, Italy

Luigi Sperti

Università Ca' Foscari di Venezia, Italy

Abstract:

For many years the Laboratory of Photogrammetry of IUAV University has been experimenting instruments and methods for the photogrammetric and laser scanning survey in the archaeological field. The laboratory was born for the study of architecture, but it also found a logical application field in the survey of archaeological architecture, that is those ancient monuments of which we still have remarkable parts in elevation. The amphitheatre is one of the typologies we studied through several surveying campaigns. The surveys undertaken until now and the ones that we are still planning will also provide data to aliment a particularly fertile research project of our University concerning the shape of the amphitheatres. In the paper we will present the results of the researches conducted on the amphitheatres of Verona and Grumentum, with particular attention to the methodologies used and the new applications currently in progress using UAVs on some other amphitheatres.

Keywords:

Amphitheatre, Laser Scanner, Aerial Photogrammetry, UAV

1. Introduction

A natural instrument of gaining knowledge is measurement; it is an action that associates quantity information with quality information in order to define a shared and intersubjective knowledge of the real environment (Achille and Monti 2001, 78-90).

Many of the numerous on-going and past research activities have had, and still have, the purpose of studying, exploring and developing the techniques for dimensional survey and 3D reconstruction of objects and surfaces. Special attention has been given to the metrological aspects of the problem, by comparing the "traditional" survey methods and more recent applications.

With the purpose of achieving the most accurate and reliable understanding of the works of art, we found the possibility of using optical technology for studying cultural objects, which were explored in recent years thanks to the progress made in the computing and optical electronics fields. In particular, we have witnessed the rapid development of 3D optical scanning technologies to understand the form of highly complex structures,

Corresponding author: martinaballarin@hotmail.it

a development that has confirmed its effectiveness as a method for analysing and conserving Cultural Heritage (Bitelli 2002; Peloso 2005). Also, in the archaeological field the use of these techniques is developing very fast, in spite of their costs. In fact, often archaeologists feel the need for a fast and non-invasive documentation of objects, artefacts and sites.

For years, the Laboratory of Photogrammetry of the System of Laboratories that belongs to IUAV University in Venice has been working in the archaeological field, trying to apply the methodologies of the Geomatics to Cultural Heritage, and archaeology in particular. On the other side, the Laboratory has been involved in the experimenting of instruments and methods per se, always trying to keep up with the latest technological innovations and evaluating them in terms of operability and feasibility, as in this field it is very important to maintain a tight connection to the practice.

This paper concerns some results of a several year work of the Laboratory of Photogrammetry. We will present two case studies set in two different parts of Italy: the Arena of Verona, in Northern Italy, and the Amphitheatre of Grumentum, in Southern

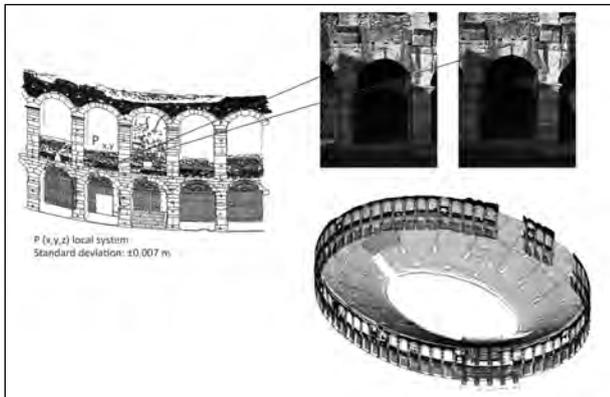


Figure 1. Close range photogrammetry for the Arena.

Italy. They represent a methodological path, from the traditional photogrammetric survey to the new image modelling software products, through laser scanner and aerial photography. These two examples are particularly significant because, even if they belong to the same typology, they are two completely different monuments, as one is still entirely standing, and the other one is just ruins.

2. The Arena of Verona

The Arena of Verona chronologically is the first of the two surveys the Laboratory performed. We started the survey campaign in a time when analytic close range photogrammetry was the standard method. Soon, when the laser scanner started to develop, we used the Arena as an application field to check the validity of this new technique. In fact, the survey of the external part of the building went on for seven years, from 1998 till 2004. In 2008, it ended with the acquisition of the ground level and underground tunnels.

As a first thing, we realised a main topographic net that covered the whole Arena, in order to fix a reference system also for all the following surveys. This net was composed by 16 vertices, distributed in the three dimensions, and was the point of departure of a second net, used as a support for the photogrammetric survey. As it is well known, close range photogrammetry is the technique through which we can reconstruct 3D coordinates from at least a couple of images (Fig. 1). In particular, these two photographs have to be taken from two different points in such a way that the same region is shown on both photographs. Therefore, they have to overlap by at least 60% and they have to be as



Figure 2. Special "tools" used for the acquisition of the images.

orthogonal as possible. The topographic net was also used here to acquire the coordinates of points in order to orient the images (Balletti et al. 1999).

The special conformation of the place and sometimes the difficulties concerning the operations required particular instruments and methods. In order to obtain the best results, we tried to acquire the photograms in the best possible way, which involved two special "tools" (Fig. 2):

- Boom truck to have photogrammetric images as orthogonal as possible, even for the highest areas;
- Other types of elevation systems to reach the highest areas also inside the vaulted parts of the Arena.

In the application field of the Arena of Verona, we used all the latest developments of digital photogrammetry that means not just the stereoscopic restitution of digital images, auto correlation, automatic DEM, profile extraction, etc., but also geometric elaboration of digital images – from rectification to orthoimages and 3d models (Balletti et al. 1999).

In 2004, after the Laboratory acquired the laser scanner Riegl LMS-Z360i, the arches of the Arena seemed like the perfect application field to evaluate the opportunities given by this instrument. The purpose of the survey was to confront the products of the laser scanner with those of photogrammetry. The main problem, from the point of view of the applicability, was to understand whether the laser scanner could give us results that were as good

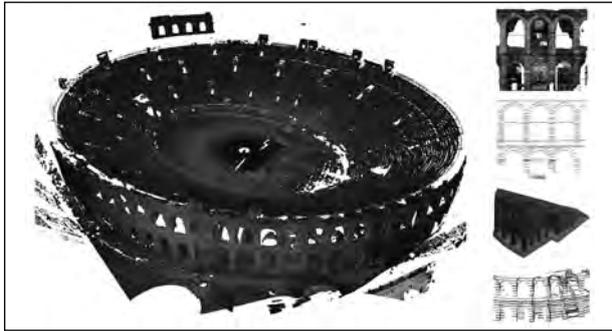


Figure 3. Some of the products obtained by the laser scanner.

as the ones obtained by traditional close range photogrammetry, using less resources.

Close range photogrammetry for a monument like the Arena was a well-known technique: we knew the schedules, the workflow, the costs and the results we could obtain. The laser scanner was instead a new technique and there were not a big number of case studies so that we could plan an efficient survey (Guerra et al. 2004).

The results made clear that laser scanners had great potential, as they could acquire 3D data of structures of any complexity in a fast and non-invasive way. The point clouds produced by these tools are almost continuous data, through which we obtain sufficient information for the elaborations. During post-processing we can achieve documents such as: profile extraction, cross-section, 3D views, 3d models, and even orthophotos, thanks to the digital calibrated camera mounted on the scanner (Fig. 3).

Moreover, the introduction of laser scanners brought another benefit in the surveyor's job: the separation of competences between the acquisition and processing phases and the restitution one. This last one can now be easily left to the scholar that needs the representation, as it is easier and it does not require specialised technicians.

These applicative experiments lead us to conclude that laser scanning is best viewed as a survey technique that does not replace the others, but incorporates them: to achieve a "geometrically" accurate survey, the surveyor must use topography, photogrammetry, direct survey and laser scanning, each used according to its characteristics.

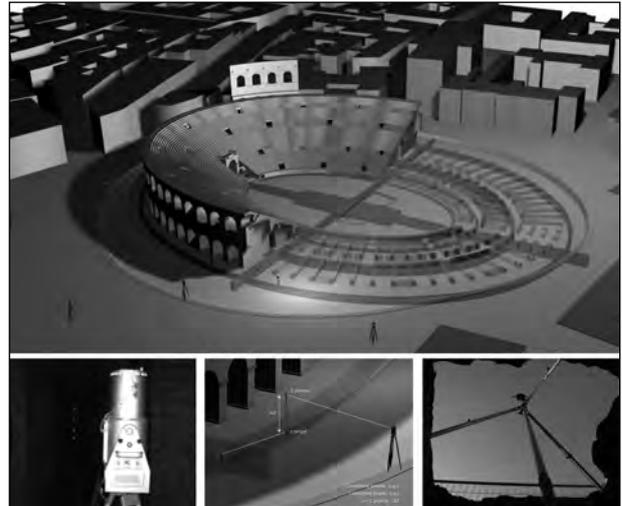


Figure 4. The survey of the underground tunnels of the Arena.

Once we had realised how this tool could be used and its strength and weaknesses, we used it for a new application field where the standard methods could not be utilised: that is the underground tunnels below the Arena. Here, we had to operate in a narrow, dark space, with a constant presence of water; there was no visibility between topographic stations, and so topography could not be an option.

The survey of the main axes and the external tunnel was realised with a laser scanner, through 55 scans that were then registered in the same reference system. However, the biggest problem was the registration of the measurements in the reference system of the surface, in order to link it to the rest of the survey. We overcame this problem by using prism poles with reflective prisms – collocated in the external concrete sumps – that were long enough to link the underground areas to the topographic net on the surface. In this way, at the ground level we used topography, inside the tunnels we used the laser scanner and we collimated the same prism pole both with topography and with laser scanner, so that the coordinates of the two points were the same part from a Delta Z that we could easily calculate (Fig. 4).

3. The Amphitheatre of Grumentum

The second case study is the Amphitheatre of Grumentum (Fig. 5), where we performed a terrestrial laser scanner and a photogrammetric survey, which included aerial photos captured

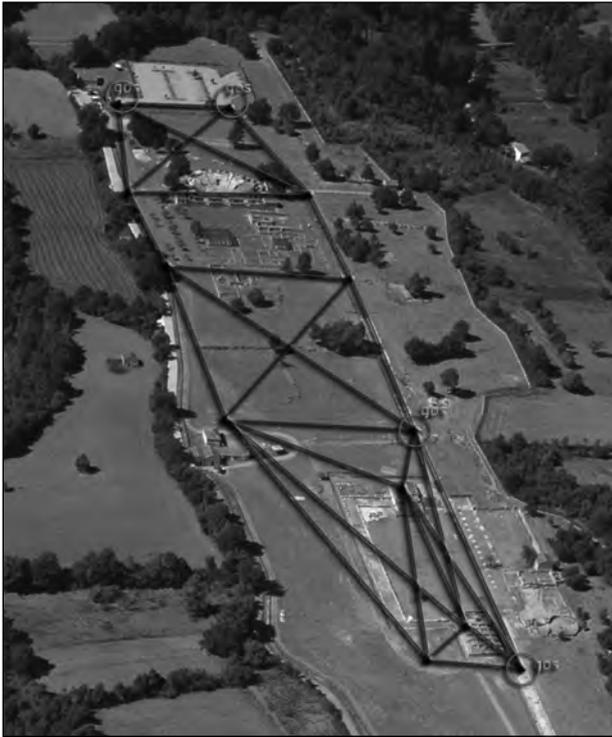


Figure 5. *The site of Grumentum.*

from a remote controlled aircraft. The purpose of the survey procedures we followed was to integrate the laser scanner data with those resulting from the photogrammetric survey, in particular for the horizontal portions, such as excavated parts and the collapsed sectors that are very common in archaeology.

A big part of the survey campaign was dedicated to the analysis of the amphitheatre. Here, photogrammetry was used in different ways, but we privileged the choice of aerial images, because they often facilitate the reading of both site and buildings, being taken from a favoured position: the zenith.

During the first campaign (summer 2005) we decided to use moored balloons to perform a photogrammetric survey with high resolution calibrated digital cameras (Fig. 6), because at the time it surely was one of the least invasive and most controllable methods (Balletti, Pillon and Sartorelli 2007).

The system was composed of a helium aerostat constrained by some cables, in order to control it from the ground, and a radio remote control system



Figure 6. *The moored balloon.*



Figure 7. *Aerial Images of the amphitheatre.*

to acquire the images we needed. At the bottom of the balloon we collocated a swivel camera mount, a camera (a Fuji S3, 12 Megapixel) with calibrated lenses (50 e 20mm) and a video transmitter on the aircraft in order to know what the camera was displaying when it was taking a picture.

As the areas we wanted to survey (forum and amphitheatre) were very big, we planned a flight in order to obtain images on two scales: 1:50 and 1:100. The final results of this research were remarkably interesting, as we were able to produce a mapped 3D model, aerial ortophotos and photoplans of the wall surfaces (Balletti et al. 2006) (Fig. 7).

However, this experience highlighted some logistic problems connected to the management of the balloon. In fact, the limits of this system are strictly connected to the weather conditions in which the surveyor has to work. In particular, wind conditions affect not only the trajectory the balloon has to follow, but also –and mostly– the flight altitude. This caused a last-minute change of plans: originally we wanted to take the images at two different altitudes, 50 m and 20 m, in order to have different resolutions, but the highest altitude we could reach, maintaining the control and stability of the balloon, was 30 m.



Figure 8. Point cloud of the amphitheatre and aerial image acquired by the plane.

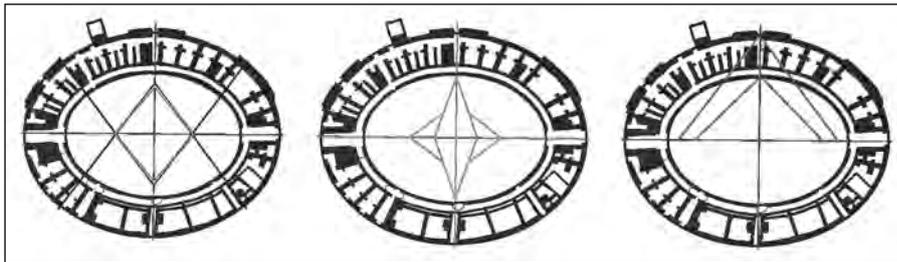


Figure 9. Construction of the amphitheatres: oval with 4 centres; oval with 8 centres; ellipse.

In the second campaign (summer 2006) we integrated the previous aerial survey with another one, which required the use of stereo couples acquired by a small two-seated touristic plane from a 200 m of elevation. Even if the plane was not specifically equipped for aerial photogrammetry (as we used an ultra light aircraft) we were still able to perform an aerial survey with almost vertical photogrammetric strip-runs, in the manner of traditional photogrammetric aerial survey.

These two applications of stereoscopic photogrammetry demonstrated the flexibility of these techniques: even if acquired with different means of transport (moored balloon and ultra light aircraft) the products obtained were easily integrated.

For Grumentum too, we performed a laser scanning survey, which had demonstrated before its potentials in the acquisition of large areas. For this tool, we thought of specific applications in order to obtain the best results both for the documentation at a medium scale and for the survey of the details (Fig. 8).

The laser scanner used for the survey of the amphitheatre was –again– the Reigl LMS-Z360i integrated with digital camera Nikon D100. In the areas we were most interested in, we tried to perform scans that were as orthogonal as possible,

in order to reduce the noise of the point cloud and to obtain a homogeneous distribution of points. The point clouds were then registered together using the ground control points surveyed with the topography and tie points distributed inside the scan area (Balletti, Pillon and Sartorelli 2007).

Unlike the Arena, in Grumentum the laser scanner was already a well-known technique, of which we knew times, costs and possible results. The innovation of this survey was the integration of this technique with aerial photogrammetry. As for the latter, we felt the necessity of a methodology or, even better, of tools that allowed us to obtain suitable products for a medium scale project, with low costs and high operation speed. The moored balloon was the first answer to these needs, even with the limits we have already underlined. Hereinafter, we will show how we are now trying to overcome the difficulties in the operations connected to the management of the aircraft.

4. Metrical Analysis

Using the data collected via these operations both the amphitheatres were analysed, following the path of a particularly fertile research project of our University that studies the shape of these monuments. The main developer of this research is prof. Camillo Trevisan, associate professor at the same university.



Figure 10. Some UAVs.

Prof. Trevisan proposed a method to analyse the curves of the *cavea* and the *arena* of those amphitheatres (Trevisan 1999). First, the points of every curve are interpolated with ellipses and ovals, which are polycentric curves constituted, in this case, of 4 and 8 centres; then, for all of them the standard deviation is calculated (Fig. 9).

The first construction (the one with 4 centres) was soon rejected, as curve and points did not interpolate very well. Also, the distribution of the standard deviation was very important to determinate the invalidity of this construction: the surveyed points were distributed alternatively inside and outside this curve. In fact, we cannot expect to have a perfect interpolation for many reasons: the original small deviations of stones from the plan; the deformations due to the degenerative effect of time on the structures, such as the damaging of the buildings; the adjustments made during the construction itself. However, here the distribution of these errors was so clear and regular that it allowed us to refuse this theory.

The ellipse and ovals with 8 centres interpolated the curves much better, and the standard deviation's distribution was more random. This allowed us to say that one of these two geometric figures was used for the construction of these buildings. Although we still cannot be certain, the oval with 8 centres is more plausible, as ellipses are very hard to draw, especially in the construction site. In fact, in order to build an amphitheatre, Romans architects needed to create two lines (one for the *cavea* and one for the *arena*) always equidistant from each other. The drawing of the first one was quite simple, as they just needed to tie a rope on the two foci and follow the perimeter design by it (the sum of the distances from any point on the ellipse to those two foci is constant). To draw the second one, though, they needed the equation

of the ellipse, and for that the Cartesian plane and some knowledge of analytic geometry that they still did not have. Moreover, the analysis of the partition walls showed that they all converge towards the hypothetical centres of a polycentric figure, more than the foci of an ellipse.

Finally, we can say that the curve of the amphitheatre is a curve with continuity, which is a smooth line without kinks or breaks. This means that the left and right tangents of every point are the same and it can say something about the construction processes as well. In fact, the ellipse has a line that changes its curvature with great continuity; the oval, instead, has a constant curvature that suddenly changes when the centre of the arch changes. However, here the arches result welded together, as they have the same tangent. This result is obtained by making sure that the radii of the two circles overlap exactly when the centre changes. For this reason, by these points the curves look squeezed, and sometimes this characteristic makes the ovals unpleasant. This might explain the efforts to draw ovals that looked as similar as possible to the ellipses.

5. Conclusions and future perspectives

For the future we want to focus our attention on new data acquisition systems. We want to test the applicability of new aerial photogrammetry technologies to architectonic and archaeological surveys. These techniques use mobile cameras installed on remote-controlled aircraft (UAVs) to register vertical or oblique digital images at a low altitude, in order to attain a very high resolution on the object that has to be surveyed (Fig. 10).

The research will concern the application, on the photos, of the orientation and multi-image

restitution algorithms that nowadays are commonly used instead of laser scanning. Also, in order to contain the weight of the instruments transported, we will use compact cameras that guarantee a high spatial resolution.

Once again, then, we will compare well-established techniques with the new experimentations allowed by recent technological developments. On one side, there will be the classic aerial photogrammetry, which will consist in the creation of a definite flight path, with strip-runs as vertical as possible, and overlapping photograms. On the other side, there will be those new software products that by now allow us to obtain good results even with blocks of images that do not have a clear geometry, being taken from “random” positions. This new development creates new possibilities in terms of versatility, but the accuracy of the results is still not clear.

The same can be said about cameras used for these surveys. Traditional aerial photogrammetry used cameras that cannot be utilised on these aircrafts, as the payload of micro Unmanned Aerial Vehicles is usually quite small. We will then have to find the good balance among weights, resolution, sharpness, etc., in order to obtain adequate results, with compact, light, low cost cameras.

In other words, the purpose will be to verify whether the fundamental components of photogrammetry affect significantly the surveying accuracy or not. Well-established techniques, for which the possible outcomes are known, will be compared to new ones, more versatile, but still at a testing phase, pursuing the methodological path that the Laboratory has been following for the past decades.

References

Achille, C., and C. Monti. 2001. *Nuove metodologie di rilievo. Tecniche di rilevamento. Nuove frontiere delle tecnologie*, edited by R. A. Genovese, 78-90. Napoli: Edizioni Scientifiche Italiane.

Balletti, C., F. Guerra, and L. Pilot. 1999. “Experimental Survey of the Arena of Verona.” In *CIPA XVI International Symposium Proceedings, 3-6 October 1999, Olinda, Brasil*. Online Proceedings: <http://cipa.icomos.org/fileadmin/template/doc/olinda/99c306.pdf>.

Balletti, C., F. Guerra, and D. Miniutti. 1999. “The Surfaces of the Arena of Verona: representations.” In *Photogrammetric measurement, object modeling and documentation in architecture and industry, Volume XXXII, Atti del 18° meeting ISPRS, Salonico, 1999*, edited by P. Patias, 16-22. Thessaloniki: The Aristotle University of Thessaloniki.

Balletti, C., F. Guerra, and L. Pilot. 2006. “Integrazione di tecnologie 3d multiscala per la conoscenza, l’interpretazione e la diffusione di un sito archeologico.” In *X Conferenza Nazionale ASITA, Federazione delle Associazioni Scientifiche per le Informazioni Territoriali ed Ambientali, 14-17 novembre, Fiera di Bolzano*. Online Proceedings: http://circe.iuav.it/labfot/pubblicazioni/01_balletti_guerra_pilot.pdf.

Balletti, C., M. Pillon, and L. Sartorelli. 2007. “L’arena di Grumentum: Misura, Geometria, Forma.” In *E-ArCom07 Workshop Proceedings, “Archivi e musei di architettura, Tutela e promozione del patrimonio, Design e progetto di architettura. Le nuove frontiere della rappresentazione 3D” 17-19 maggio, CLUA Edizioni, Ancona*, 66-71. Firenze: Alinea Editrice.

Bitelli, G. 2002. “Moderne tecniche e strumentazioni per il rilievo dei Beni Culturali.” In *VI ASITA Conference Proceedings, vol. 1, IX-XXIV*, Varese: Artestampa sas.

Guerra, F., L. Pilot, and P. Vernier. 2004. “Fotogrammetria e laser scanning: alternativa?” In *E-ArCom04 Workshop Proceedings, “Tecnologie per comunicare l’architettura”, 20-22 maggio, CLUA Edizioni, Ancona*, edited by E. S. Malinverni, 356-361. Ancona: Clua Edizioni Ancona.

Peloso, D. 2005. “Tecniche laser scanner per il rilievo dei Beni Culturali.” *Archeologia e Calcolatori XVI*: 199-224.

Trevisan, C. 1999. “Sullo schema geometrico costruttivo degli anfiteatri romani: gli esempi del Colosseo e dell’Arena di Verona.” *Disegnare* 18-19: 117-131.

Why Hunter and Gatherers did not Die More Often? Simulating Prehistoric Decision Making

Florencia Del Castillo and Juan A. Barceló
Autonomous University of Barcelona, Spain

Abstract:

We have created a computer model of economic and social cooperation among hunter-gatherers that simulates how prehistoric people survived. This is an agent-based model in which agents simulating hunter-gatherer households move through a territory for resources irregularly distributed. The probabilities of their survival not only depend on the availability of resources but on the probability to establish cooperation links with other agents in such a way that the quantity of labor is enhanced and the probability of success in hunting increases. We expect to be able to discern if survival is conditioned by social decisions only, or if it is the result of the constraints on mobility generated by geography and the irregular distribution of resources, both in space and time. Results suggest that small sized groups (less than 10 families) died by starving when agents could not build a high enough number of social ties with other groups. Cooperation appears to be filtered by the specific social (cultural) identity of agents, a parameter that changes constantly, because it is being negotiated at run time as a consequence of previous stances of successive cooperation.

Keywords:

Hunter-Gatherer, Economic Decision Making, Social Reproduction, Computer Simulation, Agent-based

1. Towards a Probabilistic Model of Economic Decision-Making in Prehistory

Prehistoric hunter-gatherers have been studied many times from the point of view of animal foraging behavior, stating that human agents also forage in such a way as to maximize their net energy intake per unit time. In other words, it is assumed they should find, capture and consume food containing the most calories while expending the least amount of time possible in doing so. The understanding of many ecological concepts such as adaptation, energy flow and competition hinges on the ability to comprehend what food items such human agents selected, and why (Grove 2009, Winterhander and Smith 1981, Smith 1983). If such an assumption were true, we would say that hunter-gatherers survival would depend just on the availability of resources, given the adaptive nature of economic behavior.

We have created a hypothetical model in which social agents survive only if they have success in acquiring subsistence available in the environment

Corresponding author: florenciadecastillo@hotmail.com

by hunting. In the case of gathering, we can assume that posterior probabilities for gathering success, and hence of survival, are completely defined by the probability of plants availability (Fig. 1). These prior probabilities are the probabilities of the influence plant availability may have on survival before any evidence is entered about a concrete historical situation. That is, prior probability of individual survival equals life expectancy at the moment of birth. In case the environment is full of available resources ("rich world hypothesis"), the probability of finding enough plants to eat and make instruments is very high, and prior probabilities for survival are also high. Once the influence of additional factors has been entered in the model, all the probabilities are called *a posteriori*. Technically speaking, it is the marginal posterior probability distribution of the hypothesis, given the available evidence and the factors dependencies (the model) (Oaksford and Chater 2007, Singpurwalla 2006). This follows directly from standard probabilities theory (Darwiche 2009, Koller and Friedman 2009, Neapolitan 2003). Then, the probability of survival can be computed from the probability of gathering success conditioned on the availability of plants and territorial mobility (external factors), multiplied by what we know already about the probabilities

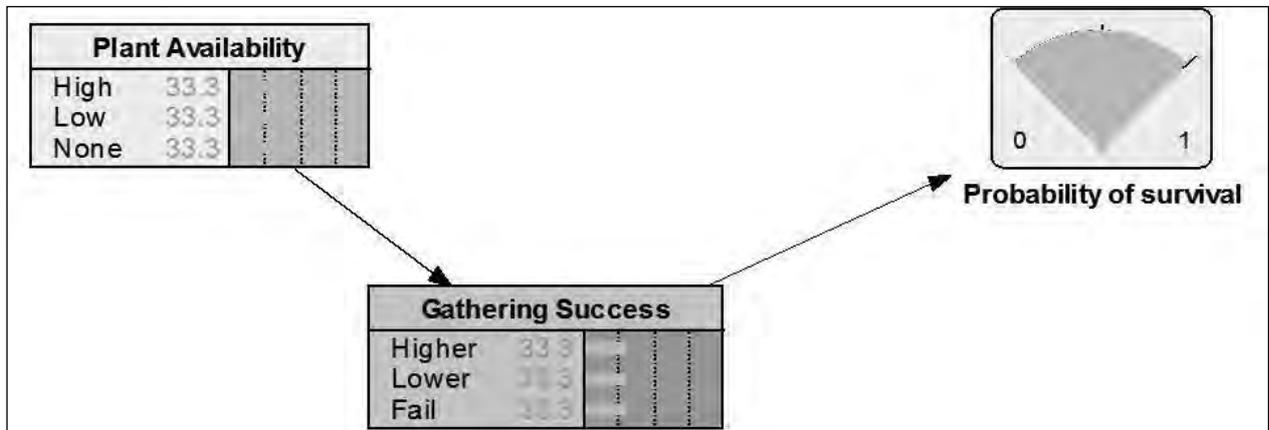


Figure 1. Survival is completely defined by the probability of plants availability in the case of gathering.

of gathering taking place at this moment and the probability of survival in those circumstances.

Hunting seems to be a much more complex activity, whose success and hence the posterior probabilities of survival are less deterministically affected by the availability of animals in the area (Fig. 2). For instance, if a social agent cooperates with another agent, the chances of hunting success are higher, even in the case of low animal availability, and so on. Therefore, the network model needs additional factors and dependencies to be able to calculate posterior probabilities for hunting success and hence of survival.

Our preliminary hypothesis suggests that labor force availability is the most relevant factor

when means of production remain poorly developed. Therefore, territorial mobility is not only important for locating resources but for finding partners with which to cooperate acquiring resources that would be impossible without the contribution of additional labor. Cooperation in hunting is of basic importance to understand the probabilities of surviving, because in the absence of high technological efficiency and efficacy, the only way to increase the probability for hunting success is by increasing the number of hunters collaborating in looking for preys and killing them. The question is how such cooperation may emerge? Why social agents can take the decision of cooperating? We are considering two main factors, the first one, is just a consequence of human density, that is, the amount of people that can be contacted in the area. The second is the

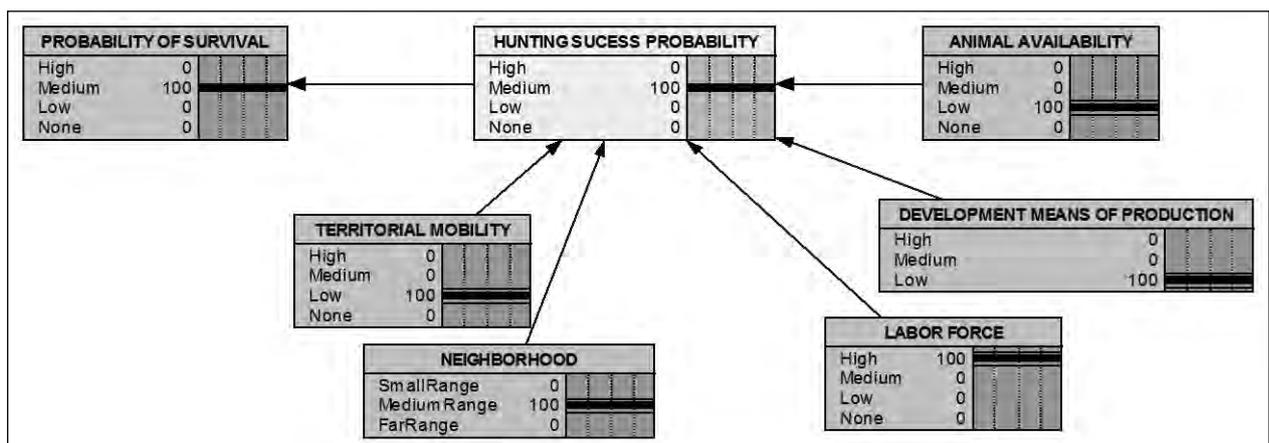


Figure 2. The success of hunting is affected by the availability of animal resources, but also by the quantity of labor force, the state of development of means of production (instruments), and the frequency of mobility and the size of the mobility area.

existence of communalities in motives, actions, behavior and mediating artifacts (like language) among interacting agents.

Our preliminary hypothesis can be expressed using a probability network (Fig. 3) in which the probability for cooperating with members of other groups appears specially affected by the probabilities of finding someone with the necessary level of cultural similarity, that is, someone who has the same motivations, not very dissimilar behaviors and recognizable mediating artifacts.

We suggest that people restrict cooperation to people from the same lineage, people speaking a similar language, or people having cooperated in the past, because these three are the most important ways of increasing cultural consensus in terms of social interactions built among different individuals. The probability for cooperating with members of other groups is specially affected by the probabilities of finding someone with the necessary level of cultural similarity, that is, someone who has the same motivations, not very dissimilar behaviors and recognizable mediating artifacts (for instance, language) (Barceló et al. 2012).

2. An Agent-based Simulation Model of Prehistoric Decision-Making

We have implemented the above model into a computer simulation framework to explore some logical consequences of our assumptions and hypotheses of what could have occurred in the prehistoric past. Ours is an agent-based simulation in which agents are surrogates of households moving through a territory looking for irregularly distributed resources (Barceló et al. 2010, Del Castillo 2012).

We have modeled an impossible world, where resources exist everywhere at random quantities and with the same natural reproduction rate all around. Carrying capacity also varies randomly. There are not any topographic barriers. Therefore, agent geographical isolation is hardly probable. In the system, agents are modeled as a number of individuals acting as a single decision unit, that is, what we usually name “family” in real live, and what anthropologists define as “domestic unit” or “household”. Each virtual “household” behave as a single entity, interacting with other similar

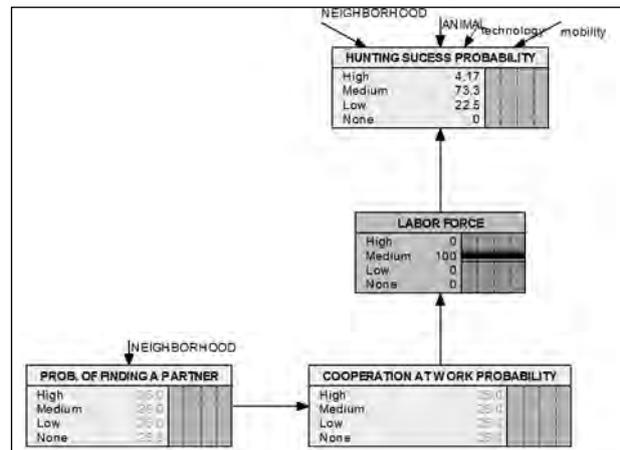


Figure 3. The probability for cooperating with members of other groups is specially affected by the probabilities of finding someone with the necessary level of cultural similarity, that is, someone who has the same motivations, not very dissimilar behaviors and recognizable mediating artifacts.

households, and/or they are affected by external environmental factors. Agents are characterized by the following attributes:

- AGE: number of lived time-steps at each time-step.
- MAX-AGE: threshold defining death. Refers to the maximum number of steps allowed for each household. It is initialized and configured as a Poisson distribution measure (up to 80). When the agent arrives to such threshold, it dies.
- SURVIVAL: Probability of survival for each agent at each time-step. It depends on the actual value of energy acquired from environment through hunting and/or gathering. It determines if the household survives or dies.
- MY-NEIGHBORHOOD: list of agents around. Although direct mobility is restricted to the patch’s area, each agent has the ability to interact and cooperate with other nearby agents; hence this attribute reflects the territory covered by the agents in a given time-step.
- ENERGY: A counter of acquired resources through hunting and gathering.
- IDENTITY: Each household has a vector of k

features (cultural dimensions). The length of the cultural vector defines the social complexity of the population. We have defined a simple and computable vector of 10 terms generated randomly at the beginning. Such a vector is modified at successive time steps because of the addition of some random noise simulating internal dynamics of cultural change, and because it should fit the identity vector of all households within a social aggregation.

- **CAPACITY:** Quantity of labor to be delivered by a household. It is normally distributed parameter counting the quantity of labor of each social unit. It is initiated randomly when households are created, and it can be modified through the simulation.
- **LEADERSHIP:** A counter of the number of households following another household.

Environmental factors and processes are constant in this simulation, and we are not representing any dynamics below the annual cycle (there are no seasons in our artificial world). There are not climatic cycles, nor climatic catastrophes: the cycle of natural regeneration is constant, given our purpose of extreme simplicity.

Seven processes are responsible for all system dynamics: agents *move*, *survive*, *reproduce* and *die*; to survive they hunt and gather and they use existing interaction flows to decide whether *cooperate* in hunting or not. Consequently, they need to *identify* other agents and build a social network, which evolve to configure a *social aggregate* if it is sufficiently stable through time. Within such aggregate, identity evolves and updates, and it is transferred to new born agents when *reproducing*. In the simulation, these processes are performed by agents simulating “households” which are assumed to be homogenous, and consequently the model does not implement anything below that level of social dynamics.

In the simulation only two kinds of economic activities have been taken into consideration: gathering, which is an individual household task, and hunting, which is only possible when the members of different households cooperate. There are increasing returns to cooperation, i.e. households get more resources working together

than individually. Nevertheless, hunting is also affected by *diminishing* marginal returns:

a) relative to the number of households exploiting the same resources at a similar place. If agents remain too much time at a single location, their immediate environment cannot regenerate resources -they do not know agriculture nor herding-, and the probabilities for survival decay according to the quantity of resources remaining at place;

b) the need to share all what has been acquired collectively, even in case there is a very small amount. The less the probability of finding animals at a particular area, the worst results are forecasted for collective hunting, given the need to share among participants. In such circumstances, individual household hunting is rationally more profitable, although more insecure.

Looking for resources at a nearby “grid cell” is the first process in the agent schedule at the beginning of a new time-step. Once in a new grid cell, households hunt and gather for surviving. When they find other households in the neighborhood and some degree of cultural consensus already exists between them, they may add their labor force to increase the probability of survival. To find other agents that may cooperate, agents should interact with others within a single local neighborhood –its geographical radius is a model parameter-, within the limits allowed by perceived cultural consensus. Identifying other households as prospective members of the same group means to be able to observe similarities in social activity between agents, and it depends on how the identity traits look to the social observers.

To simplify, we have only considered two sources of agent interaction:

- cooperation in acquiring subsistence work,
- cooperation to reproduce (production of new households).

Therefore, in the present state of the simulation, the probability of survival is only affected by:

1. the quantity of labor force involved in hunting,

given the assumed historical continuity of low technological efficiency, and

2. the spatial mobility across the environment, looking for resources and potential households partners for subsistence acquisition and/or reproduction.

Households cooperate if they have appropriately similar identities, that is to say, if some existing cultural consensus is above a critical threshold. We have approached the very idea of “group identity” in terms of perceived similarities in social activity, instead of a global measure of homogeneity. That is, instead of listing all needs, motivations, beliefs, behaviors and mediating artifacts active at each step, we have preferred to simplify calculations and creating a surrogate of the agent’s conceptual space in the form of a 10 dimensional metric space where all aspects of social life can be situated. In the simulation, the particular identity of each agent is created at birth and it is identical to that of its parents, simulating what relates each agent with its historical lineage. In principle, such an identity vector can be perceived by all other agents, who interpret the social personality and group membership of agents which they interact in the present based on it. Different situations can be imagined where all or only a part of this identity vector is accessible to agents out of the group. That is, in many cases, agents are only partially aware of what identifies the “other”.

The current value of each agent identity vector influences the probabilities of cooperating within the current time-step, but when cooperation is successful, its current value will change adaptively to fit the newly built cultural consensus. In this way, the necessary degree of cultural consensus acts as an *intolerance* parameter. We expect that when the need of cultural consensus is stronger and households prove to be very sensitive to different identities, no one is able to cooperate and chances of survival are reduced. When cultural consensus is below 70% (higher tolerance) most neighbors should theoretically cooperate, irrespective of the expressed identities, and the population expands in size.

“Identity” is socially built by individual households through a local imitation process.

Inherited at birth from parents, it evolves, changing and adapting itself to fit local features at the current time step. There are two main mechanisms for identity change:

1. Internal change, supposed to be random at the scale of a population.
2. Adaptive transformation, which fit individual household identity to collective identity when social aggregation proves to be economically advantageous.

We assume that the higher the cooperation between different households, and the higher the cultural consensus among them, the higher the probability that new households be formed within the group (descendants), so that the group maintains through time. The idea is that once the new social aggregation has emerged and survival of households has been assured, hybridization mechanisms began to act because inherited identities (ethnicity) should be modified to maintain the newly built consensus (Fig. 4).

Household actions are oriented to foraging and food gaining through mobility across a territory, conditioned by available technology and agent density, and the establishment of cooperation between agents when direct survival is not possible. However, what they have acquired as subsistence has a short temporal duration, and given the low degree of storing technology, agents should begin the process anew at the beginning of each time-step. In the model, the availability of resources is fixed as a global probability parameter. Each household has the possibility to move camp/settlement location and interact with other households in order to decide whether cooperate or not in survival or in reproduction. Each agent has the goal of optimizing the probability of survival and reproduction by gaining enough food (energy reserves) to meet a threshold of energy necessary for successful reproduction (Fig. 5).

In the simulation, households move randomly because we want to test the consequences of agents’ movement into a non-constrained space defined by the lack of topographic barriers. They are allowed to follow any direction within a restricted neighborhood. Each household defines at first the

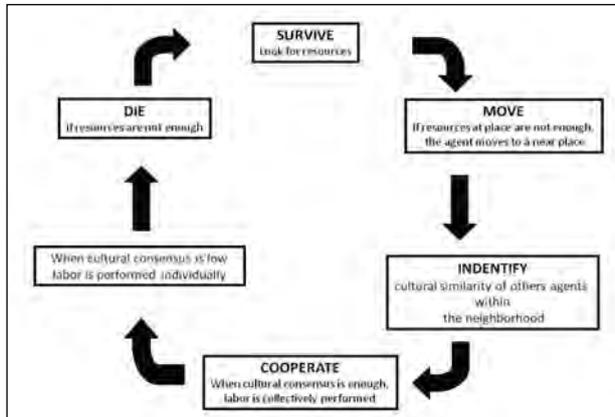


Figure 4. Overview of the cycle of cooperation causes and effects

group of neighboring households it can cooperate with (social network). The household agent has not information about all agents in the world, but only those within a reasonable geographic distance. When moving, agents interact with other agents around. A process “identify-agents” is then called to create such a network of connected households. The procedure is based on a calculation of the number of common identity traits between agents perceived within a neighborhood. Once within an aggregate, the household’s subsistence output can be enhanced adding to the agents’ capacity to work, the capacities of other agents within the group.

We have defined an output function for each household j at a time period t as a weighted average of the sum of the capacities of its local group (Fig. 6). Fig. 7 shows the equation used to calculate the amount of surplus at each time step (Fig. 7).

Identity is the key aspect for survival, given that it restricts or allows the ability to cooperate and to enhance each household capability to work. Identity is received at birth, and there is a specific process (reproduction) whose specific function is to create a copy of the parent households’ identity within the new born agent. Such identity received at birth evolves according two different sub processes:

1. A constant random rate of change (identity drift), simulating the occurrence of internal processes. From an external perspective, such decisions at the local level may appear as internal shocks perturbing the apparent linearity of a given trajectory. Therefore, although such identity changes

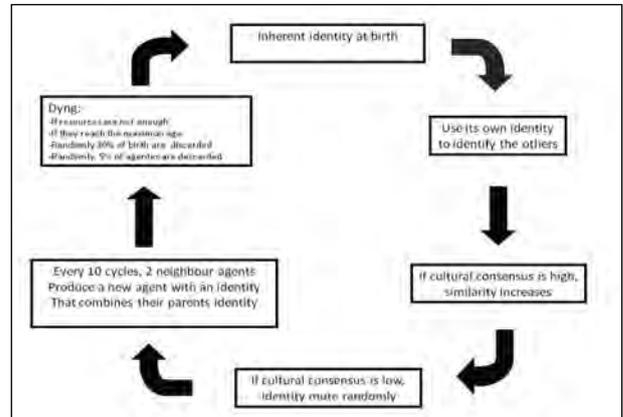


Figure 5. Overview of the identification cycle for each agent.

$$O_j(t) = \frac{c_j}{\sum_{k \in G_j(t)} c_k} \left(\sum_{k \in G_j(t)} c_k \right)^\theta = c_j \left(\sum_{k \in G_j(t)} c_k \right)^{\theta-1} \quad \text{with } \theta \geq 1$$

Figure 6. Output function for each agent j at a time period t . It is a weighted average of the sum of the capacities of its local group $G_j(t)$ raised to the power of $q > 1$, which modulates increasing returns to cooperation. Here, c_j is the agent’s capacity, and c_k introduce capacities of a group of agents working conjointly (social aggregate).

$$S_j(t+1) = S_j(t)(1-\rho) + O_j(t) - S_0$$

Figure 7. “Economic” function for each agent. Considering that a time step is roughly equivalent to one year, an agent j will obtain an amount of energy S_j which is equal to:

- the amount of available resources at that specific patch ($S_j(t)$),
- minus a depreciation parameter $\rho \in [0,1]$ simulating the low technical efficacy and efficiency of technology for storing subsistence during long periods of time,
- plus the output of its capacity to work (including contributions from other agents in the group),
- minus the minimum subsistence requirements (subsistence needs, another user-configurable parameter before starting up).

are not random at the level of each household, they can be modeled as random at the level of the population. With a fixed probability level, each household mutates one of its identity traits. Different scenarios may be imagined with different rates of internal change, in some cases, strictly random, in other exponential, once the amount of change has arrived to certain threshold.

2. By fitting the current vector to the statistical mode of all agents within the group identities. After successful cooperation based on previous cultural consensus, a new consensus should be built based on the statistical mode of all households with the cooperative network at current time, and each household will change its identity vector accordingly. With a fixed probability level an agent copies some identity trait of the mode of its group, in such a way that consensus increases and “culturally” homogenous groups tend to emerge.

Therefore, what future generations will inherit is not the old identity from ancestors, but the new communality rebuilt at each time-step. We assume that the higher the cooperation between households, and the higher the cultural consensus among them, the higher the probability two particular households reproduce within the group.

3. Discussion

As it is the case for most social simulations, we intend to use the model to experiment with a theory of social factors influencing hunter-gatherer survival in prehistory. User manipulation has been allowed for agent-level parameters to test the global implications of behavioral assumptions of our previous hypotheses, but we have also manipulated global parameters. Of course, the limit of the approach is the simple design of the economic behavior, in terms of probabilities for economic success. When manipulating the model we have generated different trajectories of virtual worlds by perturbing them randomly or introducing randomized adjustments in initial conditions. By carefully calibrating these experiments, different experiments have been designed.

We have processed hundreds of experiments in different scenarios, testing for the threshold values of parameters allowing survival. Here we will present only a subset of such experimental scenarios.

1. Controlling population

We have run the program with differently sized initial populations: 20, 40, 60 and 80 households as founding population. Results suggest that the probability of reproduction, and hence of

survival does not depend on the size of the founding population, provided there are more than 20 households. Below this threshold, the population dies, because there are not enough people to cooperate and compensate for the irregularities in resource availability.

2. Controlled variable: neighborhood

Our simulated households acquire subsistence directly from the environment. Given the poor efficiency and efficacy of their means of production they cannot restore what they have extracted from nature, and hence they are obliged to constant spatial mobility, given their extreme dependence to local carrying capacity and diminishing returns from labor. The frequency of movements, the travelled distances and the extent of the explored area constitute another relevant factor, not only because in bigger areas the chances of acquiring subsistence are higher, but also because in big areas is easier to find potential partners to cooperate, even independently of the amount of resources. When the effectively explored neighborhood increases, for instance because of a development on means of transportation –horse domestication–, cooperation increases and social fusion also increases, diminishing the number of differentiated groups and emerging cultural homogeneity. In the same sense, when the size of effectively explored neighborhood areas diminishes, mortality increases.

3. Controlling cooperation advantages

The hypothesis to be tested suggests that the higher the benefits of cooperation, the higher the interest for being involved in cooperation and exchanging networks of labor. Successive simulations suggest that when the size of the population is middle to low, but always higher than 20, and the probability to find enough resources is also medium to low, cooperation renders more benefit and it is rationally sought after by agents. The benefit of cooperation is not a free parameter, but it depends on the actual circumstances, so that it is used as a strategy by agents to increase the chances of survival. Furthermore, the higher the predictable benefits of cooperation, the lesser the probabilities of cultural differentiation and the emergence of differentiated groups. The highest probability for cultural differentiation appears when predictable benefits

are low to medium. If the advantages of cooperating in the acquisition and processing of subsistence is far enough, all agents decide rationally to exchange their labor force with anyone in the neighborhood. In the long run, cultural differences disappear given the constant rate of cultural transmission within an integrated group. When cooperation is not seen with positive eyes, agents hardly exchange labor, and in the long run cultural bias imposes its law increasing cultural differences and differentiated ethnicity emerges. But when cultural differences are too big, cooperation is infrequent and mortality increases. An additional result is that for very high rates of predictable cooperation benefits the effect saturates, emerging a minimum number of different groups as a result of the random character of constant cultural drift

4. Controlling identity

Our simulation shows that this “cognitive aspect of human behavior affects the chance of survival, because it has important influence on the conditional probability for cooperation. The simulated experiments show that decisions about whether to cooperate depend on the level of cultural consensus, consequently, the more complex and diversified the identity, the more difficult to define cultural consensus, and hence that harder the decision about cooperate or not with an agent entering in the same neighborhood. When identity is fairly simple to establish because it is based on very few cultural dimensions, cultural consensus can be easily defined and maintained, and cooperation facilitates hunting success.

5. Controlling tolerance

“Tolerance” to other difference has been traditionally considered as an additional cognitive factor to explain why agents decide to cooperate. Traditional sociological analysis has considered it as an external “innate” factor. Nevertheless, our simulations reveal that it is intrinsically related with the predictable future benefits of cooperation. When resource availability is scarce and prior probability of hunting success is low, cooperation is really sought after, and cultural consensus is built with any potential partner in the neighborhood. In such circumstances, the minimum threshold of cultural similarity is lower than in the case cooperation is

not regarded so positively. There are cases, however, in which resources are scarce and they tend to be monopolized by the agents first arrived to the area. In these circumstances, cooperation is regarded as something negative, because agents assume there is not enough for sharing, even given the increased probabilities of hunting success derived from investing more quantities of labor force. Our results show that the benefits of cooperation are higher when the more coincident is the degree of intolerance to the attained level of cultural consensus. That is to say, the highest chances of surviving are approached when cultural consensus is easy to build and to reproduce through generations. The more emphasis gives the group to their own identity, and the more complex is the cognitive space of identity, the lesser are the chances of survive in the conditions typical of hunter-gatherer bands. These results coincide with the empirical observation that strong territoriality among hunter-gatherers is not related with strong cultural differentiation: groups may exist, and even politically defined territories but all share the same material culture. Again in the same scenario, our simulations suggest the adaptive relevance of cognitive factors. Accepting or denying cultural difference is not an innate property of human beings, but a result of survival strategies. We have experimented successive changes in the threshold of intolerance, starting with minimum values (only 20% of common cultural features necessary for beginning a cultural consensus) and lowering the requisites for cultural consensus every 2500 steps. Result shows an arithmetic progression: the founding population segregates at medium-to-low levels of communality, and the chances of open cooperation diminishes accordingly for each additional 10 % increase in the threshold of cultural consensus. The consequence is an expected increase in mortality caused by starvation, and a substantial decrease in the chances of survival.

4. Conclusions

Our simulations contribute to sustain the view that prehistoric people did not die as often as imagined. Instead of the traditional image of prehistoric hunters fighting for survival in hard environments, we suggest that social exchange networks were easy to build and negotiate, allowing hunting success even in the case of low availability of resources and the poor efficacy of working

instruments. The simulation suggests that small sized groups (less than 10 households) died by starving because the impossibility to build a high enough number of social ties.

Furthermore, cooperation is not only the result of bounded rational decision making, but it is filtered by the specific social (cultural) identity of agents, a parameter that changes constantly because it is probabilistically conditioned by a number of social factors. Among hunter-gatherers, cooperation depended on the predictable benefits of working together and sharing the results of collective work. The higher the cooperation, the higher the probability of hunting success. We are not arguing that the higher the number of hunters, the higher the amount of meat. In some experiments, the total amount of energy per person can be lesser, but hunting success is more frequent in the long run, that is, it is more probable. In the usual circumstances of small bands with hardly efficient instruments for hunting and transport, the absence of cooperation made uncertain the probability of survival, given the increasing risk for hunting failure, even in the case of high animal availability in the area.

What archaeologists have called “culture” and is used to distinguish different groups of people that may have existed in prehistory one of the material consequences of this complex process of socially and politically mediated survival. Our simulation shows that “cultures” are not innate features of human groups, but the negotiated consequence of the social mechanisms human groups adopted when the circumstances in which subsistence acquisition changed. Obviously, the mechanisms we have so far explored are characteristic of hunter-gatherer groups, and the effects of cooperating at work would be different in Neolithic societies.

Acknowledgments

This research has been funded by the Spanish Ministry of Science and Innovation, through Grant No. HAR2009-12258 awarded to J.A. Barceló and Project CSD2010-00034 «Social and environmental transitions: Simulating the past to understand human behaviour (SimulPast)» (CONSOLIDER-INGENIO 2010 program by Spanish Ministry of Science and Innovation, see: [es\). We also acknowledge the contribution from our colleagues Laura Mameli, Francesc J. Miguel Quesada and Xavier Vila \(Universitat Autònoma de Barcelona\), José Cuesta \(Universidad Carlos III, Madrid\), Juan M. Galán and José I. Santos \(University of Burgos\). The authors thank the referees and editors for suggestions that have increased the readability of the paper. Authors are responsible by any mistakes or misunderstandings that may remain in the text.](http://www.simulpast.</p></div><div data-bbox=)

References

Barceló, J. A., F. Del Castillo, L. Mameli. 2012. “Simulating the Emergence of Social Networks of Restricted Cooperation in Prehistory. A Bayesian network approach.” Paper presented at the symposium The connected past: People, networks and complexity in archaeology and history. University of Southampton 24-25 March 2012. Accessed April 2012 <http://coursecast.soton.ac.uk/Panopto/Pages/Viewer/Default.aspx?id=6e63616c-39b6-406e-b5bf-c0571be6fdod>.

Barceló, J.A., J.A. Cuesta, F. Del Castillo, J.M. Galán, L. Mameli, and X. Vilà. 2013. “Simulating Prehistoric Ethnicity. The case of Patagonian hunter-gatherers.” in *Fusion of Cultures. Computer Applications and Quantitative Methods in Archaeology 2012*, edited by F.J. Melero and F. Contreras. Granada Spain: Publicaciones de la Universidad de Granada.

Darwiche, A. 2009 *Modeling and Reasoning with Bayesian Networks*. Cambridge: Cambridge University Press.

Del Castillo, F. 2012. “Modelando la heterogeneidad Etnica y la Diversidad Cultural en Arqueología de Cazadores Recolectores Patagónicos. Aproximaciones desde la Simulación Computacional y los Modelos basados en Agentes.” PhD diss., Autonomous University of Barcelona.

Grove, M. 2009. “Hunter-gatherer movement patterns: Causes and constraints.” *Journal of Anthropological Archaeology* 28: 222–233.

Koller, D., and N. Friedman. 2009. *Probabilistic Graphical Models: Principles and Techniques*. Cambridge: MIT Press.

Why Hunter and Gatherers did not Die More Often? Simulating Prehistoric Decision Making
Florencia Del Castillo and Juan A. Barceló

- Neapolitan, R. 2003. *Learning Bayesian Networks*. Englewood Cliffs, NJ: Prentice-Hall.
- Oaksford, M., and N. Chater. 2007. *Bayesian Rationality: The Probabilistic Approach to Human Reasoning*. Oxford: Oxford University Press.
- Singpurwalla, N.D. 2006. *Reliability and Risk: A Bayesian Perspective*. London: Wiley.
- Smith, E.A. 1983. "Anthropological applications of optimal foraging theory: a critical review." *Current Anthropology* 24:625–651.
- Winterhalder, B., and E.A. Smith (Eds.). 1981. *Hunter–Gatherer Foraging Strategies: Ethnographic and Archaeological Analyses*. Chicago: University of Chicago Press.

Reconstruction of Ruined Archaeological Structures Using Structural Analysis Methods

James Miles

University of Southampton, United Kingdom

Aykut Erkal

Istanbul Kemerburgaz Universitesi, Turkey

Dina D'Ayala

University College London, United Kingdom

Simon Keay and Graeme Earl

University of Southampton, United Kingdom

Abstract:

Examination of material remains of previous human societies help study past human behaviour in archaeology. Some of the most important remains are the ruins of historic buildings. The use of computerized recording equipment and three-dimensional drawing are essential tools for researchers to recreate a ruined structure for reconstruction and further analysis. However, the reliability of the precision of the reconstructed model can be controversial due to the absence of sufficient historical information and fabric loss. In this study, structural analysis has proved to be an alternative or confirming tool for reconstruction in addition to site survey and historical investigation. On the basis of the stability of the ruined structures possible hypothetical forms can structurally be analysed under the associated loads and the resulting behaviour can be assessed to determine its actual form. Therefore, based on a multidisciplinary methodology, a novel approach to reconstruction of a historic structure has been presented in this paper. Following laser scanning to measure the complex geometry of the fallen structure and geophysical techniques to estimate the properties of inaccessible parts, a finite element model has been generated and some original forms of the construction has been evaluated based on the static analysis results.

Keywords:

Cultural Heritage, Reconstruction, Original Form, Structural Analysis, Archaeological Techniques, Pozzolana Concrete

1. Introduction

The historic environment provides a unique record of past human activity. The vision, abilities and investment of successive societies have formed the distinctive qualities of this environment. Therefore, it is a valuable resource, providing a sense of continuity and identity (Heritage 2008). Amongst the cultural heritage components, historic buildings have a significant place. Thus, documentation of heritage buildings is critical for their protection and restoration (Yastikli 2007). In archaeology, various methods are used. Computerized recording equipment and three-dimensional plotting are

Corresponding author: jm1706@soton.ac.uk

widely used tools (Sinning-Meister, Gruen, and Dan 1996; Park et al. 2007). However, documentation can be challenging work depending on the extent of the remains of the past structures. Absence of sufficient historical information and loss of building components and fabric may substantially decrease the reliability of the reconstructed models.

Structural analysis can be an alternative or confirming tool to assess the reliability of reconstructions. On the basis of available data relating to the ruined structure, possible hypothetical forms can be analysed structurally under their associated loads and the resulting behaviour can be assessed to determine its likely form. Thus, based on

a multidisciplinary methodology, a novel approach to reconstructing a historic structure is presented in this paper. The use of geophysical techniques for site survey, computer modelling tools for generation of plausible hypothetical forms and finite element modelling for structural assessment account for the additional benchmarking of the research performance.

2. Previous Studies

Archaeology can be regarded as a quintessentially visual discipline, because visual perception leads to the awareness of the fundamental properties of objects and to the discovery of how objects were produced and used in the past (Barceló 2010), especially from a computational archaeology point of view. Unfortunately, visual representations provide limited interpretation. Actually, many three-dimensional models produced through the interpretation of raw site data, comparisons with standing structures and common architectural understandings have a common limitation of insufficient validation. For example, the detailed study of the ancient Egyptian temple of Kalabsha and its illumination (Sundstedt, Chalmers, and Martinez 2004), the study on Roman Imperial architecture (Packer 2006) and the Rome Reborn project (Frischer 2008) all provide visually stimulating interpretations of the past that incorporate extensive literature research. However, the models can be critiqued in terms of structural accuracy. Computational archaeology within the scope of historic building evaluation relies heavily on creating models that can be used to visualise certain aspects and to some extent these reconstructions may not have a fundamental basis from which they are reconstructed. The data is used to evaluate and test various academic theories, in turn helping academics and researchers to interpret archaeological sites and contribute to the education of the public (Eppich and Chappi 2006). Therefore, to ensure the correctness of the models to a maximum degree, these interpretations need to be evaluated considering the other decisive aspects. Examination of the structural stability of a building can be counted as one of the reliable confirmation approaches. Investigation of whether the reconstructed building can withstand the gravitational loads forms the foundation of additional validation of the model.

Unlike hypothetical three-dimensional models of ruined buildings, models of standing buildings are produced as an exact imitation of the real. Laser scanning through the work of (Balzani et al. 2004), (Shih et al. 2007) and (Sturzenegger and Stead 2009) and photogrammetry through the work of (Lerma, García, and Pérez 2005), (Sauerbier et al. 2004) and (Haneberg 2008) provide accurate and reliable models that are exact and can be used as a tool to investigate and analyse these buildings. Archaeology is based on interpretative data, through the lens of the archaeologist. These interpretations are not factual with an utter certainty and are constantly open to new interpretations as the understanding of the past changes. In the case of part or complete loss of superstructure, created hypothetical buildings can therefore benefit from analyses that test how structurally stable the buildings are.

One of the analysis techniques available is Finite Element Method (FEM). FEM divides the body to be studied into finite elements, connected to each other at their nodes, and obtains an approximate solution by solving the equilibrium equations of the system and obtaining the nodal values. In stress analysis for instance, the field variables are the displacements (O. C. Zienkiewicz et al. 2005). Several researchers have used FEM together with some archaeological methods. For instance, Lubowiecka et al. investigated a masonry historic bridge (2009). Complex geometry was obtained through laser scanning and ground penetrating radar was used to understand material homogeneity and heterogeneity with the structural model analysed by FEM. Similarly, close range photogrammetry for the generation of masonry arches was used for more accurate three-dimensional geometric models within Riveiro et al's work (2011) which incorporated FEM to find failure load and distribution of internal stresses.

3. Aims

The aim of this paper is to present a new methodology for more reliable reconstruction of ruined historic buildings. The systematic novel methodology includes:

1. Site Survey: Acquisition of raw data through archaeological techniques using:

- 1.1. Excavation,
- 1.2. Laser scanning recording,
- 1.3. Topographic survey,
- 1.4. Geophysical survey,
2. Documentary investigation,
3. Computational modelling for the generation of plausible hypothetical forms,
4. Finite element modelling for structural analysis to test the hypothetical forms.

Archaeological techniques initially provide data upon which three-dimensional hypothetical models of a given building can be generated. Documentary investigation provides a means to bridge the gaps. Various hypothetical interpretations of the data can then be produced by examining how these buildings might have been constructed and how they may have looked in the past. In turn, analysis of the structural performance of these three-dimensional models through FEM allows refinement of possibilities by suggesting feasible alternatives.

In order to assess the effectiveness of this methodology the procedure has been applied to an important Roman building known as building V, recently discovered at the site of Portus, Italy (Keay, Paroli, and Strutt 2005). From a holistic perspective, all of the available data for the building are used to create visualisations that are both structurally stable and seem visually to fit with parallels and interpretations.

4. Case Study

4.1. History of Portus

Portus is located on the west coastline of Italy near the mouth of the Tiber near the harbour of Ostia (Fig. 1a). The site functioned as a principal port of Rome from the first century AD and was established under the rule of Emperor Claudius in around 42AD (Keay, Paroli, and Strutt 2005). Its creation involved the construction of two concrete moles bending towards each other as they stretched into the sea, creating a large artificial anchorage as seen in Figure 1b. This phase also saw the creation

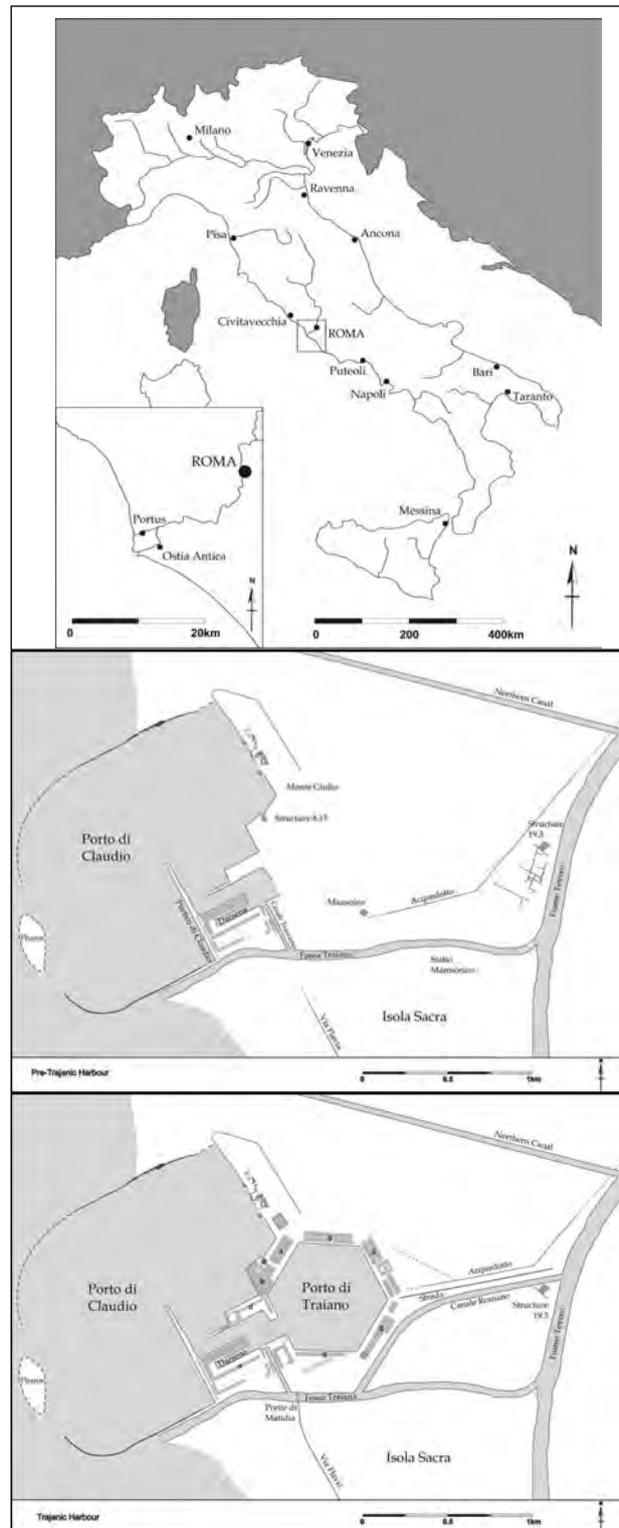


Figure 1. a) Location of Portus within Italy, b) Plan of Pre-Trajanic Portus, c) Plan of Trajanic and Severan Portus (Keay et al. 2005).

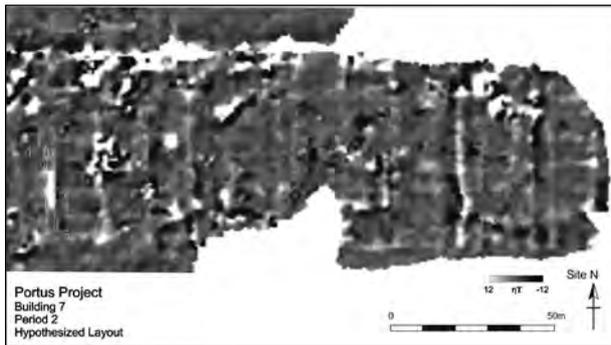


Figure 2. Magnetometry survey results.

of a lighthouse and a small inner basin, the Darsena and other buildings to facilitate the movement and storage of goods.

The site was later expanded and renovated by the Emperor Trajan in the early 2nd Century AD as seen in Figure 1c. A second inner hexagonal basin was created just to the south of the Claudian basin and incorporating the Darsena which linked to the still functioning original harbour. The Trajanic reconstruction also saw the construction of a series of buildings surrounding this harbour, (A) Building V, (B) Palazzo Imperiale, (C) Amphitheatre, (D) Severan Warehouse, (E) Portico di Claudio and Trajanic warehouse and (G) Warehouses (Fig. 1c). Further expansion of the site occurred in the reign of the Severans, with the site later being sacked by Alaric the Visigoth in 409AD.

The site was known about from the 16th Century onwards and appears in numerous engravings. It was not until the 19th century that a comprehensive account of excavations appeared through the Torlonia excavations of 1864-7 by Lanciani in 1868 with further research seen through Calza (1925) and Lugli (1935) work prior to the second world war. From the late 1990s the Soprintendenza per i Beni Archeologici di Ostia have been undertaking conservation of the site with the Universities of Southampton, Cambridge and Durham and the British School at Rome in a project of extensive survey techniques. This project ran from 1997 to 2005 using topographic and geophysical surveying. The project lead onto another by the Universities of Southampton and Cambridge and the British School at Rome, funded in two phases by the UK Arts and Humanities Research Council

(AHRC)². This developed from the previous surveys as well as starting a series of excavations and other analyses.

The structure known as building V is one of seven large scale constructions excavated by the Portus Project. The structure lies between the two harbours, to the north of the palazzo imperial. Building V is a rectangular building over 180m in length and is aligned with one edge of the hexagonal basin with its northern façade incorporated into the late Roman wall to the north of the hexagonal harbour. Brick stamps and building techniques place the earliest phase of construction to the early second century AD, in the reign of Trajan or Hadrian. It seems probable, therefore, that this building is part of the renovation of the port that begun under Trajan that saw the construction of the hexagonal harbour and the palazzo imperiale.

4.2. Archaeological investigation

4.2.1. Extant remains

A great deal of what is now understood about the structure of building V can be seen in remains extant above ground level on the site of Portus. These remains contain building elements which have been incorporated into the Late Roman wall known as the Mura Costantininae which dates to the 5th century AD. The section of the wall which separates the built-up area of Portus from the Claudian harbour incorporates the façade of building V within its structure as a series of brick-faced concrete pylons at differing intervals along the length of the wall.

4.2.2. Geophysics

In 2000 an extensive magnetometry survey of the area of building V was undertaken. The results (Keay, Paroli, and Strutt 2005) reveal numerous structures on the alignment indicated by the façade preserved in the Late Roman wall. There are several anomalies within the data that match the positions of the pylons preserved in the wall, indicating that the walls or rows of pylons run longitudinally which

² www.portusproject.org; www.heritageportal.eu The Portus Project is directed by Professor Simon Keay and is a collaborative research initiative involving the University of Southampton, the British School at Rome, the University of Cambridge and the Soprintendenza Speciale per i Beni Archeologici di Roma.

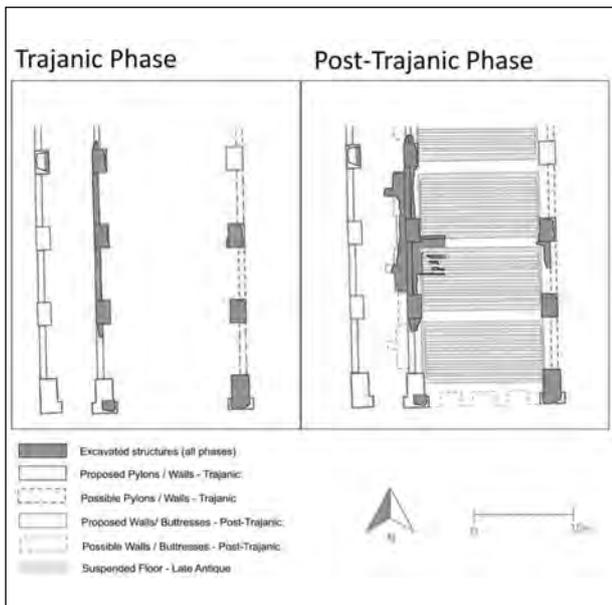


Figure 3. South western extent of the building V showing excavated remains and proposed layout in both the Trajanic and Post-Trajanic phases (Harrison 2011).

suggests that the transverse walls divide each bay into smaller rooms as seen in Figure 2.

4.2.3. Excavation

Excavations to the north of the Late Roman wall provided more detailed information about the building in 2007. A small excavation revealed the brick levelling course of the foundation and the concrete below. In 2011 a large area within the building was excavated, providing key information relating to both its structure and phasing (Fig. 3). Three rows of pylons running lengthways were present in the south west corner and followed the pattern seen in the Late Roman wall. The excavation focussed on one of the smaller spans of c. 4.5m, and a larger span of 12m, with brick stamps indicating that the earliest phase of construction dates to the early 2nd Century AD and fits in with Trajanic remodelling of the site. Further to these pylons, narrow lengthways walls of 58cm width survive and are faced in opus mixtum which is indicative of an early second century date, indicating that they may have been built immediately following the construction of the pylons. The central row had a further set of lengthways walls, built up with the wall forming transverses which divide the larger bay.

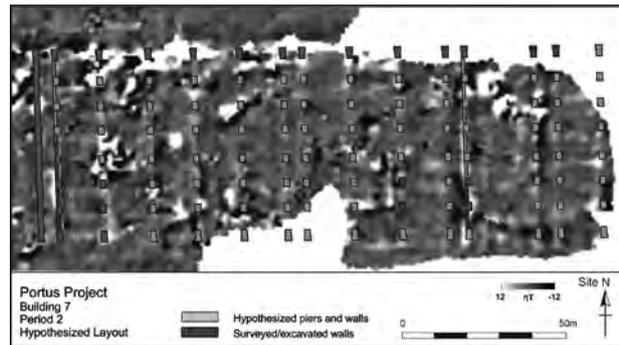


Figure 4. Layout of building V based on the combination of excavation, magnetometry survey and extant remains.

4.3. Documentary analysis

Through the comparison of the different techniques used within the previous section we can clearly produce the basic layout of building V (Fig. 4).

It has been shown that the building consists of at least two different sorts of internal elements which appear to be internal spaces that traverse the building from northern to southern façade. The first is generally 4.3m wide whilst the other is generally 12.3m wide. The layout of the individual pylons have been repeated and projected as far as the survey extends. Whilst the remains, surviving above ground level, have been surveyed, it cannot be determined whether all the remains relate to building V in its initial Trajanic phase or through later conversion of the building. Likewise, this can also be applied to the geophysical data. One possible interpretation of building V is that it was used for the construction of ships. Similar forms to building V have been identified (Harrison 2011; this volume).

5. Investigation of the Building Form

Having completed the necessary data collection through archaeological investigation and documentary evidence a series of distinct models of building V were defined:

- The first hypothetical model was based on a barrel vaulted roof that has a higher front façade leaving the rooftop bare to reduce the weight of the roof;
- The second hypothetical model deals specifically

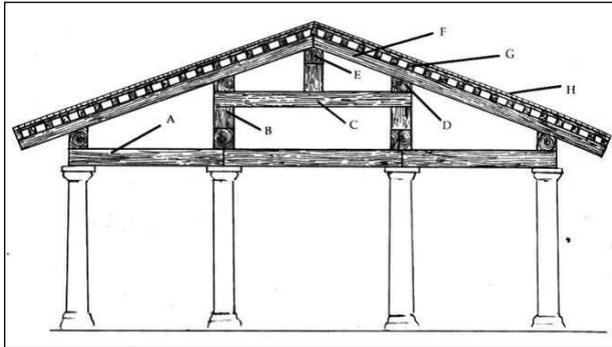


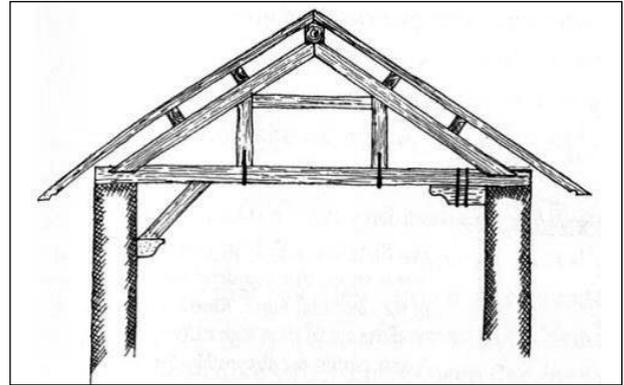
Figure 5. a) Top. Prop and lintel timber roof b) Left. Tie-beam truss roof with queen posts (Ulrich 2007).

with an alternative barrel vault which has a consistent height throughout and contains a solid flat rooftop;

- The final model focusses on a timber framed roof which has a pitched front.

The models' dimensions have been chosen based on fieldwork and desk-based studies. Specifically, the length and width of the building are 200m and 60m respectively, which were identified through geophysics, excavation and extant remains. The heights of the pylons were indicated by use of a laser scanner to measure accurately the size of a fallen pylon next to the remains of building V. This was found to be 7m in height. From these pylon heights, necessary arches were created using the width of the bays found within the façade and the internal rooms which create the necessary height of the building. There is little evidence of what type of roofing may have existed and as such the most likely vaulting and timber frame construction techniques have both been considered.

The remains of building V principally consist of the material *opus caementicium*, often referred to as Roman poliozonic concrete. Therefore, in the first two scenarios the building material is assumed to be made of *opus caementicium*. This material is comprised of an aggregate of stone or broken bricks mixed with lime mortar with volcanic sand to create a durable and strong material that can be moulded into a variety of forms through the use of wooden framework and centring. It is frequently used to form the structural core of walls and pylons, as is the case in building V. For the structural analyses, the data needed only relies on the elements that



are of structural importance, so the investigation only focuses on the *opus caementicium* as a core structural material of the structural load-bearing elements.

Regarding the third scenario, there are two types of timber roof structures known to be used in Roman buildings (Fig. 5). The first is a prop and lintel roof where the principal rafters rest on the architraves of the outer structure, and gain additional support from resting on 'props' that lie on architraves of internal walls. The second is a tie-beam truss which allows for a wide space to be spanned by a single pitched roof structure without the need for internal walls or columns to support it. The tie-beam truss uses smaller timbers and can support a greater weight that is less subject to sagging which makes it the logical choice for the roofing of building V. An example of this can be seen in Figure 6 which shows the basilica of St Paul's which was built in the later fourth century with the roof surviving in its original form until a fire in 1823.



Figure 6. Interior Of The Basilica Of St Paul's Outside The Walls, Rome (Ulrich 2007).

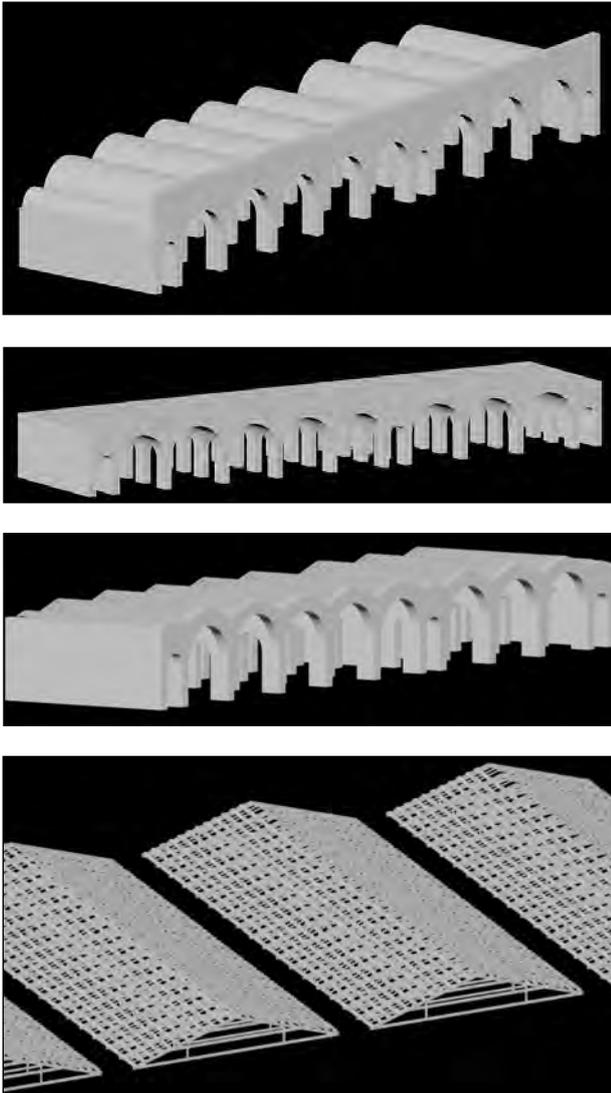


Figure 7. CAD drawings of the three hypothetical structures.

6. Creation of Models

In this study structural analysis is used to analyse the load-bearing structural system of the building in order to predict the responses of the structure under the relevant loading through linear analysis. The purpose is to test the adequacy of the structural core from the view point of stability of the structure. There are various methods to analyse a construction structurally. For instance, in order to assess the structural safety of a medieval masonry bridge, de Arteaga et al. (2012) conducted a structural analysis study where four different analytical methods were employed to see the scope of each

with a view towards their utility in collecting data in the field. In the analyses, the geometry of this bridge was obtained using terrestrial laser scanning and photogrammetry technology, which are commonly used in archaeology. One of the main conclusions was the need for multidisciplinary group work to make better structural assessment of masonry arch bridges, emphasising the incorporation of interdisciplinary approaches such as geomatic techniques. This clearly confirms the centrality of archaeological techniques in data collection for further analysis of structures.

In this study, Finite Element Method (FEM) was used for the analyses. After the acquisition of data using archaeological techniques through a site-study and documentary analyses through a desk-study, plausible interpretations of geometrical building forms, creating as CAD drawings, formed the foundation of structural modelling as part of computational archaeology. In other words, building geometry was created. Dividing the structural body into a number of finite elements was the next step before the solution of equilibrium equations incorporating structural material properties (Autodesk 2012).

6.1. Creation of hypothetical forms

The three-dimensional models were produced within AutoCAD using three-dimensional solid parts. Then the model geometry was transferred into the Finite Element Software, ALGOR (Autodesk 2012) for meshing the model into finite elements and for the assignment of the considered material properties for each part. The models can be seen in Figure 7 with each having the same basic structure. They only focus on the structural elements of the building and ignore any decorative features as the data is not needed to test the structural capabilities of the interpretations. Fixed boundary conditions have been applied at the bottom of the pylons. Importantly, in order to reduce the number of equations during the solution procedure, only half of the building has been analysed imposing boundary conditions considering symmetry for the symmetrical self-weight loading of the building.

To ensure the highest quality mesh, bricks and tetrahedral finite elements have been used. The solid mesh consisted of as many 8-node three-

Materials used in Models	Modulus of Elasticity	Mass Density	Poisson Ratio	Compressive Strength	Tensile Strength
Opus caementicium	3 GPa	1540 kg/m ³	0.2	4 MPa	0.1 MPa
Silver fir	11 Gpa	500 kg/m ³	0.22	40 MPa	80 MPa

Table 1. Material Properties of Opus Caementicium and Silver Fir.

dimensional finite elements (brick elements) as possible with respect to the mesh size. These 8-node elements have 3 degrees of freedom at each node. As commonly used for this type of immense masonry building, static stress analysis with linear material properties has been defined (Perucchio and Brune 2008). The static self-weight loads of the building have been applied to the model and all material strains are expected to be in the linear elastic range of the material. Resulting maximum stresses and strains have been compared with the strength of material to test the structural reliability of the models.

6.2. Material properties

There are a number of different forms of concrete vaulting found within Roman architecture with examples seen at the markets at Tivoli, Porticus Aemilia and the Basilica of Maxentius. Significantly Lancaster (2005) provides a complete survey of vaulted constructions within Imperial Rome through which it can be seen that opus caementicium is a suitable material that can be used as the material property in the structural analysis. In addition, the study by Perucchio and Brune (2008) also confirms that opus caementicium is the most realistic material that can be used within the case study of building V at Portus. Linear elastic isotropic material properties of opus caementicium needed for the static linear stress analyses are given in Table 1 and with the material properties taken directly from Perucchio's analysis of the Trajan's Market (ibid) as no test have been completed on the Portus remains.

Roman builders were aware of the superior properties of timbers from further afield. There are four types of tree that could have been used for the largest spans, namely fir, mountain pines, larch and cedar. Larch from the Alps was renowned for being slow to catch fire, and cedar from the Levant was thought to be extremely strong and long-lasting (Ulrich 2007). The firs from the Appenines, principally silver fir, grow to up 60m tall with a

trunk diameter of 1.5m, providing a sufficient length timber for any truss. In addition to these advantages, silver fir, as a common and readily available timber (Slavid 2006), was used in the structural analyses with the material properties (MATBASE 2012) shown in Table 1.

7. Results

The analysis took place using a fine mesh with the first two models using a mesh size of 0.2m and the third using 0.2m within the building and 0.1m within the timber roof. This mesh size was used in order to increase the accuracy of the FEM results; the displacements that occurred have been exaggerated to identify the subtle movements that can be seen within the results.

In continuum mechanics, stress is a measure of the average force per unit area of a surface within the body on which internal forces act. At every point in a stressed body there are at least three planes, where the corresponding stress vector is perpendicular to the plane, and where there are no normal shear stresses. The three stresses normal to these principal planes are called principal stresses. The results focus on two of the principle stresses,

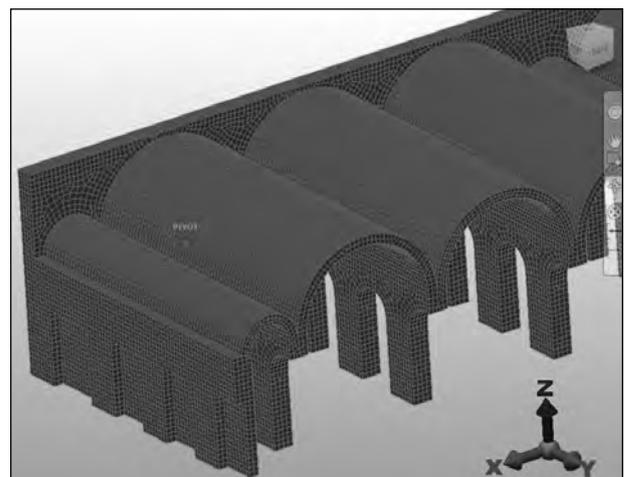


Figure 8. Meshing of the first model.

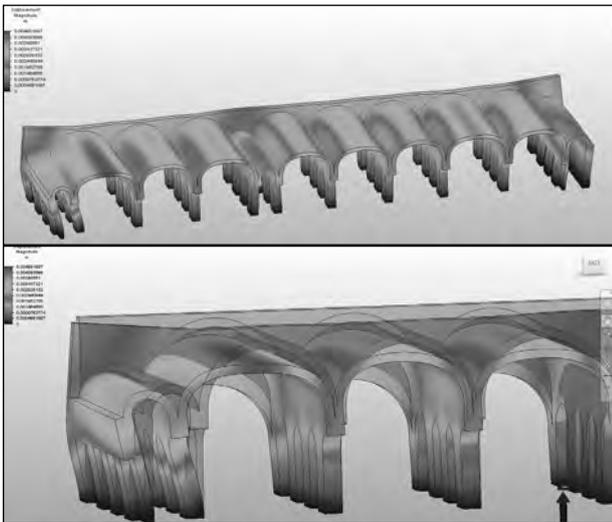


Figure 9. Displacement of the first hypothetical model.

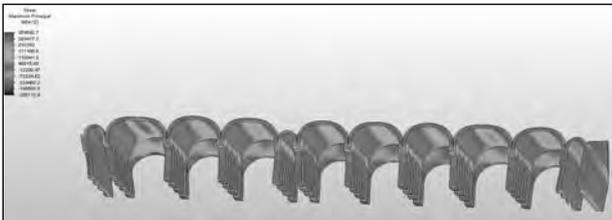


Figure 10. Maximum principal stresses of the first model.



Figure 11. Deflected shape of the second hypothetical building.

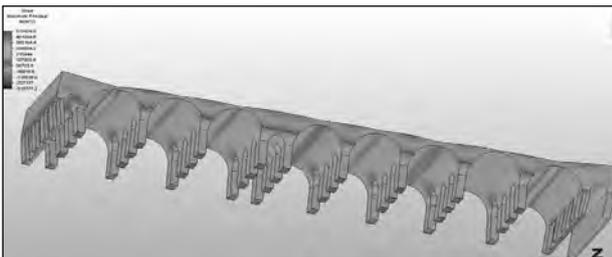


Figure 12. Maximum principal stresses of the second model.

namely minimum which tests compression and maximum which tests tension.

The weakness of Roman pozzolana concrete is its very low tensile strength, which is, to a certain degree, a control point to decide on the probability of

the selected building form. Therefore, the maximum tensile stresses are significant to see the critical regions on the model and their magnitude.

7.1 First hypothetical structure

Figure 8 illustrates the mesh of the model and Figure 9a shows the displacement results from the first hypothetical building under its own gravitational force. In general, the maximum vertical displacements occur at midpoints of the vaults. The maximum displacements are at the midpoint of the first left vault due to its span being slightly larger than the others and are supported with a less stiff side wall in comparison to the other side of the structure. As expected, the displacements get smaller and smaller at lower levels close to the ground since the structure is fixed at the bottom of the pylons. A close up image (Fig. 9b) of the left side vault highlights the deflected and non-deflected geometry. Due to the load transfer along the vault, the narrow side vaults are pushed outwards by the adjacent vaulted roof due to the stiffness of the left side of the vault being less than the other side.

Figure 10 shows the maximum principle stress contours on the model. Due to the bending of the vaults under gravity loads, tension stresses develop at the bottom of the vaults. The maximum value here is seen to be 354542.7 N/m^2 which is 3.5 times larger than the tensile strength of the material. Due to the maximum value being higher than the tensile strength of the material, this specific interpretation presents very low probability as this form is not structurally strong enough to carry itself.

Although the structure is implausible due to the tensile strength of the Roman pozzolana concrete, it was also important to test if the structure had sufficient strength to accommodate compression forces. As a result, the maximum compressive stress of 40907.7 N/m^2 developed in the entire structure is less than the material compressive strength of 4000000 N/m^2 . Analysis results suggest that the tensile stresses control the failure behaviour in this model.

7.2. Second hypothetical structure

The model was again tested through self-loading and was analysed using linear stress analysis

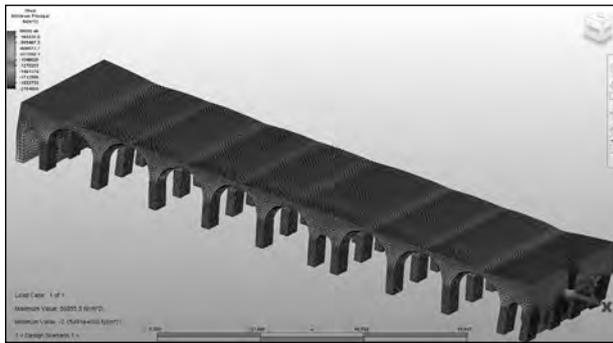


Figure 13. Minimum principal stresses of the second model.



Figure 14. Deflected shape of the third model.

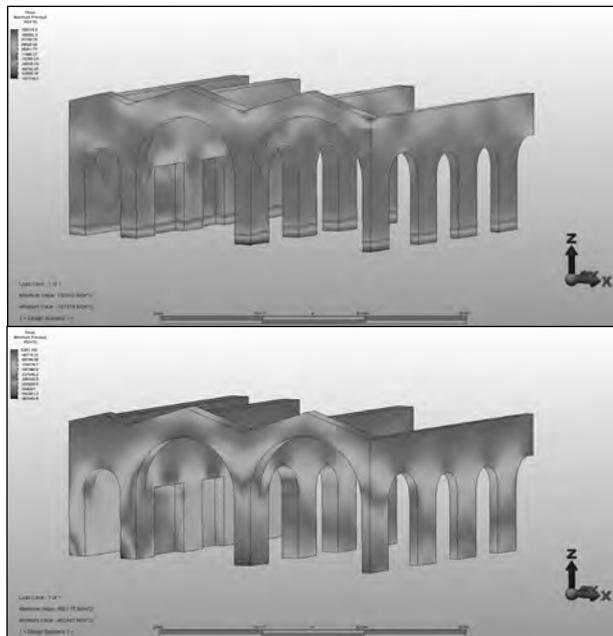


Figure 15. Principal stresses on the third model.

with linear material properties. Figure 11 highlights the deflected shape of the structure as well as the mesh of the model. The model has quite fine meshing in order to replicate geometry in a more realistic

state. Expectedly, analysis results demonstrate that relatively larger displacements take place at the mid-points of the arches of the barrel vaults. The maximum displacement occurs over the first left large arch as it is slightly bigger than the others; also the arch thickness at the top is less than that on the other side of the building. Therefore left side vertical structural elements bend towards right with the effect of the heavier roof unlike the previous model. These displacements have again been magnified to highlight the movement that occurs as the maximum displacements are as large as 3.6 mm.

The maximum principle stress shown in Figure 12 is 200000 N/m² which is twice the tensile strength of the material (100,000 N/m²). The result although not as high as the first model again proves that the selected building is structurally unstable as the tensile strength is double that of the material strength. The minimum principle stress being 2154910 N/m² is lower than the compression strength of the material (Fig. 13) however tension stress results again suggest that this hypothesised structure would not be viable.

7.3. Third hypothetical structure

Under self-loading and using linear stress analysis, this time, the model is based on two different linear materials: fir for the timber roof and Roman pozzolana concrete for the building. In order to test structural reliability of this hypothetical structure there was a need to examine both materials separately. Figure 14 depicts the model geometry and the deflected shape under self-weight.

Figure 15 focuses on the maximum and minimum principal stresses of the Roman pozzolana concrete. The result for the concrete is 85358 N/m² which is less than the tensile strength of the material. The result for the fir is 475976 N/m² and is again less than the tensile strength of the material. Regarding minimum principal stresses of both materials, the results are considerably lower than the compression rate allowed for both.

The results provide a model that is structurally stable and suggests that this could have been a possible form of the building. Although it cannot be said with absolute certainty due to the various possibilities of structural form, material

heterogeneity, inherent approximations of the analysis method and model idealization, it can be concluded with a high probability that the hypothesized third structure is the most probable structure amongst three of the buildings, based on the mathematical examination.

Importantly, geometric variations within the modelling of the first two models could be tested to see the structural building response. This could be done, for instance, through altering the roof thickness and material properties of these two structures. These alterations would provide an assessment of whether the updated shape could be structurally viable or not.

Linear analysis rather than non-linear analysis was used as a basis to test the hypothetical forms of the building, as it would have to withstand its own static stress before extreme events, such as earthquakes, would affect it. Although non-linear analyses can be performed for a more in-depth investigation, given the involved uncertainties linear analysis has been considered sufficient alongside the other assumptions.

8. Conclusions

For more reliable reconstruction of ruined historic buildings a systematic and novel methodology has been suggested. Data acquisition through archaeological field survey is crucial. Excavation, laser scanning, topographic survey and geophysical survey are the effective techniques to obtain an insight in the structural remains. In addition, documentary investigation completes the various gaps during the overall analysis procedure. The data can be used to create a model of a past structure and provide information about its physical properties. Once sufficient data is produced, computational modelling for the generation of plausible hypothetical forms is the next step towards finite element modelling for structural analysis to test them.

Structural analysis provides suggestions as to the basic form of past buildings. In terms of the case study the results suggest that two hypothetical structures may be discredited with a third providing a viable structure. However, there are many other possible structurally stable

forms. Since the loads and their application can vary between hypothesised buildings depending on many environmental factors, different analysis types will be incorporated in future work to evaluate fluid flow, thermal dynamics, vibration and critical buckling, to complement the existing work focus on gravitational load. The resulting models can then be used to create photorealistic models for visual representation.

Chalmers and Debattista (2005) suggested that structural analysis could be used as a way of producing viable models for interpretation within archaeology. Their research on prehistoric Maltese temples provided an innovative approach, but was limited by the lack of a complete material property database and the software available. The method produced within this paper builds on their work, and confirms that there is a need for archaeological modelling to focus equally on the collection of volumetric and surface survey data, and their deployment within a structural analysis framework, alongside historical knowledge and conventional site survey.

Acknowledgments

Thanks must go to the AHRC for funding the Portus project; Matthew Harrison for the prior work that he completed; Grant Cox who along with Matthew produced a number of hypothetical models of building V; Penny Copeland for her help outlining the shape of the building as well as providing the necessary survey data and Christina Triantafyllou for her help in outlining the form of the building.

References

- Autodesk. *Autodesk Simulation Mechanical and Multiphysics* 2012. Accessed 15 June 2012. <http://usa.autodesk.com/adsk/servlet/pc/index?siteID=123112&id=13773836>.
- Balzani, M., N. Santopuoli, A. Grieco, and N. Zaltron. 2004. "Laser Scanner 3D Survey in Archaeological Field: the Forum Of Pompeii." In *International Conference on Remote Sensing Archaeology*, edited by Joint Laboratory of Remote Sensing Archaeology, 169-175. Beijing: Joint Laboratory of Remote Sensing Archaeology.
- Barceló, J. 2010. "Visual Analysis in Archaeology.

- An Artificial Intelligence Approach Morphometrics for Nonmorphometricians.” In *Morphometrics for Nonmorphometricians*, edited by Ashraf M. T. Elewa, 93-156. Berlin and Heidelberg: Springer.
- Calza, G. 1925. “Ricognizioni topografiche nel porto di Traiano.” *Notizie degli Scavi di Antichità* 6 (1):54-80.
- Chalmers, A, and K Debattista. 2005. “Investigating the Structural Validity of Virtual Reconstructions of Prehistoric Maltese Temples.” Paper presented at the *6th International Symposium on Virtual Reality, Archaeology and Cultural Heritage, Pisa, Italy*.
- de Arteaga, I., and P. Morer. 2012. “The effect of geometry on the structural capacity of masonry arch bridges.” *Construction and Building Materials* 34 (0):97-106.
- English Heritage. 2008. *Conservation principles policies and guidance for the sustainable management of the historic environment*. London: English Heritage.
- Eppich, R, and A Chappi. 2006. “Recording and documenting cultural heritage- 3D modeling for conservation in developing regions.” In *Recording, Modeling and Visualization of Cultural Heritage*, edited by M. Baltsavias, A. Gruen, L. Van Gool and M. Pateraki, 11-20. London: Taylor & Francis Group.
- Frischer, B. 2008. *The Rome Reborn Project. How Technology is helping us to study history, OpEd*. University of Virginia.
- Haneberg, W. 2008. “Using close range terrestrial digital photogrammetry for 3-D rock slope modeling and discontinuity mapping in the United States.” *Bulletin of Engineering Geology and the Environment* 67 (4):457-469.
- Harrison, M. 2011. “Shape Grammar Modelling for Archaeology: Visualisation of the possible Ship-Shed at Portus.” Msc Thesis, University of Southampton.
- Keay, S., M. Paroli, and K. Strutt. 2005. *Portus: An Archaeological Survey of the Port of Imperial Rome*. London: British School in Rome.
- Lancaster, L. 2005. *Concrete Vaulted Construction in Imperial Rome: Innovations in Context*. New York: Cambridge University Press.
- Lerma, J. L., A. García, and C. Pérez. 2005. 3D Reconstruction and Modelling of Architectural Domes. In *CIPA 2005 XX International Symposium*. Torino, Italy, edited by Sergio Dequal, 902-905. Torino: CIPA 2005 Organising Committee.
- Lubowiecka, I., J. Armesto, P. Arias, and H. Lorenzo. 2009. “Historic bridge modelling using laser scanning, ground penetrating radar and finite element methods in the context of structural dynamics.” *Engineering Structures* 31 (11): 2667-2676.
- Lugli, G. 1935. *Il porto di Roma imperiale e l'Agro Portunense*. Bergamo: Officine dell'Istituto Italiano d'Arti Grafiche.
- Matbase. 2012. “Silver Fir.” Accessed 15 June 2012. <http://www.matbase.com/material/wood/class4-5-10-years/silver-fir/properties>.
- Packer, J E. 2006. “Digitizing Roman Imperial architecture in the early 21st century: purposes, data, failures, and prospects.” In *Imagining Ancient Rome: Documentation - Visualization - Imagination*, edited by L. Haselberger and J. Humphrey, 309-320. Portsmouth: Journal of Roman Archaeology Supplementary Series.
- Park, H. S., H. M. Lee, Hojjat Adeli, and I. Lee. 2007. “A New Approach for Health Monitoring of Structures: Terrestrial Laser Scanning.” *Computer-Aided Civil and Infrastructure Engineering* 22 (1): 19-30.
- Perucchio, R., and P. Brune. 2008. “Concrete vaulting in Imperial Rome: A structural analysis of the Great Hall of Trajan’s Markets.” Paper presented at the *6th International Conference on Computation of Shell and Spatial Structures IASS-IACM 2008: 28-31 May 2008, at Cornell University, Ithaca, NY, USA*.
- Riveiro, B., J. C. Caamaño, P. Arias, and E. Sanz. 2011. “Photogrammetric 3D modelling and mechanical analysis of masonry arches: An approach based on a discontinuous model of voussoirs.” *Automation in Construction* 20 (4): 380-8.
- Sauerbier, M., M. Kunz, M. Fluehler, and F. Remondino. 2004. *Photogrammetric reconstruction of adobe architecture at Túcume, Peru*. Zürich: ETH, Eidgenössische Technische Hochschule Zürich, Institut für Geodäsie und Photogrammetrie.

- Shih, Naai-Jung, Huey-Jiun Wang, Chen-Yan Lin, and Chai-Yuan Liau. 2007. "3D scan for the digital preservation of a historical temple in Taiwan." *Advances in Engineering Software* 38 (7): 501-512.
- Sinning-Meister, M., A. Gruen, and H. Dan. 1996. "3D city models for CAAD-supported analysis and design of urban areas." *ISPRS Journal of Photogrammetry and Remote Sensing* 51 (4):196-208.
- Slavid, R. 2006. *Wood Houses*. Kansas: Abbeville Press.
- Sturzenegger, M., and D. Stead. 2009. "Close-range terrestrial digital photogrammetry and terrestrial laser scanning for discontinuity characterization on rock cuts." *Engineering Geology* 106 (3-4): 163-182.
- Sundstedt, V., A. Chalmers, and P. Martinez. 2004. "High fidelity reconstruction of the ancient Egyptian temple of Kalabsha." In *Proceedings of the 3rd international conference on Computer graphics, virtual reality, visualisation and interaction in Africa. Stellenbosch, South Africa*, edited by Lynette van Zijl and Patrick Marais, 107-113. New York: ACM Press.
- Ulrich, R.B. 2007. *Roman Woodworking*. Michigan: Yale University Press.
- Yastikli, N. 2007. "Documentation of cultural heritage using digital photogrammetry and laser scanning." *Journal of Cultural Heritage* 8 (4): 423-427.
- Zienkiewicz, O. C., R. L. Taylor, Robert Leroy Taylor, and J. Z. Zhu. 2005. *The Finite Element Method: Its Basis and Fundamentals*. Oxford: Elsevier.

Telling the Story of Ancient Coins by Means of Interactive RTI Images Visualization

Gianpaolo Palma, Eliana Siotto
Visual Computing Lab, Pisa, Italy

Marc Proesmans
Katholieke Universiteit Leuven, Belgium

Monica Baldassari
Museo Civico, Montopoli in Val d'Arno, Italy

Clara Baracchini
MIBAC, Pisa, Italy

Sabrina Batino and Roberto Scopigno
Visual Computing Lab, ISTI CNR, Pisa, Italy

Abstract:

We present a system to provide an enhanced access of ordinary people to a collection of ancient coins, preserved in the National Museum of San Matteo in Pisa (Italy). Those coins have been digitized to produce RTI representations and are presented to museum visitors together with data telling their story.

The interactive presentation system, designed to be deployed either by a museum kiosk or by a web site, is composed of two integrated sections: a brief introduction to the different subsets of coins; an interactive RTI viewer. The coins are organized in different subsets to better present them to a public of inexperienced visitors. The viewer supports the interaction with the RTI image through a set of functionalities: changing the light direction; zooming; panning over the image; flipping the coin; visualizing hot spots, which link specific areas on the coin with descriptive HTML hypertext.

Keywords:

RTI Images, Ancient Coins, Interactive Kiosk, WebGL, SpiderGL

1. Introduction

Reflectance Transformation Imaging (RTI) is a computational photography method that, starting from a set of images taken from a single view under varying lighting conditions, encodes the subject's surface shape and colour to enable the interactive re-lighting of the subject from any direction. RTI encodes this redundant data, the same scene sampled under many different lighting conditions, in a compact way, using view-dependent per-pixel reflectance functions.

The virtual examination and study of Cultural Heritage artefacts is taking advantage from the use of the RTI techniques. In this field the way light interacts with the object is very important because the characteristics of the material, reflectance

behaviour, and texture offer major perceptual and cognitive hints for the study of a Cultural Heritage object. In many cases the ability to interactively play with the light is often more useful than the manipulation of an accurately sampled 3D shape, that is hardly able to capture all the interesting aspects of the artwork. Furthermore, there are several advantages of RTI technologies compared to 3D scanning techniques: RTI techniques use inexpensive and widely available hardware (in many cases, just a digital camera and light), scale well to both large and very small objects, and are able to easily achieve a sampling density and a precision that most current 3D scanners are unable to reach, even under optimal acquisition conditions. For those reasons, RTI techniques are widely used in the Cultural Heritage field for documentation tools and to support detailed visual analysis.

Corresponding author: gianpaolo.palma@isti.cnr.it

RTI techniques were designed to support easy documentation and inspection of Cultural Heritage artworks, giving a precious instrument to the specialists in the analysis and interpretation process. The recent advances of the web visualization instruments are increasing our capability to disseminate data and to support remote visual inspection of both scholars and ordinary public.

We present an interactive kiosk for the interactive presentation and virtual inspection of coins collection using HTML 5 technology and RTI images; the specific test case, the coin collection of the National Museum of San Matteo in Pisa, gives us the opportunity to present the underlying technology in the framework of a concrete and practical project.

2. State of Art on RTI Acquisition and Visualization

The first RTI image format was introduced in 2001 (Malzbender, Gelb, and Wolters 2001), with the Polynomial Texture Maps (PTM). This image encodes the per-pixel reflectance function using a biquadratic polynomial. Already in this original work, the PTMs are presented as a powerful tool to improve the study of ancient writings and inscriptions. Authors proposed two contrast enhancement methods which allow improving the readability of the ancient inscription by mathematical manipulation of the biquadratic polynomial. A new encoding for RTI data was proposed in 2009 (Gunawardane et al. 2009), the Hemispherical Harmonics Map. This new format uses a linear combination of the first nine hemispherical harmonics for the reflectance function, enabling a higher rendering quality due to the better sampling properties of the new per-pixel function.

In the last 10 year RTIs were employed in several applications. They were used in Palaeontology, to provide noticeable improvement in imaging of low colour contrast, like high relief fossils (Hammer et al. 2002). The application of PTM method on ancient stone tools revealed fine details of concoidal knapping fractures, use scarring and stone grain (Mudge et al. 2006). A joint work done by National Gallery and Tate Gallery of London showed that PTMs provided additional information about the surface textures of oil paintings (Padfield, Saunders, and Malzbender 2005). Cuneiform

tablets were analyzed using both 2D (PTM) and 3D (structured light scanner) information; the PTMs were texture mapped on the model, and a special 3D viewer was created (Mudge et al. 2006). The application of PTMs and scanning techniques on a large numismatic collection permitted the creation of a more complete documentation than the traditional photographic methods and the communication of this information with ease through digital media (Mudge et al., 2005). PTMs were used to study the Antikythera mechanism, an ancient mechanical computer designed to calculate astronomical positions (Freeth et al. 2006). Here, the analysis of the different fragments using PTMs increased the readability of the inscriptions, allowing a more complete understanding of the mechanism operation. The potential and the advantages of the application of the RTI images in the archaeological context was analyzed in (Earl, Martinez, and Malzbender 2010; Earl et al. 2010), especially in the field of the conservation, analysis of the material and representation of the archaeological data.

Improved methods were proposed for the acquisition of RTI images. A method to acquire the RTI of large object, using a single moveable light and an acquisition plan without the employ of a light dome, was presented in (Dellepiane et al. 2006). An automatic method for RTI generation based on the tracking of the highlight position of the light source on glossy spheres was proposed in (Barbosa, Sobral, and Proença 2007). In this way the user is free to move a single light source around the object and, after the acquisition, he can use an automatic tool to estimate the encoded light direction of each photo.

Advanced shading enhancement techniques were proposed in (Palma et al. 2010). The authors proposed a set of enhancement operators for RTI images that improve the perception of details, features, and shapes depicted in the image and provide better and more flexible visual inspection capabilities.

3. The San Matteo Coins Project

The San Matteo coins project started from the request of the curators of the National Museum of San Matteo in Pisa to allow presenting the ancient coins collection of the museum in an innovative way, to better capture the interest of the visitors and

to give them enhanced information. The attention of the museum curator towards the coin collection is due to the current way to expose and present it at the public. A coin is a very small artwork, which in a standard museum exposition is presented to the public from a distance (typically at least 50cm far from the observer eyes). This distance does not allow to the visitors to note some small and interesting details on the legend or on bas-relief and, moreover, the coin is usually visible only from one side. Furthermore, coins have a lot of hidden knowledge that is difficult to transfer to the visitors in an easy, effective and understandable manner.

The main challenges of this project were:

- to allow the virtual manipulation of the coins to inspect them in detail;
- to bring some of the hidden knowledge of the coins to the ordinary public in an easy and understandable way.

We proposed the design and implementation of an interactive kiosk to allow the presentation and the virtual inspection of the coins collection. The kiosk must be easy and intuitive to use for the ordinary public of the museum, and it must allow the real-time manipulation of the coin with a set of basic operations like zooming, panning, flipping the coin, changing of the light direction. Finally the kiosk must tell the story of the coins using multimedia data, with a combination of text, images and videos. For this last purpose the coins are organized in several subsets, each one characterized by a feature that can be the historical collection to which the coins belong to or a common thematic subject (for example the coin of a specific geographic area or epoch). In addition, we put on the surface of the coins some hot-spots that contain additional multimedia information useful to better understand the most important and interesting details depicted on the coin engraved decorations.

The management of the virtual inspection of the coins is one of the most important aspects in the design of the kiosk. Several scientific researches (Ramachandran 1988; Adelson and Pentland 1996) proved that the human brain is able to infer more cognitive data from the dynamic reflection and shading of an object. Then it is a fundamental

requirement to give to the user the possibility to rotate the coin under, at least, a directional illumination. This means that we need a virtual representation of the coin that can simulate the illumination effects in real-time and in an accurate way, in order to obtain a photo-realistic rendering. In the choice of this virtual representation we have to pay attention because typically the production of a photo-realistic rendering of a coin is extremely complex, due to the reflection effects of the different types of materials. For example in the museum collection we have both high reflective and specular gold coins, and bronze coins, that are more opaque and present on their surface different kind of patinas and degradation processes that alter their appearance (Fig. 1).

There are two possible options for the virtual representation of the coin: a complete 3D model or a RTI image. The creation of a 3D model requires to acquire both the 3D geometry and the surface appearance. This task could be quite complex because scanning coins introduces several problems and drawbacks. The acquisition of the geometry can present problems with the scanning of the coin's border, that can be very thin, and the right alignment of the two sides of the coin. The acquisition of the reflectance must be done in a following step with special setups to sample in an accurate way its dimensionality. Even if several solutions have been proposed for the acquisition of the surface appearance of real object, all ones have some drawbacks. They require a very intensive data acquisition, that must sample in accurate way both the light and the view directions, and complex reflectance models that, due to the assumptions on the reflection effects that we want to capture, are not able to reproduce all type of materials. Finally, the processing of the acquired data for the creation of the final 3D model is time consuming and the manipulation of a 3D model is still more complex to understand and control for the user.

On the other hand, RTI techniques produce a 2D representation of each coin that encodes both the surface normal and the appearance in a single image. This image can be dynamically relighted by the user reproducing the illumination-dependent effects of the surface with a higher quality and a higher resolution that is not usually provided by 3D scanned model. The acquisition and processing step are cheaper than 3D scanning and the final

representation simplify the interaction because the user is more accustomed to interact with an image than with a 3D model. For these reasons our choice was to adapt RTI images for the San Matteo kiosk.

4. Acquisition and Processing of RTI data

The first step of the project was the acquisition and the generation of the RTI images. For the acquisition the museum curator selected a subset of 41 coins from the museum's collection, following value and storytelling criteria. The coins cover different epochs, from the Roman Empire to the Grand Duchy of Tuscany (XVI –XIX centuries).

For the digitalization we used a minidome designed by the University of Leuven (Willems et al. 2005) cuneiform tablets, either manually, by moving lights around the tablet to maximize readability, or by studying photographs (or drawn copies). The dome is composed by 4 shells that can easily assemble and disassemble to simplify the transport. It has 260 white LEDs and an overhead CCD camera, and it is computer controlled to allow a completely automatic acquisition (Fig. 2). For each coin the acquisition takes about 10 minutes, required to shot and store 520 photos (260 photos for each side). The data was recorded with 3 different integration times, to provide the necessary dynamic range for HDRI imaging and further analysis. Different sets of lenses and macro rings have been used to cope with the variety of sizes in coins. All the coins were digitalized in a single working day. It was made on site in the museum thanks to the features of the used light dome (it can be dismantled and transported easily). The acquisition phase was done by staff of the University of Leuven, but it is quite easy to do even for a person without any specific expertise in this field. In the specific the hardware setup requires only to mount the shell and the camera and to put the coins in the middle of the dome, while the software tuning needs only to set the exposure time in an interactive way with visual feedback.

The processing of the acquired raw data to produce the final RTI image involved 3 steps:

- the transformation of the raw image taken with the minidome from the Bayer Pattern to a RGB format;



Figure 1. Two example of coins inserted in the kiosk: (Left) a Roman bronze coin; (Right) a modern gold coin.

- the generation of an RTI images for each coin side using the tools provided by Cultural Heritage Imaging corporation (Cultural Heritage Imaging 2012);
- the construction of a multiresolution streamable RTI format that permits the asynchronous loading of the image, allowing to the user to interact immediately with the coin without the awaiting of the complete loading of the data.

For the generation of the RTI images we chose the HSH format because it guarantees a better reproduction of the specular reflection with results that are more photorealistic with respect to a PTM, as showed in (Mudge et al. 2008).

For the construction of the multiresolution format, we subdivided the RTI image in nine layers, one layer for each HSH coefficient (Fig. 3). The i -th layer contains the i -th coefficient of the three RGB colour channels. For each layer we created a multiresolution tree and we cut each level of the tree in tiles using a quad-tree structure (Fig. 4). Finally we saved each tile in a different PNG image. This means that to visualize a specific pixel we need to load the nine PNG images that contain the HSH coefficients of the pixel. We chose PNG format because it guarantees a lossless compression that reduces the amount of data to transfer from the server to the client without to lose any information needed for the photorealistic rendering. The advantage of multiresolution streamable format is the out-of-core loading of the tiles making immediately available at least some low resolution data, to allow to start quite instantaneously the user interaction with the coin. In the specific at the beginning of the loading the user interacts with a low resolution version of

the coin, which is progressively refined as soon as the higher resolution data are loaded. The loading of the tiles at the different resolutions is guided by the zoom and pan operation of the user.

The generation of the RTI images took about 24 hours of totally automatic processing, without user intervention.

5. The San Matteo Kiosk

The kiosk is composed by two integrated sections. The first section allows to introduce and present to the user the different subset of coins. The second section permits the interactive RTI visualization. The kiosk has been implemented by using the Community Presenter, a tool developed for designing informative kiosks in the 3DCOFORM project. This also demonstrated the flexibility of the tool that has been designed for managing 3D data and has been extended to manage an innovative 2D medium such as the RTI images.

The main features of the kiosk are:

- the organization of the coins in categories and the presentation of these categories with multimedia data;
- the virtual inspection of the coin by RTI manipulation. We associate to each coin a general presentation and some hot-spots, that are located on selected areas on the surface of the coin to tell the most important and significant details;
- the possibility to run the kiosk on a web site or on a touch screen system (an interactive installation inside the museum) thanks to the technologies used for the development: HTML and JavaScript for the general structure of the kiosk and for the presentation of multimedia data; WebGL and SpiderGL for the RTI visualization.

WebGL (Khronos Group 2009) is a library that extends the capability of JavaScript to allow the interactive generation of 3D content within any compatible web browser (Chrome, Firefox, Opera, Safari). The main advantage of WebGL, with respect to the previous solutions for the integration of 3D content in the web, is the absence of external plug-ins



Figure 2. *Minidome from the University of Leuven used for the RTI acquisition.*

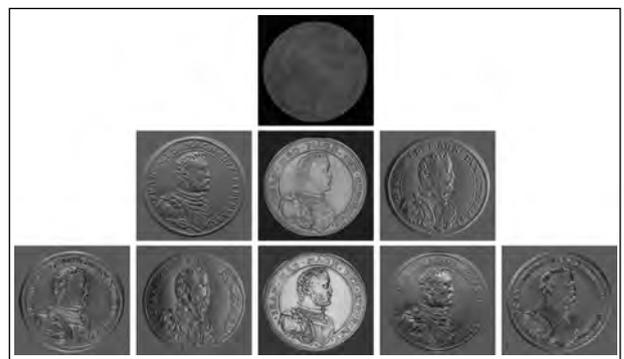


Figure 3. *Decomposition of the nine layers that compose the HSH image format.*

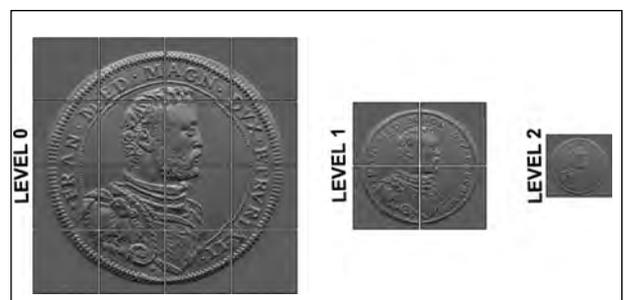


Figure 4. *Quad-tree encoding of the RTI image.*

to be installed, because WebGL is a built-in feature of modern web browsers. SpiderGL (Di Benedetto et al. 2010) is a computer graphics JavaScript library that provides a set of data structures and algorithms to simplify the development of a WebGL application. It is composed by a number of modules that allow to define and manipulate shapes, to import 3D models in several formats, to handle asynchronous data loading and to manage the user interaction. The use of WebGL and SpiderGL allows to customize the pipeline of the Graphics Processor Unit directly from the web browser with specialized shaders. The custom shaders permit to have a complete

interaction with the RTI image changing even the light direction in real time. This kind of interaction is not possible with a simple HTML viewer for 2D images because without WebGL we does not have the possibility to change the rendering of the object dynamically and in real-time.

The presentation section is characterized by a HTML template with a JavaScript engine that allows to load the multimedia content to visualize in the kiosk. The multimedia content is organized in XML files. For each category we store the multimedia HTML data and the list of its coins, while for each coin we store the general description and the position and the content of the hot-spots. The JavaScript engine loads dynamically the content from the XML files according the user choice and interaction.

The Figures 5-8 show some screenshots of the kiosk. Entering from the cover page (Fig. 5 - top left) the user has a presentation page of the project (Fig. 5 - top right). On the left there is a menu that allows to navigate among the different pages of the first section and to access directly the RTI viewer. In the middle there is a short description of the project. The items in the menu allow the access to the different subsets of coins. The subsets are subdivided into two categories: the historical collections (Figure 5 - bottom left) and the thematic subject (Figure 5 - bottom right). The San Matteo Museum's coins belong to three different historical collections:

- the Moise Supino's collection, dealer of Pisa in the 19th century that gave his coin collection to the museum;
- the Angiolo Franceschi's collection, archbishop of Pisa in the 18th century;
- the collection of coins findings, that contains the small treasures discovers in Pisa and surroundings in the 20th century.
- The thematic subjects are four:
 - the coinage techniques, that explores the different methods for the production of the coins;
 - the coins of the Tuscany, that presents the coins of the Tuscany cities in the different epoch;

- the iconography subject, that analyzes the changes in the representation of the main characters of the bas-relief in the different coins;
- the value of the coins, topic dedicated to the understanding of the economic value of the coins.

By selecting a subset, the user can read additional information about it (Fig. 6). The multimedia presentation content is showed in the middle of the page. The content is organized in several subsections that can be navigated and consulted with the arrows in the bottom of the page. On the left there is a scrollable bar with the thumbnail and the name of the coins in the subset. By clicking on a coin in this bar the user can open the visualization of the relative RTI image. In the page of RTI images (Figure 7) there is the RTI viewer in the middle, a general description of the coin on the left and the scrollable bar on the right. Using this bar the user can switch very quickly on other coins of the current subset. In the RTI viewer we have a title bar on the top with the name of the coins and the name of the subset used to access it, and a tool bar on the bottom that allows the manipulation of the images. The user can change the light direction, pan the image, zoom in and zoom out, flip the coin to see the other side, enable the visualization of the hot-spots. By clicking on a hot-spot the user can display the relative multimedia content on the left side of the page (Fig. 8). The visualization of the hot-spot's content is preceded by an automatic zoom animation on the detail associated to the hot-spot to better highlight it. The arrows in the bottom of the dialog allow to scroll among the images associated with the hot-spot.

The interactive kiosk will be installed at the end of 2012 at the National Museum of San Matteo in Pisa. The installation setup is composed by a 24-inch multi-touch screen, used for the user input and visualization, paired by a bigger screen set on the side that shows the same content of the touch screen, allowing a clean vision for the other visitors that don't interact directly with the kiosk.

6. Conclusions

We have presented the design and the implementation of an interactive kiosk for the visualization and inspection of a coins collection



using multimedia data and RTI images. The kiosk is designed to be easy and intuitive to use for the ordinary public of the museum, and to transfer a part of the knowledge of the coins using multimedia data, with a combination of text, images and videos. For this purpose the coins are subdivided in several subsets, to better present them according different criterions and aspects, and each coin is complemented by a set of hot-spots useful to better understand the most important and interesting details depicted on the engrave decoration. The overall methodology was successful since:

- coins are 3D, but usually most of the interest is in their front and back faces, rather than on the side;
- visual inspection of coins is performed usually by changing the light and gathering 3D information from the dynamically change of the reflection and shadows; this is exactly what is supported (at high quality) by the interactive inspection of the RTI medium, more faithfully than what is possible with the interactive rendering of 3D models;
- RTI 2D acquisition is faster and more automatic than the corresponding 3D acquisition, thus results in savings for the digitization phase;
- finally, we have showed how to include also this representation type in a tool for interactive presentation of 3D models (the Community Presenter), extending its capabilities and flexibility.

Figure 5. (Top-left) Cover page. (Top-right) Presentation page. (Bottom-left) The collections. (Bottom-right) The thematic subject.

Figure 6. Category content.

Figure 7. RTI viewer: (Top-left) starting page; (Top-right) activation of the hot-spots; (Bottom-left) coin detail; (Bottom-right) coin detail with a different light direction.

Figure 8. Visualization of the hot-spot's content.

To understand the real effectiveness of the current organization of the kiosk, we have implemented a data logging framework that allows to record the interaction of the users with the system in order to allow following analysis. From this analysis we will be able to evaluate different things, as the easiness of the interface, the relation of the user with the new RTI technology (time spent in the interactive manipulation session), simple statistics on the most viewed categories and coins, and so on.

Acknowledgements

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 231809 (IST IP "3DCOFORM").

References

- Adelson, E. H., and A. P. Pentland. 1996. "The Perception of Shading and Reflectance." edited by D.C. Knill and W. A. Richards. *Perception* 1: 409–423.
- Barbosa, J.G., J.L. Sobral, and A.J. Proença. 2007. "Imaging Techniques to Simplify the Ptm Generation of a Bas-relief." In *The 8th International Symposium on Virtual Reality Archaeology and Cultural Heritage VAST2007*, edited by D. Arnold and F. Niccolucci and A. Chalmers, 28–31, Brighton: Eurographics Association.
- Di Benedetto, M., F. Ponchio, F. Ganovelli, and R. Scopigno. 2010. "SpiderGL: a JavaScript 3D Graphics Library for Next-generation WWW." In *Proceedings of the 15th International Conference on Web 3D Technology* 1 (212), edited by D. G. Aliaga, M. M. Oliveira, A. Varshney and C. Wyman, 165–174. Los Angeles: ACM.
- Cultural Heritage Imaging. 2012. "RTI Software." Accessed February 2013. http://culturalheritageimaging.org/What_We_Offer/Downloads/index.html.
- Dellepiane, M., M. Corsini, M. Callieri, and R. Scopigno. 2006. "High Quality Ptm Acquisition: Reflection Transformation Imaging for Large Objects." In *The 7th International Symposium on Virtual Reality Archaeology and Cultural Heritage VAST2006*, edited by M. Ioannides, D. Arnold, F. Niccolucci, and K. Mania, 179–186. City: Eurographics Association.
- Earl, G., G. Beale, K. Martinez, and H. Pagi. 2010. "Polynomial Texture Mapping and Related Imaging Technologies for the Recording, Analysis and Presentation of Archaeological Materials." *Archives XXXVIII*: 218–23.
- Earl, G., K. Martinez, and T. Malzbender. 2010. "Archaeological Applications of Polynomial Texture Mapping: Analysis, Conservation and Representation." *Journal of Archaeological Science* 37 (8): 2040–2050.
- Freeth, T., Y. Bitsakis, X. Moussas, J. H. Seiradakis, A. Tselikas, H. Mangou, M. Zafeiropoulou, R. Hadland, D. Bate, A. Ramsey, M. Allen, A. Crawley, P. Hockley, T. Malzbender, D. Gelb, W. Ambrisco and M. G. Edmunds. 2006. "Decoding the Ancient Greek Astronomical Calculator Known as the Antikythera Mechanism." *Nature* 444 (7119): 587–591.
- Gunawardane, P., O. Wang, S. Scher, I. Rickard, J. Davis, and T. Malzbender. 2009. "Optimized Image Sampling for View and Light Interpolation." In *The 10th International Symposium on Virtual Reality Archaeology and Cultural Heritage VAST2009*, edited by K. Debattista, C. Perlingieri, D. Pitzalis, and S. Spina, 93–100. Malta: Eurographics Association.
- Hammer, Ø., S. Bengtson, T. Malzbender, and D. Gelb. 2002. "Imaging Fossils Using Reflectance Transformation and Interactive Manipulation of Virtual Light Sources." *Paleontologia Electronica* 5 (1): 1–9.
- Khronos Group. 2009. "WebGL - OpenGL ES 2.0 for the Web." Accessed February 2013. <http://www.khronos.org/webgl/>.
- Malzbender, T., D. Gelb, and H. Wolters. 2001. "Polynomial Texture Maps." In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques SIGGRAPH 01* 52 (August), edited by Eugene Fiume, 519–528. Los Angeles: ACM.
- Mudge, M., T. Malzbender, A. Chalmers, R. Scopigno, J. Davis, O. Wang, P. Gunawardane, M. Ashley, M. Doerr, A. Proença, J. Barbosa. 2008. "Image-Based Empirical Information Acquisition, Scientific Reliability, and Long-Term Digital Preservation for the Natural Sciences and Cultural Heritage." In *Tutorial Eurographics 08*, edited by Maria Roussou and Jason Leigh. Nicosia: Eurographics Association.
- Mudge, M., J.-P. Voutaz, C. Schroer, and M. Lum.

2005. "Reflection Transformation Imaging and Virtual Representations of Coins from the Hospice of the Grand St. Bernard." In *The 6th International Symposium on Virtual Reality Archaeology and Cultural Heritage VAST2005*, edited by M. Mudge, N. Ryan, and R. Scopigno, 29–39. Pisa: Eurographics Association.
- Mudge, M., T. Malzbender, C. Schroer, and M. Lum. 2006. "New Reflection Transformation Imaging Methods for Rock Art and Multiple-Viewpoint Display." In *The 7th International Symposium on Virtual Reality Archaeology and Cultural Heritage VAST2006*, edited by M. Ioannides, D. Arnold, F. Niccolucci, and K. Mania, 195–202. Nicosia: Eurographics Association.
- Padfield, J., D. Saunders, and T. Malzbender. 2005. "Polynomial Texture Mapping: a New Tool for Examining the Surface of Paintings." *ICOM Committee for Conservation I*: 504–510.
- Palma, G., M. Corsini, P. Cignoni, R. Scopigno, and M. Mudge. 2010. "Dynamic Shading Enhancement for Reflectance Transformation Imaging." *Journal on Computing and Cultural Heritage* 3 (2): 1–20.
- Ramachandran, V S. 1988. "Perception of Shape from Shading." *Nature* 331 (6152): 163–166.
- Willems, G., F. Verbiest, W. Moreau, H. Hameeuw, K. Van Lerberghe, and L. Van Gool. 2005. "Easy and Cost-effective Cuneiform Digitizing." In *The 6th International Symposium on Virtual Reality Archaeology and Cultural Heritage VAST2005*, edited by M. Mudge, N. Ryan, and R. Scopigno, 73–80. Pisa: Eurographics Association.

IBISA 3D: Image-Based Identification/Search for Archaeology Using a Three-dimensional Coin Model

Sylvain Marchand

University of Brest, France

Abstract:

The IBISA system was designed to help the user decide, from their images, if two coins are either the same, come from the same die, share resemblance in style, or are completely different. This decision is extremely useful for the study of coin hoards, which can inform us about the original production process, chronology, geography, etc. IBISA uses computer vision methods to get rid of the viewing conditions when searching for similarities in the databases. To cancel rigid transforms (translation, rotation, zoom) inherent to the photography, the system uses a global registration method. The present work gives an overview of the original IBISA system, characterizes the errors in the estimation of the rigid transform by the registration method, as well as the impact of the light conditions. Then the extension of the system to a three-dimensional coin model is presented, yielding a more reliable estimation.

Keywords:

Numismatics, Ancient Coins, Image Registration, Image Similarity, Shape From Shading

1. Introduction

IBISA (Image-Based Identification/Search for Archaeology) is a software system (Marchand et al. 2009) that manages databases of digital images of archaeological objects, and allows the user to perform searches by examples. The objects are only required to be quasi flat (approximately two-dimensional) and produced from matrices via some striking / stamping / casting process. The original matrices are generally lost now, but many objects with their prints can still be found, with many similarities among them. The IBISA system was originally designed to work with ancient (Greek, Roman) coins. In this numismatic context, the matrix is then a die.

By taking advantage of the relations (die links, style similarities) among large finds (coin hoards), one could determine the original production process, chronology, geography, or even economical or social issues (Bresson 2005). The difficulty is to handle a large number of objects, very similar at first sight for a non-specialist, and to compare them. This task is time-consuming, quite exhausting, and thus error-prone. IBISA was designed to help the user decide, from their images, if two objects are either the same, *Corresponding author: sylvain.marchand@univ-brest.fr*

come from the same matrix, share resemblance in style, or are completely different.

IBISA uses computer vision methods to make this decision while getting rid of the viewing conditions when searching for similarities in the databases. The method works on the image in its globality, and does not rely on specific features or key-points. To avoid any colorimetric issue, only the luminance (intensity) of the pixels is considered. This enables also the comparison of coins struck using the same dies but with different metals (such as *denarii* and *aurei* of the Roman Empire). To get rid of the background context, a segmentation method based on active contours first extracts the useful part of each image. To cancel any translation, rotation, or zoom inherent to the photography, the system uses a registration method based on the Fourier-Mellin transform, which can estimate the parameters of such rigid transforms. These transforms can then be inverted, to get the best match for the images. For now, the similarity measure is given by the cosine similarity among the target image and the registered version of each candidate in the database.

The last viewing condition problem is the light source position. Indeed, the shading can change the cosine similarity by more than twenty percents.

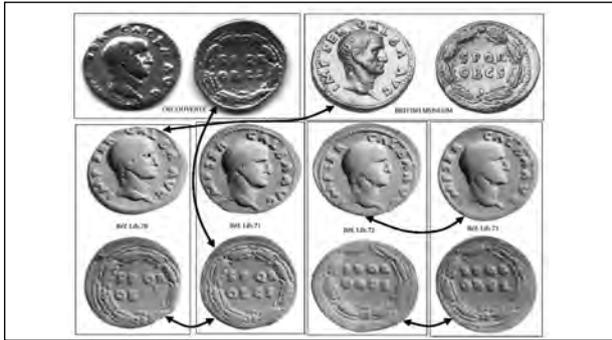


Figure 1. Studying ancient coin hoards: die links (see arrows) for 6 Roman gold coins (aurei) of the emperor Galba (68-69 AD) - an exemplar under investigation (Parisot, Prieur, and Schmitt 2008) at the upper left corner, an exemplar of the British Museum (Sear 2000) at the upper right corner, and the 4 exemplars from the Liberchies hoard (Thirion 1972) at the bottom. It turns out that the first coin is indirectly linked to the one in the British Museum.

The solution proposed here is to consider a three-dimensional model, the elevation map replacing the luminance information.

The remainder of this article is organized as follows. Section 2 gives a brief overview of the original IBISA system. It presents the archaeological objects under consideration and their digital images, together with the problematics they raise. Section 3 details the registration method and characterizes the errors in the estimation of the rigid transform. Section 4 shows the impact of the light source direction on this estimation and on the similarity, and proposes an extension of the system to a three-dimensional coin model, yielding a more reliable estimation and similarity. The problem is now to be able to get the model from images. A simple yet efficient solution to this photometric problem is also proposed. Finally, Section 5 concludes by giving directions for further research.

2. Search by Example

The IBISA system manages databases of digital images of archaeological objects. From a given image, the system is able to sort the other images by decreasing similarity. From the system will come out first the same object with different viewing conditions, then other objects coming from the same matrix, objects of the same style, and finally very different objects.

2.1 Physical Objects

The archaeological objects are only required to be quasi flat (two-dimensional) and produced from matrices via some striking / stamping / casting process. The original matrices are generally lost now, but many objects with their prints can still be found, with many similarities among them, although these objects underwent some alteration (wear, patina, or even breaks) over centuries. For now, the system works with ancient (Greek, Roman) coins.

In the case of ancient coins, a design was engraved on a metallic die, used in turn often by pair (obverse / reverse) to strike the coins. The original dies are now lost, but many coins can still be found, sharing many similarities among them. IBISA should allow the user to identify an isolated specimen and facilitate the study of coin hoards (Fig. 1).

2.2 Digital images

It is more convenient to manipulate digital images instead of the archaeological objects themselves. One can photograph or scan the objects, or even use existing pictures from books or databases. The system must be resistant to the viewing conditions, since they are generally unknown. Pictures from books often lack the chromaticity information and suffer from Moiré patterns. Also, sometimes the scale is not specified. Pictures from camera nearly always lack this scale information, and are neither perfectly centred nor aligned.

To get rid of the colorimetry and Moiré problems, our system manipulates gray-scale images in the spectral domain. Let us denote by $g(p)$ the (gray-scale) value of the image g at the point p .

2.3 Image Similarity

Although determining the similarity between two images g_1 and g_2 is a very complex problem, we have shown that using the cosine distance, given by the classic inter-correlation factor

$$F(g_1, g_2) = \frac{\sigma_{12}}{\sigma_1 \sigma_2} = \frac{\sum_p (g_1(p) - \bar{g}_1)(g_2(p) - \bar{g}_2)}{\sqrt{\left(\sum_p (g_1(p) - \bar{g}_1)^2\right) \left(\sum_p (g_2(p) - \bar{g}_2)^2\right)}}$$

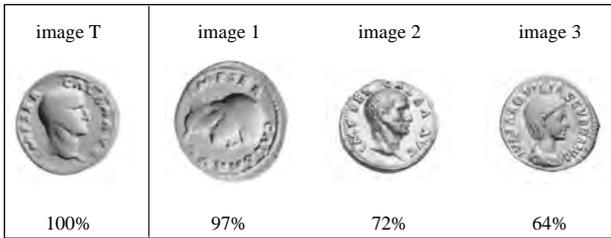


Figure 2. Similarity between images: the leftmost image is the target T , followed by the result of a search by example in our test database. The similarity factor is indicated below the images. The coin on image 1 is indeed from the same die - and the system suggests for T a translation, a rotation of angle $=\pi/2$ rad, and a homothety of ratio 1.5, for the two coins to superimpose optimally (97% of similarity). The next coin (image 2) is of the same type (but not of the same die), and the coin on image 3 is not even from the same emperor.

where $\bar{\mathbf{g}}$ is the mean of \mathbf{g} .

yields very good results in practice, as shown in Figure 2, provided that the second (target) image has been previously aligned with the first (reference) image. Thus, the core of the system is the registration method, designed to get rid of any viewing conditions.

3. Image Registration

To avoid any colorimetric issue, only the luminance (intensity) of the pixels is considered. This enables also the comparison of coins struck using the same dies but with different metals (such as *denarii* and *aurei* of the Roman Empire). To get rid of the background context, a segmentation method based on active contours first extracts the useful part of each image, see (Marchand et al. 2009). To cancel any translation, rotation, or zoom inherent to the photography, the system uses a registration method based on the Fourier-Mellin transform, which can estimate the parameters of such rigid transforms. These transforms can then be inverted, to get the best match for the images.

3.1 Rigid transformation

A geometric transformation is defined by

$$\begin{aligned} T: \\ p \mapsto p', p' = T(p), \\ g \mapsto g', g'(p') = g(p). \end{aligned}$$

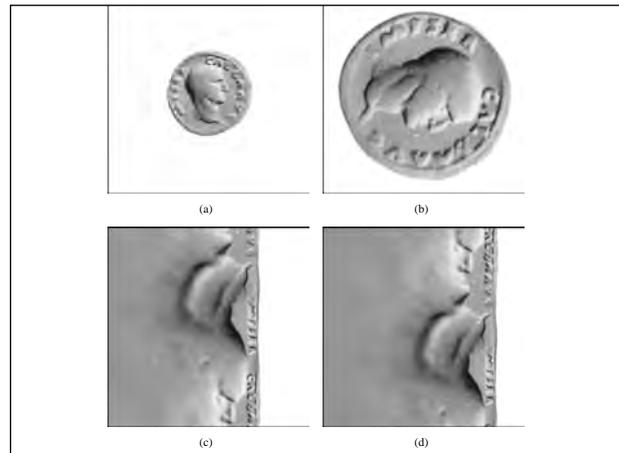


Figure 3. Effects of a rotation of angle $\pi/2$ rad plus a homothety of ratio 2 in Cartesian (top) and log-polar (bottom) coordinate systems. In the Cartesian case, the original picture (a) gets rotated and scaled (b), whereas in the log-polar case, the original picture (c) is simply translated (d) by $(\log(2), \pi/2)$.

A translation of vector $t = (\Delta_x, \Delta_y)$ can be easily expressed using complex numbers with the Cartesian form (where x is the abscissa and y is the ordinate):

$$p' = p + t \text{ where } p' = x' + iy', p = x + iy$$

$$\text{and } t = (\Delta_x, \Delta_y).$$

For a rotation+homothety of angle φ and ratio s , the polar form is more suitable (where ρ is the distance and θ is the angle):

$$p' = rp \text{ where } p' = \rho' e^{i\theta'}, p = \rho e^{i\theta}, \text{ and } r = s e^{i\varphi}.$$

As shown in Figure 3, this rotation+homothety is equivalent to a translation in the log-polar representation, since:

$$\log(p') = (\log(s) + i\varphi) + \log(p).$$

Without loss of generality, any rigid transformation - very likely to occur in photographs - can be expressed as one translation followed by one rotation+homothety:

$$p' = (p + t)r.$$

In practice, the images are discrete. To compute each sample (pixel) of the new image $g'(p')$,

the corresponding value has to be extracted from the original image at position p . This is very likely to occur outside its pixel map, but the value $g(p)$ can be reconstructed from the neighbouring samples (for example by bilinear interpolation).

3.2 Registration method

The registration method aims at cancelling this rigid transformation. It consists in estimating the parameters of the transformation in order to invert it. We use a registration method based on the Fourier-Mellin transform (Wolberg and Zokai 2000) to find the optimal superposition of the two images.

It works in the spectral domain, using the spectrum G of the image g given by the Fourier transform

$$G(\omega) = \int_p g(p)e^{-i\omega p} dp.$$

Recovering a translation $g'(p') = g(p'-t)$ considering the spectral domain $G'(\omega) = G(\omega)e^{-i\omega t}$ is possible using the phase-correlation technique, which consists in calculating the G^*G' product, G^* being the conjugate of G . More precisely, we have

$$\frac{G^*(\omega)G'(\omega)}{|G^*(\omega)G'(\omega)|} = \frac{G^*(\omega)G(\omega)}{|G(\omega)|^2} = e^{-i\omega t}$$

which inverse Fourier transform is an impulse located at t . Thus, estimating the translation t is only a matter of searching for a maximum.

Recovering a rotation+homothety $g'(p') = g(p'/r)$ is not more difficult, since it reduces in the log-polar representation to the translation case (see above).

We recall then two properties of the Fourier transform:

1. The translation does not modify the amplitude spectrum (its only effect is a phase shift).
2. The spectrum of a rotated image is the rotation of the spectrum of the original image (similar property for the homothety, but with the inverse ratio).

In practice, the latter property does not really

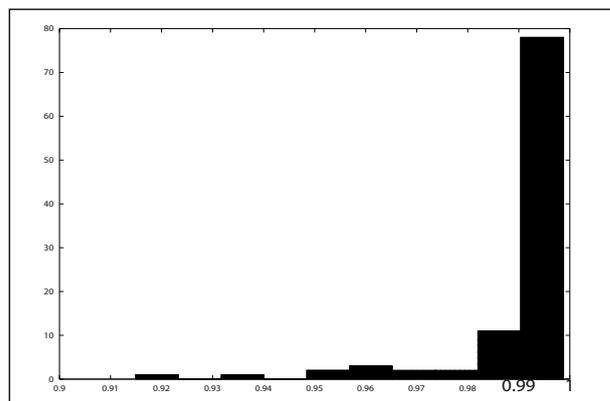


Figure 4. Distribution of the similarity factor measured with a random reference image and a target image generated from this image using a random rigid transformation. In theory, they should be regarded as perfectly similar. In practice, the similarity is excellent (above 0.99 most of the time).

hold for discrete images. However, the aliasing problem can be greatly reduced by applying a smoothing window to the images together with a high-pass filter on their spectra.

Finally, the full registration algorithm is the following:

1. Find the rotation+homothety r (estimate (φ, s))
 - By finding a translation in the log-polar system,
 - By considering the amplitude spectra of the images to ignore the effects of the translation t .
2. Invert the rotation+homothety (rotation of angle $-\varphi$ and homothety of ratio $1/s$);
3. Find the translation t (estimate Δ_x, Δ_y), now free from any rotation or homothety;
4. Invert the translation (translation of vector $(-\Delta_x, -\Delta_y)$).

3.3 Performance

In order to test the performances of the registration method, we randomly select one reference image, apply to it a rigid transformation chosen randomly (with uniform choices of the rotation angle in the $[-\pi, +\pi]$ interval, the scaling

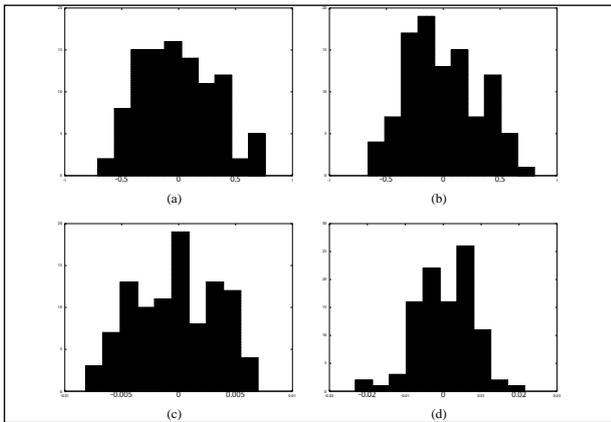


Figure 5. Distribution of the estimation errors for the rigid transformation: the horizontal (a) and vertical (b) translations in pixels, the rotation angle in radians (c), and the scaling factor of the homothety (d).

factor in the $[0.2, 2]$ interval, and sub-pixel translations) in order to get the target image. This transformed image can (rarely) fit only partly in the screen, thus simulating “chipped coins” (broken coins where a part of the flan is missing). As shown in Figure 4, the after-registration similarity is excellent (above 0.99). The resistance to chipped coins (partial images) is also demonstrated (since the similarity stays high even in those cases). However, very rare (less than one in 100 trials) but very extravagant results were noticed, which still have to be investigated.

Looking into details (Fig. 5), the estimation of the translation vector is unbiased (zero mean) and efficient (low variance). The absolute error is below 0.5, which is normal since the estimation is in pixels while the translation was sub-pixel. The error is below 0.005 rad for the rotation angle, and below 0.02 for the scaling factor, which is quite remarkable.

4. Image to Model

The results of the previous section show that the registration method is quite robust. However, these results are valid provided that the light conditions are constant. By chance, it was the case most of the time in our databases, since the coins are usually photographed using one light source at the top right of the coin, which is an excellent choice e.g. for Roman imperial coins, where the emperor has generally his portrait facing right, thus facing the

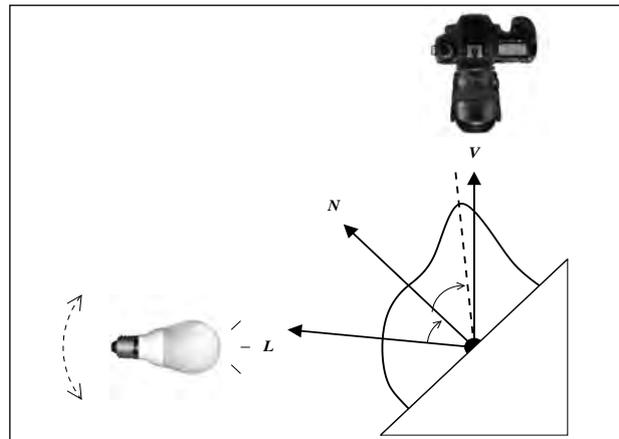


Figure 6. In all the our experiments, the camera and the object are fixed, and the light can be moved around the object. We denote by N , L , and V the normal, light, and view vectors, respectively. The observed intensity is maximal when the angles between L , N and between N , V are the same (i.e. when $L.N = N.V$).

light. But this condition does not necessarily hold for other type of coins.

4.1 Light conditions

The study of the impact of the position of the light source was studied in two ways.

First, a simulation was performed using a 3D coin model from (Zambanini et al. 2009), for the *aureus* of the Roman emperor Herennius Etruscus. More precisely, the elevation map of the obverse was extracted from the 3D model, then the normal map was derived from the elevation map, and the rendering was done using ray casting with the Ward’s model approximating the bidirectional reflectance distribution function (BRDF) for the gold material (see Fig. 10 for graphical examples).

Let us denote by N , L , V the normal, light, and view vectors. Moreover, let us denote by $H = (L + V)/2$ the “half vector”, and by δ the angle between the N and H vectors.

For each pixel, with the Ward’s model (Schwenk 2008) the RGB color is

$$C_{(R,G,B)} = a + (d + sk_s)(N.L)$$

where the specular part of the BRDF is given by

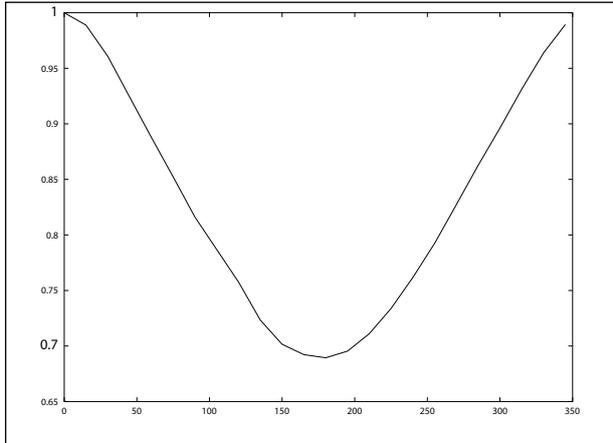


Figure 7. Simulated similarity factor degradation as a function of the relative light source azimuth. The factor goes down to 0.7, which indicates a sensibility of the similarity measure to the light conditions.

$$k_s = \frac{1}{\sqrt{(N,L)(N,V)}} \frac{\exp\left(-\frac{\tan^2 \delta}{\alpha^2}\right)}{4\pi\alpha^2}$$

the ambient, diffuse, and specular constants, as well as the roughness for the gold material being $a=(0.3, 0.3, 0.3)$ $d=(0.0468, 0.0203, 0.00167)$, $s=(0.279, 0.163, 0.0425)$, and $\alpha=0.243$ see (Ngan, Durand, and Matusik 2005).

A reference image was generated, then the light source was rotated all around the coin to get target images (Fig. 6), and for each light source angle the similarity factor with the first reference image was computed. The results are shown in Figure 7, indicating a sensibility of the similarity factor to the light / shading conditions.

Second, an experiment was performed using a real Roman coin (from a private collection), again with fixed object and objective positions, and a light source moving around the object (variable azimuth) (Fig. 8).

Again, for each light source angle the similarity factor with the first image was computed. The results are shown in Figure 9. The trend is exactly the same, with an impact on the similarity factor even more dramatic (as an effect of the small elevation angle).

Thus the light conditions have a strong impact on the luminance, and thus on the similarity factor / registration algorithm, which can be problematic.



Figure 8. Experimental setup with fixed object and objective positions, and a light source moving around the object (variable azimuth) at a fixed elevation (here 16 degrees).

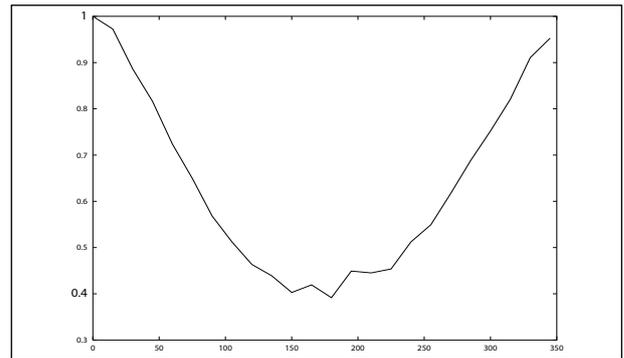


Figure 9. Experimental similarity factor degradation as a function of the light source azimuth. The factor goes down to 0.4, indicating a clear sensibility of the similarity measure to the light conditions.

Fortunately, a solution is that the registration algorithm of Section 3 can be applied on the elevation, instead of the luminance (intensity). The image registration turns into a model registration. The resulting accuracy depends on the precision of the model, and the problem is now to get the elevation information.

4.2 Model acquisition

Getting the elevation information is possible from several images with fixed object and camera but different light source positions, using a very simple shape-from-shading principle (Fig. 6), close to Reflectance Transformation Imaging - RTI



Figure 10. Synthetic data: elevation map (a), normal map (b), and rendering with different light conditions (c and d).

(Mudge et al. 2005):

- For each pixel, get the luminance distribution with respect to the light source position.
- The peak in the luminance distribution is when the light+view and normal vectors are in the same direction ($H - N$) with a special case when all the vectors are the same $L = V = H = N$.
- Since we know the light and view vectors, we can deduce the normal vector.
- From the normal map, we can integrate the elevation map.

In fact, the last step is trickier, since the integration introduces arbitrary constants. Fortunately, adding a constant or multiplying by a constant the image data has no impact on the cosine similarity factor.

Again, the validation of this algorithm was done in two different ways.

First, we ran a simulation where we used the same 3D coin model as previously, rendering images for different random light source positions (Fig. 10). We got acceptable results (in comparison to the original elevation map, known from the model) after approximately 100 images.

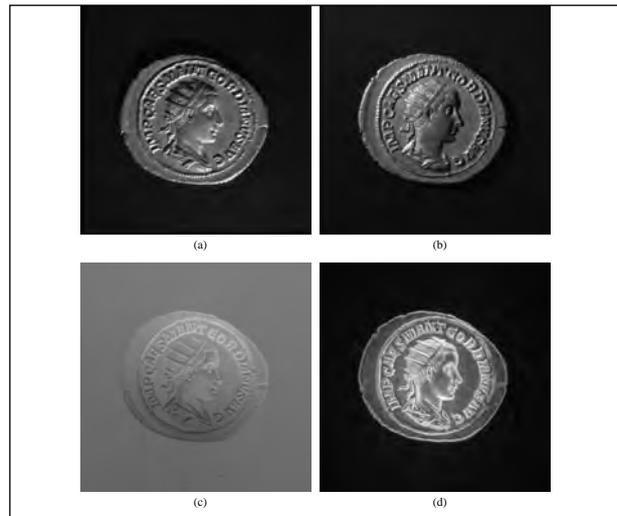


Figure 11. Real data: from photographs with different light conditions (a and b), the shape-from-shading algorithm can estimate the normal map (c), the integration of which giving the elevation map (d).

Second, we performed an experiment with 24 photographs with the same elevation (26 degrees) but different azimuths (each 15 degrees). The result is illustrated in Figure 11.

4.3 Image-Model registration

We have shown that image (intensity) registration is sensitive to the light conditions. We have also shown that model (elevation) registration can be a solution. But with the proposed method for model acquisition, an access to the object is necessary (to get several images of it). In practice, most databases contain only one image for each object. Moreover, as mentioned above, computing the elevation from the normals can be problematic.

However, it is important to note that the normals are sufficient for rendering an image with any light condition. In the intermediate situation where a model has to be compared to an image, a solution can be to render the target image from the model with the most suitable light source position so that the similarity with the reference image is the greatest. This is easy if the light source position can be estimated in the reference image, and much more complex if some exhaustive search has to be done. This opens up new research perspectives for the future.

5. Conclusions and Future Work

In this article, we have presented the current status of the ongoing IBISA project. This software system manages digital images of archaeological objects and helps the user study ancient coins hoards by discovering similarities. The key part of the system is the registration method that maximizes the similarity between the reference and target images. We have shown that this intensity-based method is quite robust, provided that the light conditions are the same though. To be resistant to these conditions, we propose to use the same algorithm but to consider the elevation instead of the intensity (luminance). The acquisition of the resulting three-dimensional coin model can be done in theory using a simple shape-from-shading technique, and in practice with a Reflectance Transformation Imaging (RTI) dome. Such a dome indeed provides multiple light conditions with fixed object and objective positions. We are now building our own dome, adapted to ancient coins. Part of our future research is also the evaluation / validation of the system on real - large - databases.

References

Bresson, A. 2005. "Coinage and Money Supply in the Hellenistic World." In *Making, Moving, Managing, The New World of Ancient Economies (323-31 BCE)*, edited by Z. H. Archibald, J. K. Davies, and V. Gabrielsen, 44-72. Oxford: Oxbow Books.

Marchand, S., P. Desbarats, A. Vialard, F. Bechtel, A. Ben Amara, B. Cicuttini, J.-P. Bost, A. Bresson, K. Konuk, and A. Beurivé. 2009. "IBISA: Image-Based Identification / Search for Archaeology." In *Proceedings of the International Symposium on Virtual Reality, Archaeology and Cultural Heritage (VAST)*, edited by K. Debattista, C. Perlingieri, D. Pitzalis, and S. Spina, 57-60.

St. Julians, Malta: The Faculty of ICT, University of Malta.

Mudge, M., J.-P. Voutaz, C. Schroer, and M. Lum. 2005. "Reflection Transformation Imaging and Virtual Representations of Coins from the Hospice of the Grand St. Bernard." In *Proceedings of the International Symposium on Virtual Reality, Archaeology and Cultural Heritage (VAST)*, edited by M. Mudge, N. Ryan, and R. Scopigno, 29-39. Pisa: A. K. Peters Ltd.

Ngan, A., F. Durand, and W. Matusik. 2005. "Experimental Analysis of BRDF Models." In *Proceedings of the Eurographics Symposium on Rendering (EGSR)*, edited by K. Bala and P. Durte, 117-226. Konstanz, Germany.

Parisot, N., M. Prieur, and L. Schmitt. 2008. *Rome XXII*. Paris: Compagnie Générale de Bourse.

Schwenk, K. 2008. *A Survey of Shading Models for Real-time Rendering*. Accessed June 21, 2012. http://www.karsten-schwenk.de/downloads/a_survey_of_shading_models.pdf.

Sear, D. R. 2000. *Roman Coins and Their Values*, vol. 1. London: Spink and Son Ltd.

Thirion, M. 1972. *Le trésor de Liberchies*. Bruxelles: Pro Geminiaco a.s.b.l.

Wolberg, G., and S. Zokai. 2000. "Robust Image Registration Using Log-Polar Transform." In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, Vancouver: IEEE.

Zambanini, S., M. Schlapke, M. Kampel, and A. Müller. 2009. "Historical Coins in 3D: Acquisition and Numismatic Applications." In *Proceedings of the International Symposium on Virtual Reality, Archaeology and Cultural Heritage (VAST)*, edited by K. Debattista, C. Perlingieri, D. Pitzalis, and S. Spina, 49-52. St. Julians, Malta: The Faculty of ICT, University of Malta.

Using Image Analysis to Match a Coin to a Database

Sebastian Zambanini and Martin Kampel

Vienna University of Technology, Austria

Abstract:

In this paper we present a method for automatic ancient coin classification based on computer vision. The method takes a coin image as input and finds the most visually similar coin in the database. For this purpose, a flexible and powerful image matching method is used that is able to cope with the possibly large degree of variation between coins of the same type. Our system is evaluated on a dataset containing 24 types of Roman Republican Coins and achieves a classification rate of over 68 % on the obverse coin sides and over 73 % on the reverse coin sides.

Keywords:

Computer Vision, Ancient Coins, Image Classification

1. Introduction

Determining the type of an ancient coin is in general a time consuming task and needs a lot of numismatic experience. Therefore, an automatic method for this task would be of high usage for the numismatic community. 2D images can serve as input for such a system as they are easy and cheap to produce and are widely available in museum databases and digital online archives. Potentially, such a methodology is able to act as a supporting tool for numismatists and can thus enable a much faster processing of coins. In the long run, an automatic image based coin classification system could be of use for a broad range of the numismatic community, e.g. by means of a freely accessible online coin classification tool.

This paper presents a method which uses image analysis in order to determine the classes of the coins. We argue to use image matching for this task, as it is able to measure the similarity of coins between images and can consequently be used for classification: the method measures the similarity between a coin image and all coin images in the training set (the database) and finally chooses the class with the highest image similarity. We use a flexible image matching method which is able to deal with the local spatial variations of features within a class, i.e., of the kind shown in Figure 1. The method is based on local SIFT features (Lowe, Distinctive Image Features from Scale-Invariant Keypoints 2004) which are extracted on a dense field in the

Corresponding author: zamba@caa.tuwien.ac.at

image and matched by means of optimization of a cost function (Liu, Yuen und Torralba 2011).

2. Related Work

Related work in image based classification of ancient coins is scarce so far. Recent research approaches for coin classification focused mainly on the classification of present day coins (Huber, et al. 2005) (Nölle, et al. 2003) (Reisert, Ronneberger und Burkhardt 2006). However, the differences between present day and ancient coins exposed to be too large to effectively apply such methods to ancient coins (van der Maaten und Poon 2006). Due to the abrasions over the centuries and the non-industrial manufacturing, ancient coins naturally exhibit a larger variation within a class. This impedes the coin classification problem to a large degree



Figure 1. Reverse side of six coins of our database. Coins in the same row belong to the same type.



Figure 2. Matching two coins of the same type, from left to right. query image, database image, database image warped back to be aligned with the query image.

since a classification method has to determine the features common to class while ignoring the coin-specific ones. On the other side, this fact favors the automatic identification of ancient coins. For instance, the shape of the coin border has shown to be a very discriminative feature to identify individual coins (Huber-Mörk, et al. 2010).

First promising results on the classification of ancient coins were presented in (Kampel und Zaharieva 2008). Similar to our work in their paper a classifier-free approach based on SIFT matching is used for coin classification. Image similarity is measured by counting the number of matched SIFT features between two images. Compared to our image similarity based on SIFT flow, this methodology provides a much simpler and thus less robust measure of coin similarity, as only a sparse set of interest points is matched without considering their spatial relationships. The authors report a classification rate of about 90%, but the dataset used contained only three coin classes. Recently, a more sophisticated method based on feature classification has been proposed by Arandjelovic (Arandjelovic 2010). In his work the so-called *locally-biased directional kernel feature (LBD)* is proposed to encode the spatial context of local interest points. A classification rate of 57% on a testset containing 65 classes is reported which shows that spatial-context features like LBD are generally promising as they offer a powerful representation able to capture the class-specific coin appearance. However, in this work, we aim at a classifier-free approach for coin classification that can be easily extended to large databases.

3. Coin Classification Methodology

SIFT flow was recently introduced by (Liu, Yuen und Torralba 2011) as a dense image matching method that finds nearest neighbors for each pixel



Figure 3. Matching two coins of different types, from left to right. query image, database image, database image warped back to be aligned with the query image.

in the presence of large image variations. Thus, SIFT flow does not match points by means of a global transformation between images but rather aims to find the most likely pixel-to-pixel matching based on well-defined constraints: local appearance similarity, smoothness and small displacement.

The adoption of the SIFT flow algorithm for coin classification relies on the following idea: matching two coin images of the same class will likely produce a lower final result of the cost function to be minimized than matching coin images from different classes, since a smooth matching can be more likely found in the former case. As no global rigid transformation is involved in the correspondence search, SIFT flow is able to cope with the high intra-class variability arising from abrasions and manual manufacturing. An example for SIFT flow matching is shown in Figure 2 and 3. Matching the test coin image with a coin image from the same class produces a reasonable result, as can be seen in Figure 2. In contrast, matching the test coin image with a coin image from a different class produces an unsuitable result and thus a higher cost (Fig. 3). As a consequence, we achieve coin classification by matching the coin image with all coin images in the database and finally choosing the class of the image with lowest costs.

In detail, the following steps are conducted for coin classification. The first two steps are also visualized in Figure 4.

Coin Segmentation: In order to remove background clutter such as rulers in the image, the coins have to be segmented first. For this task we use a shape-adaptive thresholding method (Zambanini und Kampel 2009) that has proven to provide a robust segmentation for a variety of coin images. It uses a range and entropy filter under the assumption that the coin in the image provides



Figure 4. Preprocessing steps of our coin classification system, from left to right. original input image, segmented coin (transparent area), normalized image.

more information and details than the background. The final segmentation mask is obtained by a thresholding operation where the optimal threshold is found by means of the formfactor (Russ 2006) of the resultant binary mask. The output of the segmentation is a binary mask identifying the pixels belonging to the coin.

Normalization: To additionally achieve scale invariance, the segmented coin image is normalized to a standard size of 150x150. This is needed as we use a dense grid of SIFT features with fixed scale (i.e. fixed window size of SIFT descriptor). Additionally, the image is converted to grayscale for SIFT descriptor computation.

Generation of SIFT image: SIFT descriptors are computed at every pixel location to produce the SIFT image. Empirical tests have shown that a window size of 12x12 (i.e. cell size of 3x3) works best to capture the local coin details.

SIFT flow matching: Finally, SIFT images are matched to assess the visual similarity of the coins. As the coins do not necessarily have the same orientation in the images, the matching has to be done in a rotation-invariant way. Therefore, we ignore the small displacement term which is included in the original algorithm (Liu, Yuen und Torralba 2011) for our application in order to allow large displacements between coin images. The SIFT features itself are rotation-invariant, as gradient orientations are represented relative to the dominant orientation.

4. Experiments

For the experiments Roman Republican coins are used which have been kindly provided by the

Museum of Fine Arts, Vienna. Coins from this era have been chosen because a big stock of material is available at the museum. There is a comprehensive standard-reference work by Crawford (Crawford 1974) that defines the various types (classes) of coins. Crawford’s work has 550 distinct reference numbers and we chose a subset of 24 types from 131-102 B.C. for our experiments. For each type, three coin images are available.

For evaluation we use a 3-fold cross validation to test classification performance. The dataset is divided into three subsets, each set containing one image from each class. Three classification runs are executed whereas in each run one subset serves as testset and the remaining two serve as training set. An image from the testset is then matched with all images from the training set. The sum of SIFT flow costs of the two images of a class defines the class-costs, and thus finally the image is classified as the class with minimum class-cost. We test the single performance on the obverse side as well as on the reverse side.

To compare our method to a previously proposed classifier-free method for coin classification, we furthermore apply the standard SIFT matching method proposed by (Kampel und Zaharieva 2008) to the database. In this method, similarity between coins is measured by the number of matched interest points, extracted at Difference-of-Gaussians extrema (Lowe 2004) and described by SIFT. The same evaluation procedure as for the SIFT flow matching is applied.

The overall results are listed in Table 1. Our proposed SIFT flow method clearly outperforms the method of (Kampel und Zaharieva 2008). As the SIFT descriptor itself is potentially noisy on ancient coins due to the lighting problem, the simple

	SIFT Flow Matching	SIFT Matching
Obverse side	63.9%	25.0%
Reverse side	73.6%	33.3%
Total	68.8%	29.2%

Table 1. Classification rates achieved on the coin obverse and reverse sides using the proposed method and the method of (Kampel und Zaharieva 2008).

matching of SIFT interest points is more vulnerable than SIFT flow matching. SIFT flow introduces an additional constraint for a spatially meaningful matching with local variations. Therefore, SIFT flow provides a more robust matching and coin similarity measure.

Another observation from the results is that classification rates are higher on the reverse sides of the coin. This is caused by the typical composition of Roman coins from the investigated period. Customarily, obverse sides show the heads of gods or emperors. In the given evaluation dataset the obverse side of 15 of the 24 classes depict the goddess Roma. Reverse sides depict certain scenes and thus have a higher inter-class variability. However, the given dataset contains coins from an early stage of Roman Republican coinage and therefore this difference is less pronounced since 16 of the 24 classes show chariots, a common motive for this period (Crawford 1974).

One reason for misclassifications is the unsatisfactory handling of lighting variations due to the use of the SIFT descriptor. If different lighting effects appear on the surface of the two coins, SIFT features are too different and thus difficult to match. The second main source of error is the low variability between two coin classes. For instance, the types Cra 282/1 and Cra 282/4 both show chariots and differ from each other only by the inscription. The third main reason for errors of the classification are strong abrasions on the coins. In this case large image regions are not matchable by the SIFT flow method and finally a misclassification is produced.

5. Conclusions

In this paper we presented a classifier-free method for ancient coin classification on 2D images. The method is based on image matching and therefore does not need an offline classifier training, which is difficult to achieve for ancient coins due to the limited number of coins available per class. We use the flexible SIFT flow method which is well qualified for ancient coins as it takes account of the spatial variations that arise from the manual manufacturing of the coins. The experimental results show the potential of the method for ancient coin classification with a classification rate of 73.6% on the reverse sides of 24 coin classes. These results

are achieved on a database including all the main challenges of ancient coin classification: lighting variations, abrasions and high intra-class variability in combination with low inter-class variability. The method can be easily extended to larger databases, as it does not rely on machine learning methods and thus needs no offline training phase.

One drawback of the method is the long computational time for SIFT flow. Currently SIFT flow matching of two images takes about 5 seconds on a standard machine and thus matching a test image with all images in the database is a time-consuming task. As the time to classify a coin is linear to the number of coin classes in the database, a preselection step is a possible solution to speed up the overall process. This can be achieved by using low-level global image features that are fast to compute but are informative enough to initially reject a certain amount of classes. Such features could be, for instance, the coin color (silver, gold etc.) or statistical texture measures.

In general, we see such a hybrid system that combines feature-based classification and image matching as a promising research direction. This way, the best of both worlds can be achieved: feature-based classification like in (Arandjelovic 2010) enables a fast classification once the classifier has been trained, but the feature representation might lose its discriminative power when hundreds of coin classes are contained in the database. In contrast, coin classification based on image matching like the one presented in this paper does not transform a coin image into a less informative feature representation and is therefore able to provide an in-depth analysis of coin similarity. However, this ability comes at the cost of high computational effort.

Acknowledgement

This research has been supported by the Austrian Science Fund (FWF) under the grant TRP-140-N23-2010 (ILAC).

References

Arandjelovic, O. 2010. "Automatic attribution of ancient Roman imperial coins." In *Conference on Computer Vision and Pattern Recognition*, edited by T. Darrell, D.

- Hogg, and D. Jacobs, 1728-1734. Washington: IEEE.
- Bishop, C.M. 2007. *Pattern Recognition and Machine Learning*. New York: Springer.
- Crawford, M.H. 1974. *Roman Republican Coinage*. Cambridge: Cambridge University Press.
- Grierson, P. 1975. *Numismatics*. Oxford: Oxford University Press.
- Huber, R., H. Ramoser, K. Mayer, H. Penz, and M. Rubik. 2005. "Classification of coins using an eigenspace approach." *Pattern Recognition Letters* 26 (1): 61-75.
- Huber-Mörk, R., S. Zambanini, M. Zaharieva, and M. Kampel. 2010. "Identification of ancient coins based on fusion of shape and local features." In *Machine Vision and Applications*, edited by M. Shah, 1-12. Berlin and Heidelberg: Springer.
- Kampel, M., and M. Zaharieva. 2008. "Recognizing ancient coins based on local features." In *4th International Symposium on Advances in Visual Computing*, edited by R. Boyle, B. Parvin, D. Koracin, F. Porikli, J. Peters, J. Klosowski, L. Arns, Y. Ka Chun, T.-M. Rhyne, L. Monroe, 11-22. Berlin and Heidelberg: Springer.
- Liu, Ce, J. Yuen, and A. Torralba. 2011. "SIFT Flow: Dense Correspondence across Scenes and Its Applications." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33 (5): 978-994.
- Lowe, D.G. 2004. "Distinctive Image Features from Scale-Invariant Keypoints." *International Journal of Computer Vision* 60 (2): 91-110.
- Nölle, M., H. Penz, M. Rubik, K. J. Mayer, I. Holländer, and R. Granec. 2003. "Dagobert -- A New Coin Recognition and Sorting System." In *International Conference on Digital Image Computing: Techniques and Applications*, edited by Changmin S., H. Talbot, S. Ourselin, and T. Adriaansen, 329-338. Collingwood: Csir Publishing.
- Reisert, M., O. Ronneberger, and H. Burkhardt. 2006. "An Efficient Gradient Based Registration Technique for Coin Recognition." In *MUSCLE CIS Coin Competition Workshop*, edited by M. Nölle, M. Rubik, and A. Hanbury, 19-31. Accessed February 4, 2013. https://muscle.caa.tuwien.ac.at/coin_proceedings_2006.php.
- Russ, J. C. 2006. *The Image Processing Handbook, 5th edition*. City: CRC Press.
- van der Maaten, Laurens J.P., and P.J. Poon. 2006. "COIN-O-MATIC: A fast system for reliable coin classification." In *MUSCLE CIS Coin Competition Workshop*, edited by M. Nölle, M. Rubik, and A. Hanbury, 7-18. Accessed February 4, 2013. https://muscle.caa.tuwien.ac.at/coin_proceedings_2006.php.
- Zambanini, S., and M. Kampel. 2009. "Robust Automatic Segmentation of Ancient Coins." In *International Conference on Computer Vision Theory and Applications*, edited by Alpesh Ranchordas and Helder Araujo, 273-276. Setubal: INSTICC Press.

Pompeii Revived: Scanning Mission - Insula V 1

Nicolò Dell'Unto

Lund University, Sweden

Matteo Dellepiane, Marco Callieri, Anne-Marie Leander

National Research Council, Pisa, Italy

Stefan Lindgren and Carolina Larsson

Lund University, Sweden

Abstract:

The Swedish Pompeii Project started in 2000 as a fieldwork initiated from the Swedish Institute in Rome. The aim was to record and analyze a full Pompeian city-block, Insula V 1. This paper presents the initial results of one of the actions in the context of this project. In October 2011, two houses were acquired using 3D scanning and 3d-from-photos techniques, and the data was processed to obtain an accurate and complete model. Through the use of Virtual Reality Techniques, it will be possible to visit the Pompeian houses of Casa del Torello and Casa di Cecilio Giocondo understanding the relation between the actual archaeological context and their original outfit. The collected data will be used also to design and test a web-based access system, where the entire dataset will be available for browsing, measurement and data extraction. The features of HTML5, in particular WebGL, will be used to deliver realtime 3D content and interaction.

Keywords:

Field Recording, 3D Scanning, Pompeii

1. Introduction

Three-dimensional digital surveys have become, in the last few years, a standard tool for cultural heritage studies. The advancements in both the scanning devices and the computers used in data processing and visualization have enabled the cultural heritage operators to carry out large-scale surveys and data gathering campaigns.

This technological advancement, however, cannot, by itself be a game-changer in the field of cultural heritage studies.

In order to effectively use all the data produced by these technologies, it is necessary to carefully consider all the aspects of the digitization campaign and the available tools for the processing of the data, and to have a clear idea of the possible uses of the generated dataset.

While in recent literature it is easy to find examples of quite extensive 3D scanning and data gathering campaigns, what is still, in most cases,

Corresponding author: nicolo.dellunto@gmail.com

missing is a plan to effectively manage and use all the gathered data.

2. The Swedish Pompeii Project

During autumn 2000 the Swedish Institute in Rome started a campaign of field documentation of the Insula V, 1 in Pompeii with the aim of recording and analyzing a full Pompeian city-block. From the very beginning, different types of documentation techniques and technologies were tested; this multiple approach of methods was realized in order to provide an accurate, conjoint description of as many different aspects of the ancient buildings as possible. The possibility to develop a project for the documentation of an entire insula allowed the researchers to consider its different constituents (houses), not as separate entities but as part of a total (Fig. 1). This approach underscored the importance of the relation between the different kinds of entities that characterize Pompeian domestic architecture in order to comprehend the development of Pompeian building and social history (Leander Touati 2010; Staub 2009).

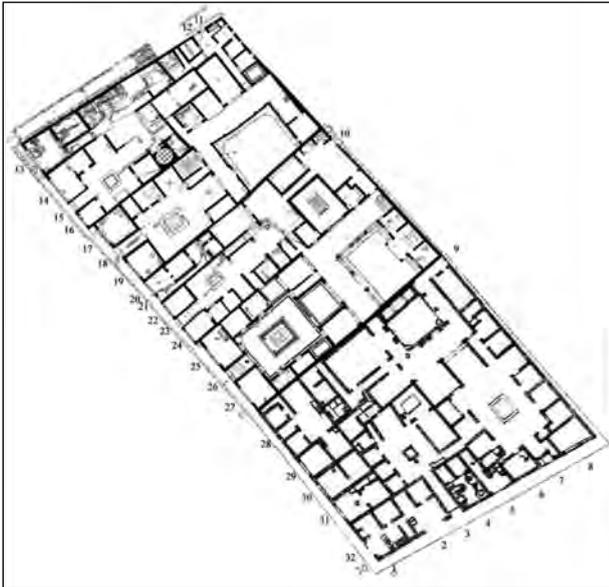


Figure 1. map of the insula V.1 compiled by the Swedish Pompeii Project.

During spring 2011, a collaboration between the National Research Council of Italy (Institute of information Technology and science “A.Faedo”) and Lund University (Institute of Archaeology and Ancient History together with the Humanistic Laboratory), allowed to start testing the use of integrated methods of 3D data recording. The integration of acquisition techniques like 3D Scanning or Dense Stereo reconstruction will increase the knowledge of the relations between the technological infrastructures of the *insula* i.e. water pipes system, illumination etc. and the distribution of the public and private spaces.

The goal of this ongoing sub-project is the documentation of the entire insula in three dimensions. The result of this work will be mainly used as a platform to visualize and highlight with accuracy the research so far developed by the Swedish Pompeii Project team.

To reach such result, a two-steps process have to be carried out: data acquisition of structures and data post-processing. Acquisition technologies, i.e. phase shift laser scanners and 3D from images techniques, and partial post-processing have already been successfully performed. The remaining work consists of an intense work of virtual interpretation of the buildings before the eruption that destroyed

Pompeii. In particular, the documentation produced in the last two centuries will be used to reconstruct part of the architecture and the decoration, today not anymore available *in situ*.

Aim of this work is assessing if, and eventually how, the use of 3D models affects the interpretation of the Pompeian domestic architecture.

During the progression of the archaeological data collecting in the field, the Swedish Pompeii project developed a digital research platform characterized by its unique transparency. The data elaborated during the field campaigns are published online. The archaeological results are organized into a website (www.pompejiprojektet.se/insula.php) that allows rapid access to different kinds of information, advancing from general information towards detailed data, presented both in text, graphic and photograph. It includes high-resolution ortho-mosaics of the ancient structures, plans, elevation and reports (Fig. 1).

For the reason that the maximum transparency should include also the 3D-data we experimented the use of WebGL, to visualize the gathered three-dimensional data directly through web browsers, in order to connect this new experimental approach with the classic documentation disseminated during these years through the Internet.

The development of such web access for the visualization of the 3D data would provide the opportunity to anyone interested in studying the insula V,1 to access directly the information elaborated by the project team.

3. Previous Work

Three-dimensional digital surveys have become, in the last few years, an important part of the documentation and communication process of archaeological sites; the examples presented in literature show how acquisition techniques and methods have been successfully used so far in similar case studies. An overview of all the most successful 3D scanning campaigns goes well beyond the scope of this paper, but we can cite two examples where 3D Scanning was used as a valuable support to provide data for reconstruction projects.

The first one, *Rome reborn*, dealt with the case where the original structures have been partially destroyed (Dylla et al. 2009). The second one, *The Parthenon*, took into account a monument which fragments are now scattered in different parts of the world (Stumpf et al. 2003).

Another example is provided by the Flaminia project (<http://www.vhlab.itabc.cnr.it/flaminia/>), which describes the development of a workflow for the three dimensional documentation and the communication of several archaeological sites, through the combination of different acquisition techniques such as Laser scanner and photogrammetry.

The result of such projects can be easily employed to build visual infrastructures of cultural communication such as Virtual museums or museum installations. Among a number of possible examples, The Virtual Museum of the Western Han Dynasty (Forte et al. 2010) uses several acquired structures and objects to build a virtual museum.

Pompeii and, more in general, the entire area affected by the eruption of 79 AD has been subject of study and mapping way before the actual birth of modern archeology. A great deal of maps and technical surveys are available for this area, with quite various levels of accuracy and detail. Recently, many different survey actions have been carried out using modern, dense-sampling devices like laser 3D scanners. This has been done, for example, in Herculaneum (Brizzi et al. 2006) to increase spatial resolution in surveys and provide effective geometric data for conservation purposes. In other cases, like (Hori et al. 2007), the aim was to validate (and possibly correct) older surveys. Other surveys covered specific areas of the Pompeii area (Poehler and Ellis, 2012) (Balzani et al., 2004) for specific conservative actions.

4. Data Acquisition

The acquisition campaign took place during the first week of October 2011; a team of six people (three from Pisa –CNR- and three from Sweden –Institute of Archaeology and Ancient History-) started the acquisition campaign of the insula through the employment of two different phase shift laser scanners: a Faro Focus 3D and a Faro



Figure 2. This image shows the extension of the area acquired at the end of the campaign and the extension of the area expected before the beginning of the campaign.

PHOTON 120. Despite these two instruments have similar characteristics in terms of acquired data quality, their different sizes and weight affect their use in different situations; Faro focus 3D is more light and easy to use in fragile and unstable areas, Faro PHOTON 120 is more sturdy and stable, thus more suited for the acquisition of higher structures from scaffoldings. Moreover, the employment of two scanners allowed also a more efficient data acquisition process, since two teams could work at the same time in different areas of the insula.

Before starting the campaign, several acquisition strategies were discussed in order to optimize the process in the limited amount of time available, and to be able to highlight parts of the archaeological features considered more challenging for the instruments. As a consequence, markers were excluded from the process due to the complexity of the houses geometrical features, and a manual alignment of the point clouds was performed instead.

We decided to start our work from *Casa di Cecilio Giocondo* and document its structures. Once completed the acquisition of this part, we planned to start acquiring a small portion of *Casa del Torello di Bronzo (ninfeum)*. In particular, the structures that we experienced being troublesome in terms of acquisition were the *cubicula* of *Casa di Cecilio Giocondo* and some of the corridors that connect the public with the private areas of the house; the small size of these rooms was almost at the limit of the acquisition range of the instruments and we were not sure if it would have been possible to use the laser to acquire such environments. At the end of the week, the campaign resulted extremely successful: after only three days we were able to acquire the

two houses entirely (*Casa del Torello di Bronzo* and *Casa di Cecilio Giocondo*) plus the streets that surround the structures (Fig. 2)

The final laser scanner dataset covered around 1330 (620 + 710) square meters, acquired using 110 scan positions, each one with a 360 degree coverage, with a quite dense sampling rate (1cm at 10 metres). The size of this three-dimensional dataset is more than 200 GB of raw data. The quality of the acquired data is visually very good. Having a precise measurement of the precision and accuracy of the sampling (beside accepting the declared specifications of the devices) would require a ground truth geometry, which is unavailable. On the other hand, since the scanning has been carried out using two different devices of the same class, by comparing the data coming from the two devices in overlapping areas, it is possible to evaluate the coherence of the two samplings (like if we were using the data from scanner #1 as the ground truth to evaluate scanner #2 and vice-versa): the value of this incoherence will be an upper bound of the sampling error. In this dataset, we obtained very low values: 90% of the measured overlapped areas had a disparity below 2.5mm and 50% below 1.5mm. This low error level may be due to the nature of surfaces (mostly non-reflective and optically cooperative) and the closeness of the sampled surfaces to the scanner.

Despite the efficiency of the Laser scanner in acquiring the main structures of the insula, several important features such as the water pipe system of the *Casa del Torello di Bronzo*, turned to be extremely complicated to document with the Laser scanner. In fact, the location of this structure, which was used to regulate the distribution of the water inside the house, and the complexity and size of its geometry, would have required a quite different kind of scanner (possibly, a triangulation one).

Therefore, we decided instead to use Computer Vision techniques to generate resolute 3D models to align afterwards with the laser scanner model of the house (Fig. 3a-b).

This technique combines algorithms of structure from motion and dense stereo matching in order to build a 3D model of a scene starting from an uncalibrated set of images.

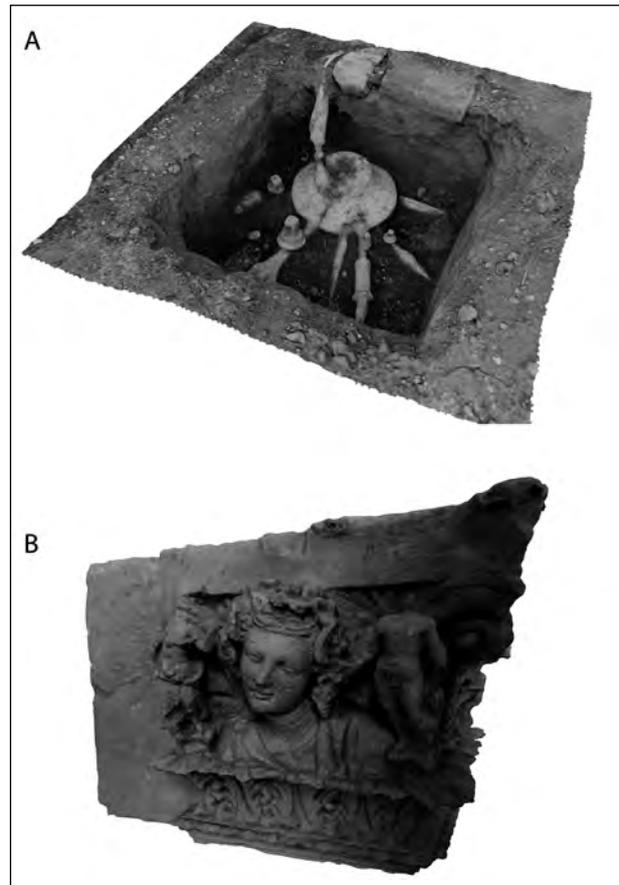


Figure 3. 3D models created using Computer Vision techniques: A water pipe device of the fountain of the *Casa del Torello Di Bronzo*, B architectural decoration found during the archaeological investigation of the *Casa del Torello Di Bronzo*.

This technique was also tested to acquire architectural materials found inside the houses and stored in a different area of Pompeii. Due to their fragility and the considerable size, architectural decorations are often very difficult to transport and therefore not so easily acquired. In fact, the narrow space of the storage rooms does not allow the use of any kind of Laser scanner or calibrated camera. On the contrary, the possibility to use uncalibrated pictures to get 3D information, allowed the elaboration of a resolute three-dimensional model to be transferred inside the house and simulate the relationship between environment and decorative elements characterizing the structures.

We have, in a previously published work (Callieri et al. 2011), carried out tests on 3D-from-

images technologies, assessing their practical use on the field and the processing of the recovered data for documentation purposes. The data in this project has been gathered and processed following the methodology defined there, and obtaining similar quality in the resulting geometries. Please refer to this work to have more details.

5. Data Processing

Once the raw data captured by the scanner were available, the processing of the data was carried out inside MeshLab. MeshLab (Cignoni et al. 2008) is an open source tool for the visualization and processing of 3D models. It is oriented to the management and processing of large, unstructured triangular meshes and point clouds, and it provides a set of tools for measuring, checking, cleaning, healing, inspecting, rendering and converting 3D meshes. MeshLab is freely available, distributed under the GPL licensing scheme and it is available for all the major platforms (Windows, MacOS, Linux).

The first step of data processing is the alignment. This step brings all the captured data in the same reference space.

Normally, TOF data are aligned using markers placed on the scene. However, in this case, the use of markers was quite complex, due to the large amount of different, interconnected rooms. This would have required a large number of markers, and additional time to manage them. On the other hand, the amount of walls and the same connected topology make easy the alignment based on geometric redundancy between scans.

MeshLab provides a Geometric Alignment filter based on the well-known ICP algorithm (Rusinkiewicz and Levoy, 2001), enhanced with all the optimization and tricks available in literature. This alignment filter works on triangulated surfaces (like in many analogous tools) but also on raw point clouds. Being able to work directly on the point dataset is a great advantage in terms of time and required memory, since there is no need to generate a triangulation.

The alignment is a two-step process: initially, the user picks corresponding points between scans, placing the scans in an approximate position; then,

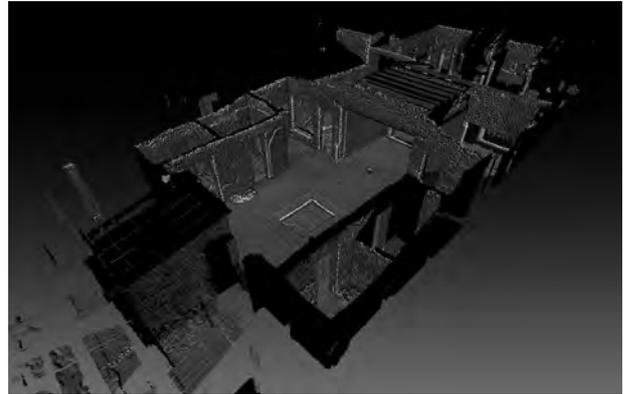


Figure 4. Some of the range scans of the House of Cecilio Giocondo, aligned inside MeshLab (each color is a different scan).

the system computes a precise alignment using the whole overlapping area between the scans.

Even if alignment, due to the manual input required, is still the most time-consuming part of processing (along with range map cleaning), after some practice, a user can obtain optimal results with a few hours of manual work. The final alignment accuracy is on par to the one obtainable using markers. This alignment method provides data on the residual errors of the single cloud matching (local error of the alignment of each cloud with every other overlapping), and on the final residual error (after the global optimization and bundle adjustment). The first value varied in the dataset between 1mm and 5mm, while the final residual was below 2mm.

After all the scans have been aligned, the resulting point cloud can already be used as a metric documentation, for taking measurements, or for visualization purposes (Fig. 4).

However, to fully exploit the visual potentiality of the data, it is necessary to transform this point-based representation into a triangulated surface. This will provide a much more

The process of computing a triangulated surface from a series of individual scans is called merging. This is a completely automatic process, where the user has only to decide the final geometric resolution of the output model.

MeshLab does implements various algorithms for the merging step, able to accurately generate

triangulated surfaces starting from pointclouds or triangulated range scans.

Given the extent of the dataset, it is impractical to think of creating a single triangulated model for the entire area. This will require an extremely long computation time, a lot of memory and produce a model so complex to be unusable. Moreover, given that each part of the dataset has specific coverage and geometrical characteristics, using a single set of parameters for the entire extent would reduce the amount of usable detail.

On the other hand, the covered areas are divided in rooms, and this geometrical subdivision is also a logical one, since it reflects the way the houses have been designed and built, and the way they will be studied and measured.

It made sense, then, to reconstruct independently the various rooms, making the merging step much easier, but still producing a coherent geometry for the entire dataset (thanks to the properties of regularity of the merging algorithms) (Fig. 5).

The final step of the processing is the color mapping. The produced triangulated model does faithfully represent the geometry of the building, but not its appearance. For this reason, a series of photos have been acquired in order to provide an accurate description of the appearance of all the walls of the houses.

To perform color mapping, it is necessary first to align the photos onto the 3D model: this is obtained by estimating the camera parameters associated to each image.

These parameters describe the position and orientation (extrinsic parameters) and internals of the camera, like sensor size, lens distortion and focal length (intrinsic parameters) at the moment of the shot. By obtaining these parameters it is possible to reconstruct the perspective projection that created the photo. This opens up two possibilities: being able to see the 3D scene through the same camera that took the shot (thus, exploring the photographic dataset spatially), and project back the color information onto the 3D model (to generate color mapping).

This alignment has been done inside MeshLab using a user-friendly approach, based on Mutual Information (Corsini et al. 2009). This is based on the calculation of a statistical measure of correlation between the image and a rendering of the model, and has a very simple interface, making the photographic alignment easy for the user.

The photos aligned in this way may be spatially-explored directly in the 3D space, instead of browsing a folder on the disk, looking at the photos like see-through transparencies suspended in space or projected one by one onto the 3D surface.

An interesting possibility offered by this strategy is to use also unconventional images, like photos with annotations, historical photos, hand drawings or sketches or even near-visible lighting photos (ultraviolet, multispectral, infrared, thermography, etc.). This possibility of spatial exploration of the georeferenced photographic set is a powerful tool to effectively browse a collection, and has multiple uses in the documentation and study of this dataset.

However, since our aim was to obtain photorealistic 3D models, we proceeded to map the color from the images onto the triangulated surface. To this aim, MeshLab offers different color mapping tools to better cope with the different needs of the various datasets. By using the color data from the calibrated images, it is possible to generate detailed, artifact free per-vertex color encoding, or create a texture map.

We employed both solutions, in order to produce high resolution 3D models with per-vertex color for study, documentation and measurement purposes, and lower resolution models with texture mapping, for real-time visualization (Fig. 6).

This final step is still ongoing, given the amount of images to be mapped, and the extent of the dataset.

6. Virtual Reality

During the autumn 2014 an exhibition in Stockholm about the insula V,1 will be organized. In that occasion a Virtual Reality system, developed using the results of this work, will be prepared in

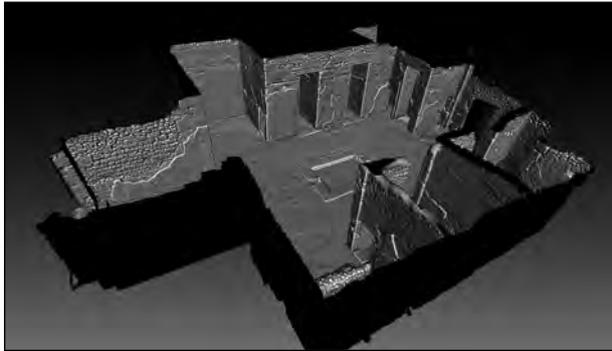


Figure 5. One of the rooms (the atrium of the House of the Torello di Bronzo), reconstructed as a triangulated surface.

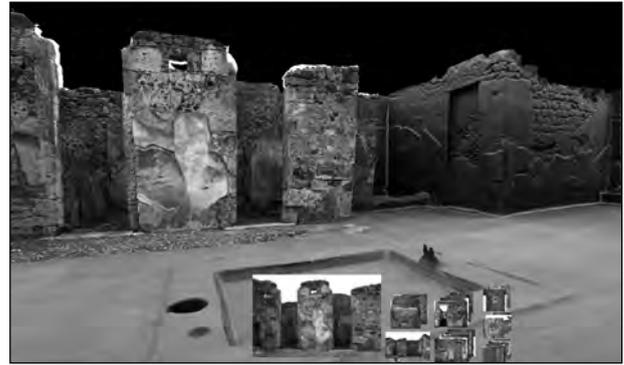


Figure 6. The aligned photos browsed spatially, over the geometry of one of the rooms.

order to show to the public the investigated area and the results of the virtual interpretation of the ancient structures. Despite the data have not been entirely processed and more than half insula has not yet been acquired, we are already discussing the possibility to build a virtual reality systems to explore -through the use of devices that allowed natural interaction such as Microsoft Kinect-the insula and having an easy understanding of the connection between the interpretation of the environments and the archaeological research developed during these years.

7. Online Publishing

One of the focal points of this project was the objective of making the entire dataset available over the net.

The web platform has acquired through the years the ability to efficiently incorporate and deliver many different kinds of digital data such as still images, videos and sound. With respect to these additions, the management of 3D content through the web still presents many problems.

HTML 5 introduced, as a new part of the standard, the WebGL component. Using this component, it is possible to display 3D content directly inside a webpage, without the use of browser plugin.

However, WebGL alone is not enough to answer the needs of people interested in web 3D visualization; following the design philosophy of OpenGL, WebGL is a very low-level API. It

is therefore necessary, to ease the use of this technology, to introduce a library able to wrap the most low level function, while giving the user the ability to dive into implementation details, when needed.

To do so, we chose SpiderGL (Di Benedetto et al. 2010), which is JavaScript library designed to provide an easily usable but powerful wrapping to the lower-level WebGL functions.

The idea of this library is to provide a complete wrapping layer to WebGL that, while hiding the details through higher-level functions, allows full access to the native API. To ease the creation of graphical applications, SpiderGL provides a series of classes and functions that cover the various aspects and levels of implementation of a modern computer graphic application. Using this technology, it has been possible to create a visualization scheme for the web platform, able to display the 3D models of the various rooms in a web page, with a simple user interface for the 3D navigation.

This visualization scheme uses, thanks to the functionalities of SpiderGL, a multiresolution model. Multiresolution is a way to encode the three-dimensional data in such a way that, for each part of the object, the geometry is available at different level of details. During rendering, for each part of the model, it is chosen the optimal resolution, the better trade of between the visible detail and the available resources, to ensure. This makes possible to show in real time very complex 3D models, composed by millions of triangles.

Another interesting feature is that a multiresolution model, being composed by small chunks of data, it is extremely well suited for network transmission. Exploiting this property, the visualization code is able to effectively stream the data from the server, resulting in a very short startup time (the model is almost immediately ready for the user) and a reduced network load (only the used data is transmitted).

The idea is to start from a standard navigation paradigm, based on plans and prospects, which exploits the natural room-based environment. This navigation method is already used in the current version of the project website. With it, the user can access datasheets of the individual rooms, making possible to visit, room by room, the entire dataset. Adding to this hierarchical visit, at the level of the rooms, a link to explore the corresponding 3D model of the room will make the integration of 3D data easy and accessible.

While the visualization and exploration part is already working, we are now adding more advanced features. The prototype of the visualization tool will be made accessible online from the website of the project (<http://www.pompejiprojektet.se/insula.php>) before the end of the year.

The idea, for the next step, will be to enrich the interactive visualization by adding the possibility to take measurement, pick points and save the current view as a bookmark or as an image. This can be done easily using JavaScript, and the early tests of this capabilities have been promising.

A more complex option, which will require server-side support, is the creation of high-resolution snapshots using the complete dataset (and not only the current room), or more sophisticated documentation like cut-through sections or maps views. It is easy, in the Javascript visualization scheme, to define the parameters of such documentation using a simple interface. However, it will then be necessary to forward the request to the server, which will compute the required data on a separate process and, later on, delivered to the users.

This tool, once complete, will enhance the possibilities for the analysis of the documentation

of the archaeological remains, proposing a starting point for additional actions like annotation, complex interaction, collaborative work.

8. Conclusions

In this paper, we presented the first results of the scanning campaign of the Insula V 1. We described the aim of the campaign, and the technical choices taken during the planning, execution and data processing.

More than the description of specific technical details and a series of figures describing the dataset, we do believe the main goal of this presentation is to justify technical choices in the framework of an articulated documentation and study project.

We showed how, with a clear idea of the needs of the project and the aim of effectively use all the collected data, we set-up a complete software pipeline for the processing and visualization of this large three-dimensional dataset.

We believe that the integration on the web platform is an important and almost mandatory feature for a modern documentation project, to ensure ample access and easy data interaction.

We plan for this next fall a new acquisition campaign that, if it proves to be as fast as the one here described, could succeed at completing the sample of the entire Insula.

The development of the web-enabled visualization component will continue and, hopefully, the new measuring features will be ready when the final, complete insula dataset will be available for the scholars.

References

- Balzani M., N. Santopuoli, A. Grieco, and N. Zaltron. 2004. "Laser Scanner 3D Survey in Archaeological Field: the Forum of Pompeii." In *International Conference on Remote Sensing Archaeology, Beijing, October 18–21*, 169–175. Beijing: Joint Laboratory of Remote Sensing Archaeology.
- Brizzi M., S. Court, A. d'Andrea, A. Lastra, and D. Sepio. 2006. "3D Laser Scanning as a Tool for Conservation: The

- Experiences of the Herculaneum Conservation Project.” In *Proceedings of the 7th International Symposium on Virtual Reality, Archaeology and Cultural Heritage VAST*, edited by M. Ioannides, D.B. Arnold, F. Niccolucci, K. Mania, 111-117. Nicosia: Eurographics Association.
- Callieri, M., N. Dell Unto, M. Dellepiane, R. Scopigno, B. Soderberg, and L. Larsson. 2011. “Documentation and Interpretation of an Archeological Excavation: an Experience with Dense Stereo reconstruction Tools.” In *VAST The 11th International Symposium on Virtual Reality Archaeology and Cultural Heritage*, edited by F. Niccolucci, M. Dellepiane, S. Peña Serna, H.E. Rushmeier, L.J. Van Gool,, 33-40. Prato: Eurographics. <http://vcg.isti.cnr.it/Publications/2011/CDDSSL11>.
- Cignoni, P., M. Callieri, M. Corsini, M. Dellepiane, F. Ganovelli, and G. Ranzuglia. 2008. “MeshLab: an Open-Source Mesh Processing Tool.” In *Proceedings of Eurographics Italian Chapter Conference*, 129-136. <http://c2.etf.unsa.ba/file.php/71/ZavrnsniRadovi20092010/SelmaRizvic/tema3.doc>.
- Corsini, M., M. Dellepiane, F. Ponchio, and R. Scopigno. 2009. “Image-to-Geometry Registration: a Mutual Information Method exploiting Illumination-related Geometric Properties.” *Computer Graphics Forum* 7 (28): 1755-1764 .
- Di Benedetto, M., F. Ponchio, and F. Ganovelli. 2010. “SpiderGL : A JavaScript 3D Graphics Library for Next-Generation WWW.” *Library* 1 212: 165-174. <http://portal.acm.org/citation.cfm?id=1836075>.
- Dell'Unto, N., F. Galeazzi, and M. Di Ioia. 2006. “Via Flaminia project: relief and post processing data techniques.” In *Proceedings of the “From space to place” Conference, Rome, Italy, 4-7 December*, edited by S. Campana and M. Forte, 52-60. Oxford: Archaeopress.
- Dylla, K., P. Müller, A. Ulmer, S. Haegler, and B. Fischer. 2009. “Rome Reborn 2.0: A Framework for Virtual City Reconstruction Using Procedural Modeling Techniques.” In *Proceedings of Computer Applications and Quantitative Methods in Archaeology*, edited by B. Frischer, J. Webb Crawford and D. Koller, 62-66. Oxford: Archaeopress.
- Forte M., N. Dell'Unto, P. Di Giuseppantonio Di Franco, F. Galeazzi, C. Liuzza and S. Pescarin. 2010. “The Virtual Museum of the Western Han Dynasty: 3D Documentation and Interpretation.” In *Space, Time, Place, Third International Conference on Remote Sensing in Archaeology, 17th-21st August 2009, Tiruchirappalli, Tamil Nadu, India*, edited by S. Campana, M. Forte and C. Liuzza. Oxford: Archaeopress.
- Hori Y., O. Ajioka, and A. Hanghai. 2007. “Laser Scanning in Pompeian City wall: A comparative study of accuracy of the drawings from 1930s to 1940s.” In *3D-Arch'2007 3D Virtual Reconstruction and Visualization of Complex Architectures*, edited by F. Remondino, S. El-Hakim. Zurich: International Society for Photogrammetry and Remote Sensing.
- Leander Touati, A.-M. 2010. “Water, well-being and social complexity in Insula V 1. A Pompeian city block revisited.” *Opuscula* 3: 105-161.
- Poehler E., and S. Ellis. 2012. “The 2011 Season of the Pompeii Quadriporticus Project: The Southern and Northern Sides.” *Fasti On Line Documents & Research* 249. Accessed January 2013. <http://www.umass.edu/classics/PQP.htm>.
- Rusinkiewicz, S., and M. Levoy. 2001. “Efficient variants of the ICP algorithm.” In *Proceedings Third International Conference on 3D Digital Imaging and Modeling*, 145-152. Quebec: IEEE Computer Society 2001. <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=924423>.
- Staub, T. 2009. “Decorative effects and room functions. Evidence of thresholds studied in the residential quarters of Insula V 1, Pompeii.” *Opuscula* 2: 205-218.
- Stumpf J., C. Tchou, N. Yun, P. Martinez, T. Hawkins, A. Jones, B. Emerson, and P.E. Debevec. 2003. “Digital Reunification of the Parthenon and its Sculptures.” *VAST*: 41-50.

A Metrology Tracking System Used as a 3D Drawing Tool for Faster Excavation Recording

Maarten Smeets, Wouter Yperman
Studiebureau Archeologie bvba, Belgium

Geoff Avern
University of Southampton, United Kingdom and Nikon Project

Abstract:

The iSpace metrology-grade tracking system from Nikon Metrology can also be used as a digital 3D drawing tool. A new archaeology-specific version iSpace was released in April 2011 at CAA-Beijing, soon after which, Studiebureau Archeologie, Belgium, became the first archaeology company to use it for recording excavations. A brief explanation of how iSpace works is followed by a report on the combined experiences of the authors with the system. Specifically we will present the results of 2 sets of trials which pitted iSpace against other forms of drawing of excavations, namely hand drawing and drawing with Robotic Total Station. A more general discussion will follow on various aspects of the iSpace system and its potential impact on commercial archaeology, based on experiences of using the system in over a dozen excavations.

Keywords:

Digital, Drawing, Commercial

1. Introduction

In the very competitive field of commercial archaeology excavations are conducted within pressing time frames. Any new method, technique or technology which significantly reduces the total time of a project should have a considerable impact upon commercial archaeology.

For over a decade a number of researchers have investigated ways to improve the speed and quality of excavation recording using various 3D data acquisition methods; such as photogrammetry (Barcelo, et al., 2002), orthorectification of photographs (Reali & Zoppi, 2001), tracing from photomosaics (Avern, 2001a), using total station points for drawings (Schaich, 2002), 3D Modeling (Avern, 2001b) and laser scanning (Doneus & Neubauer, 2004). For various reasons none of these methods have become mainstream techniques. The technology considered in this paper differs from these in that it does not produce 3D surface models but, rather, 3D vector drawings, that is, the typical form of recording currently employed on most excavations, with an added dimension.

One of us (GA) recently investigated many tracking systems for their potential as a 3D drawing tool on archaeological excavations. The premise for this use is to track the position of a hand-held probe which is simply used as a “digital pencil” to trace the borders of archaeological features. Potentially, drawings could be made much faster, while delivering more accurate drawings than could be achieved by hand or total station, and significantly reducing post-processing times.

The review of available tracking systems revealed that, almost without exception, they are designed for indoor use and are, consequently, impractical or impossible to use outdoors on an excavation, either because the system will not work in sunlight or because its implementation relies on some impractical physical arrangement, e.g. ceiling-mounted sensors. Virtually the sole exception is the iSpace tracking system from Nikon Metrology (the metrology division of the Nikon Corporation). A collaboration between one of the authors (GA) and Nikon Metrology, first at field trials at Mont Beuvray (ancient Bibracte), France, in 2010 (Avern & Franssens, 2012) and, subsequently, in the development of an archaeology-specific version of the system, saw “Archaeology iSpace” released in April 2011 at the Beijing CAA conference.

Corresponding author: geoff.avern@gmail.com

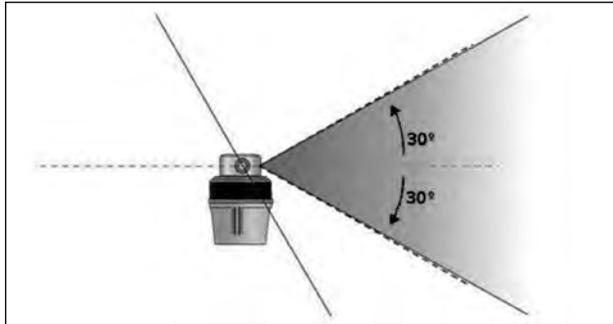


Figure 1. The two laser fans are also rotated by 30° from vertical in opposite directions. The time difference between detection of each fan is a function of the vertical angle from the “horizontal” projection axis to the sensor.

The trials at Mont Beuvray (Avern & Franssens, *ibid.*) demonstrated that drawing with iSpace was up to 10x faster than drawing by hand. Other important advantages were:

- With up to 40 points collected per second, each archaeological feature was drawn by at least an order of magnitude more points than could be practically achieved by hand, resulting in more accurate shapes of the drawn features,
- With tracking accuracy less than 1mm, the positional accuracy of the resulting drawings was far greater than could be achieved by hand,
- All drawings were inherently geo-referenced since the lines of the drawing were constructed from many (hundreds or thousands) of geo-referenced 3D coordinates,
- The system was simple and intuitive to use. This suggested that an “archaeology” version of iSpace would not require a specialist operator and so keep recording in the hands of the excavator.

Soon after its release, Studiebureau Archeologie bvba, Belgium, became the first archaeology company to use Archeology iSpace on commercial excavations and are, now, clearly the most experienced users of this new technology (Smeets and Avern, in press). In the current paper the authors describe the working of the system (GA) and report on their experiences of using this new recording tool in the area of commercial archaeology (MS, WY).

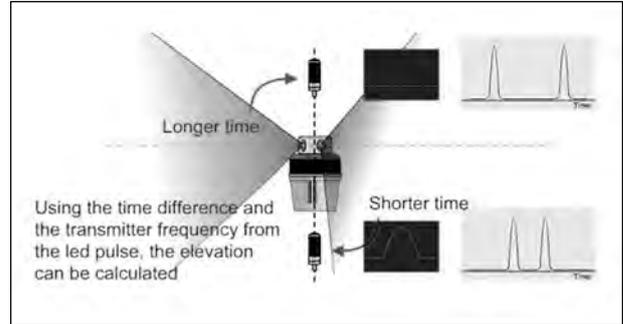


Figure 2. The time between the strobe pulse (left on the graph) and the mean of the signals from the two rotating fans (to the right of the graph) is a function of the position of the sensor relative to the transmitter and gives the azimuth.

2. The Nikon Archaeology iSpace System

The Archaeology iSpace is a large-volume, outside-in tracking system. Four tripod-mounted transmitters are set up around the area to be recorded to create a “tracking volume”. They can cover an area up to 40mx40m. Two probes, which carry arrays of sensors, are the hand-held “digital pencils” which will record up to 40 coordinates per second with sub-millimetre accuracy.

2.1 Timing Signals and Calculation of Coordinates

Each transmitter emits two infrared signals, viz. a pair of rotating infrared “fans” and a strobed timing pulse. The fans are tilted 30° from vertical, one clockwise, the other anticlockwise (Fig.1).

Given the fixed rotational velocity, the measured time difference between detection of the two rotating fans gives the vertical angle (elevation) between the transmitter and the receiver. The horizontal angle (azimuth) is calculated from the time difference between the 40Hz strobed timing pulse (visible for 360° around the transmitter) and the mean of the two timing fan signals as they rotate (Fig.2).

Together, these two angles provide a ray from the transmitter, along which the sensor lies. The position of the sensor will be at the intersection of two or more of the four calculated rays (Fig.3).

While a result is obtained from the intersection of only two rays, a third and fourth ray allows, first,

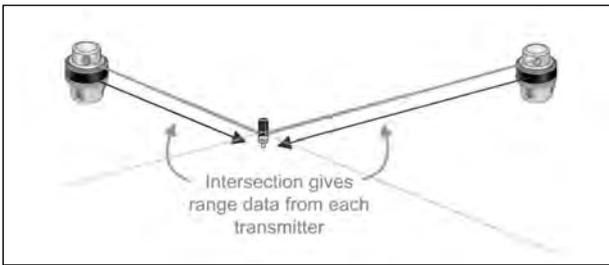


Figure 3. Sensors lie along a calculated ray from each transmitter. The position of the sensor will be at the intersection of these rays.



Figure 4. The iJavelin is a 1m configuration of 4 sensors, a trigger and radio communications, under which is fitted a 1m-long probe tip.

an estimation of accuracy and also accommodates any temporary loss of line-of-sight between a sensor and any of the transmitters. The archaeology version of iSpace manages sub-millimetre accuracy, far greater accuracy than rtkGPS and standard total station theodolites.

2.2 Sensors

The sensors are small, “lipstick-sized” cylinders connected to radio frequency transmitters. They collect timing data from the transmitters and transmit them to a tablet workstation which



Figure 5. The iProbe is a smaller configuration of 4 sensors with trigger and radio communications, under which a 40cm-long probe tip is fitted (Photograph: Antoine Mailler, BIBRACTE EPCC).

calculates of their positions. The position of each sensor in use is calculated up to 40 times per second.

2.3 Hand-held probes

The iSpace system comes with two hand-held probes, both fitted with 4 sensors in a fixed and calibrated configuration, which allows the system to calculate both the position of the probe tip and the orientation of the probe (measurement with 6 Degrees of Freedom). Accordingly, the probes do not need to be held vertically and so eliminate the major source of error encountered when using a total station with staff-mounted prism. The probes are also fitted with buttons to trigger recording episodes. Recording can done in two modes; single point mode (for surveying, taking spot heights, find localities, etc) or in continuous mode (for tracing around outlines of contexts or features).

The two probes differ in their size; the 2 metre “iJavelin”, which is convenient for taking single points (spot heights, finds, point surveys) and for working in holes or pits (Fig.4), and the 40cm “iProbe”, useful for more detailed and precise drawings (contexts, features, sections, etc.) (Fig.5).

2.4 Drawing software

Archaeology iSpace currently includes an iPhone/iPad/iPodTouch application called MobiGage for Archaeology from Titansan Engineering Inc. which allows data to be attributed as it is collected and the result exported as an

	iSpace (minutes)	Robotic Total Station (minutes)
Setup	13	5
Other	3 (calibration)	10 (geo-referencing)
Recording times	14	27
TOTAL	30	42

Table 1. Results of first test of recording times between Nikon iSpace and Robotic Total Station.

AutoCAD .dxf file. Clearly, while iSpace saves a great deal of time in acquiring the drawing data, the output of attributed .dxf files represents further substantial time savings compared to the digitising of hand drawings and manual adding of attributes in post-excavation processing.

3. Nikon iSpace versus Conventional Recording Methods

3.1 Manual drawing

Before Archaeology iSpace, Studiebureau Archeologie used digital plans obtained either by digitising hand-drawn plans or by drawing with a total station. Clearly, registering plans on paper takes a long time in the field (usually by at least two people) and even more time to make a digital, geo-referenced plan which can be used in publications. For example, in an excavation at the Fochplein in Leuven, registration of all hand-drawn traces (we use the term “traces” for evidence of ditches, postholes, pits, etc) and the walls took about one third of the total work (28 days) while the subsequent digitalization of the drawings took about 40 days.

3.2 Robotic Total Station (RTS)

The use of a robotic total station is usually much quicker than manual drawing or using a normal Total Station. This means of drawing required only one person, saving further time and obviating any communication errors between an operator of a traditional Total Station and the person with a prism staff. At first we deployed RTS drawing in trenches, later on entire excavations, although often excluding urban archaeology since it involved the recording of a lot of brick and stone walls, which were still drawn by hand. Despite some manual drawing, considerable time was gained. We next implemented a data coding system which allowed us to automatically generate lines between points, and to insert the trace number in the

drawing. This resulted in further gains, this time in post processing.

The disadvantages are that the accuracy of the system is affected by the experience of the topographer and the care taken to keep the prism staff vertical, that it is quite slow if very precise drawings of complex traces are required, and that there still remains some work in the subsequent editing of the plans. At the site of Kontich-Babelkroonbeek, the registration of 2500 square meters took 3 days (from a total of 21 working days) and the editing another 2 full days.

3.3 Robotic Total Station versus Nikon iSpace

The iSpace came with the promise that it would be even faster. Our general impression was that measuring with the iSpace is faster than measuring with the RTS but we decided to investigate the differences at a large assessment excavation at the rural outskirts of Maasmechelen, Belgium.

Three trials were performed comparing recording by both RTS and iSpace. The first test was in an area of about 10m x 65m containing 10 traces (for drawing) and 12 loose finds (for localising). The traces were postholes and small pits. This first test was done by the same operator having experience in both the RTS and the iSpace. Because both systems have a different setup and method of operation, each part was recorded separately. The results are given in Table 1.

Although the iSpace is faster, some explanation will point out the real time gaining factor. First, it is fair to point out that our RTS map is geo-referenced and the iSpace map is not. If the RTS map was not geo-referenced set up times would have been the same since there would be no need to walk 10min between the georeferenced points. Measuring the loose finds and the limit of excavation was equally fast with both systems, as they both consist of

capturing one or a few points from across the entire area. The real time gaining factor is, however, the drawing of the traces. The iSpace can use streaming measuring to acquire dozens of points for each trace while it is only practical to use the RTS for obtaining single points in the recording of the trace. To obtain a reasonably accurate trace, 10 points or more were taken for each trace with the RTS, taking a minute or more per trace. With 10 traces the drawing was 10min slower. Obviously the more traces, or the larger the trace, the more time would be gained with iSpace. At Maasmechelen we were also faced with recording the foundations of a large stone wall, using iSpace to draw every stone. We estimate that by using iSpace instead of the RTS on just these foundations saved more than two weeks.

In Table 1 we see the difference in set up times for RTS and iSpace. In the second trial we chose to record the traces in two adjacent assessment excavation strips which would require repositioning the iSpace (as opposed to the RTS which, with its much larger range, did not require moving). In this trial the iSpace was operated by two persons since we found it easier to record when one is drawing while the other is inputting settings/data/etc. in the iPod touch and looking for the next trace to measure. The number of features to be recorded were 59 traces and 10 loose finds. The results are given in Table 2.

Clearly, the more traces, the more time gained. In trenches where more than 50 traces (including ditches and large pits) were recorded in a single drawing, more than an hour of time was saved per drawing. We estimate that if faced with the task of recording 500 or more irregular stones (a stone foundation, paved path, etc.) using iSpace instead of RTS would save a full day.

However, the long range of the RTS makes it valuable where, for instance, long, narrow assessment trenches are opened over a large area or when surveying the topography of a site. At

large rural excavations it is useful for setting out geo-reference points for use with iSpace, and for recording when there are only a few finds or isolated traces.

In dense urban areas where the range of the iSpace is often large enough to cover the entire excavation, the RTS becomes redundant once the fixed points have been geo-referenced. And when a large number of traces (in our experience, more than ten) or stone/brick structures are encountered, iSpace is clearly the tool of choice.

4. Commercial Archaeology Using the iSpace Tracking System

4.1 Government requirements

Commercial archaeology in Flanders (Belgium) is relatively new, having begun in 2005. Although a legal framework was available since the Decree of the 30th of June 1993 regarding the protection of the archaeological patrimonial, the Flemish government only started in 2005 to set the conditions for each archaeological excavation done in advance of commercial or community developments. Importantly, these conditions include the methodology to be used, the number of excavators and the expected duration of the excavation.

Clearly, “time is money” for construction companies and there is considerable opportunity for commercial archaeology companies who are able to shorten the length of excavations. All parties should, therefore, be interested in techniques that make excavations shorter.

Studiebureau Archeologie bvba had the opportunity to renew contacts with Nikon Metrology while tendering for a large project in the grounds of the Great Beguinage of the city of Tienen (Vander Ginst & Smeets, 2011). Time was

	iSpace (minutes)	Robotic Total Station (min.)
Recording	57	137
Relocating to 2nd trench	25	3
TOTAL	82	140

Table 2. Results of the 2nd test of recording times between Nikon iSpace and Robotic Total Station.



Figure 6. Archaeology iSpace in use at the excavation at the Great Beguinage, Tienen.

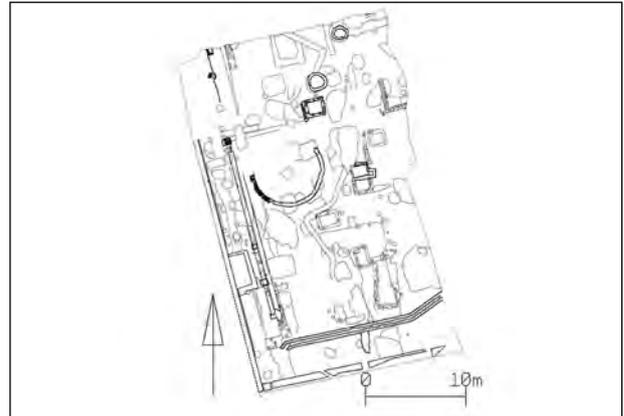


Figure 7. Archaeological plan of the excavation at the Great Beguinage, Tienen, recorded with the Archaeology iSpace.

a major problem with this project as the Flemish government had projected the excavation to last 50 days, while the construction company wanted to commence building after only 32 days. The Flemish government was approached with a proposal to use Nikon Archaeology iSpace as a means of reducing the recording times and, hence, the total excavation time, and Studiebureau Archeologie was awarded the contract. Archaeology iSpace exceeded the expectations in cutting the recording times and the project could be finished in just 30 days, that is, 40% faster than the expected time.

4.2 Experience of using iSpace

At the beginning of May, 2011 Studiebureau Archeologie purchased the Archaeology iSpace

as a means of measuring and drawing on their excavations. As mentioned previously, it was first used on the site of the Great Beguinage in the city of Tienen (Figs 6 and 7) with great success.

From May until November 2011, it has been used on twelve other excavations in central and eastern Flanders (Table 3) including excavations inside churches, in constricted urban sites and in open rural sites up to 6 hectares in size. Traces have ranged in period from Bronze Age to Iron Age, Roman and Mediaeval. With such extensive experience with this new recording system, Studiebureau Archeologie feels well-qualified to comment on its performance compared to other recording techniques.

Great Beguinage, Tienen	Medieval beguinage
Saint-Jacobschurch, Leuven	2 by 8 meter trench
Emblem-Campus Vesta	Iron Age site. approx. 3000m ²
Kontich-Groeningenlei	Iron Age/Roman site. approx. 3000m ²
Poederlee-Schrieken	Medieval site, approx. 1.6ha, rural
Saint-Willibrordchurch, Meldert	Church interior and exterior
Beerse-Holleweg	Medieval site, approx. 2.2ha, rural
Maasmechelen-Mottekamp	Bronze Age to Medieval, > 6ha
Minderbroeders, Mechelen	Part of medieval monastery
Borsbeek-Herentalsebaan	Iron Age cremation burials
Kontich-Rozengard	Roman, 2000m ²
Saint-Gudulachurch, Hamme	Church interior

Table 3. Commercial excavations recorded with Archaeology iSpace, May to November, 2011.

While Archaeology iSpace has a slower setup time than the robotic total station (one person can do this in 15-20 minutes, including calibration), drawing is done much faster. It is quite possible for one person to use iSpace but we find it easier to work with two.

Importantly, any drawing of a trace with Archaeology iSpace is far more accurate than other ways of drawing, both because each point is measured far more precisely and because the drawings contain many more coordinate points than would be used to make both manual drawings or with the robotic total station.



Figure 8. The use of the Archaeology iSpace inside a building. The transmitters are placed to 'look' down into the trench.

A big advantage is that the plan drawings are immediately available as a .dxf file. If any mistakes are made or traces missed, they can be added while the setup remains in place during check-up. Later the plan is put upon the iPad for consultation on site. The plan is inherently geo-referenced since each point in the drawing is a geo-referenced coordinate. This assumes the transmitter arrangement was geo-referenced in the setup stage but, if that was not the case, the plan can be geo-referenced later by measuring temporary survey points that are subsequently established by a topographer.

The very fast speed of drawing with iSpace has seen our archaeologists establish a new routine at assessment excavation sites. They are able to open up a zone, to plane it by shovel, then to register the entire area in just the last hour of the day ready for cross sections of traces to be done at the start of the next day. With some good planning, it is even possible to do the recording of 2 (and on one occasion even 3) different sites on one day.

In open, rural excavations we place the transmitters in the trenches in a rough square up to 40 meters between transmitters. In narrower trenches we use different configurations which, for example, can give us 10m x 80m coverage from one setup. The excavators have experienced the problem of the sensors being "flooded" by the strong signal when a probe is brought within 1,5 meter of



Figure 9. The use of the iJavelin as a giant digital pencil.

a transmitter, resulting in no measurement. The simple precaution is to ensure that no transmitter is placed closer than 2m to any trace at the setup. In confined situations where this might not be possible, our solution is to block the problem signal by placing yourself between the transmitter and the probe and using the signal from the other 3 transmitters to determine the probe coordinates.

In urban excavations there is seldom space in the trenches for the transmitters, so they are placed outside and, since they do not need to be leveled, are angled down so that they 'look' into the trench (Fig. 8). Since the surface area of urban excavations is usually much less than 50m x 50m, it is possible to repeatedly set the transmitters at the same fixed points which can be easily be geo-referenced.

Although Archaeology iSpace comes with a long and a short probe, in 95% of the cases, the archaeologists of Studiebureau Archeology are using the larger iJavelin (in the streaming mode of measuring) like a giant pencil (Fig. 9). The Javelin can be used in a standing position and so is more ergonomic than using the smaller iProbe, which is only used for drawing in more detail, as for individual stones of a wall, skeletons, etc. Up to 40 points each second are measured, but the software is adjusted only to keep 4 per second. Even for a relatively small trace, like the circumference of a large posthole, this means that around 50 points are recorded to make the drawing. This would be unrealistic to attempt by hand or by robotic total station.

During the recording we assign traces, walls, loose findings, etc. to separate layers in the CAD

file. This forces the archaeologists to think more about what they are looking at during the recording. The main benefit of attributing traces to layers at the recording is that it dramatically reduces post-processing. Just minutes after coming off the site, a plan can be finished and imported into the global plan, as at the site of Maasmechelen-Mottekamp, where new site plans were available for use barely minutes after the excavator had finished drawing the traces.

5. Conclusions

It is the experience of the archaeologists at Studiebureau Archeologie that the Archaeology iSpace tracking system has proven itself over many excavations to be a faster way to record excavations. While it is clearly much faster than manual drawing, a series of trials demonstrated that, above a threshold of amount of drawing to be done, it was also faster than drawing by Robotic Total Station. This speed can translate into a substantial reduction of the total excavation time and we estimate that at the Maasmechelen excavation (see 3.3 above) more than 2000min - almost a week - was saved by using Archaeology iSpace.

At the relative recording speeds discussed above, we reflect that if we had been able to use the system at the large Fochplein excavation in Leuven (see 3.1 above) the duration may have been shorter by some 50 days. Similarly, the recording time at the Kontich-Babbekroonbeek excavation (see 3.2 above) might have been halved.

Additional benefits of using Archaeology iSpace have been that recording becomes much simpler and more accurate, while it provides the archaeologists with an updated site plan almost immediately. Also, it makes them think more about, and understand better, the archaeology they are recording.

References

Avern, G. J. 2001a. "High-resolution computer imaging in 2D and 3D for recording and interpreting archaeological excavations." PhD Diss. Université Libre de Bruxelles.

Avern, G. J. 2001b. "A New Technique for Recording Archaeological Excavations: Research Progress Report." In *Computer Applications and Quantitative Methods in Archaeology. Proceedings of the 28th Conference, Ljubljana, April 2000*, edited by Z. Stančič and T. Veljanovski, 3-7. Oxford: Archaeopress.

Avern, G., and W. Franssens. 2012. "A Digital Drawing Tool for Recording Excavations: the Nikon iSpace System." In *Revive the Past: Proceeding of the 39th Conference on Computer Applications and Quantitative Methods in Archaeology. Beijing, 12-16 April 2011*, edited by Mingquan Zhou, Iza Romanowska, Zhongke Wu, Pengfei Xu, and Philip Verhagen, 307-317. Amsterdam: Pallas Publishing.

Barceló, J. A., O. de Castro, D. Travet, and O. Vicente. 2002. "Towards a 3D Representation of Archaeological Layers." In *Workshop 6 Archäologie und Computer, 5-6 November 2001*. CD-ROM. Wien: Phoibos Verlag.

Doneus, M., and W. Neubauer. 2004. "Digital Recording of Stratigraphic Excavations." In *Computer Applications and Quantitative Methods in Archaeology. Proceedings of the 31st Conference, Vienna, Austria, April 2003*, edited by K. F. Ausserer, W. Börner, M. Gorjany and L. Karlhuber-Vöckl, 113-116. Oxford: Archaeopress.

Realli, E., and T. Zoppi. 2001. "Computerised Techniques for Field Data Acquisition." In *Computer Applications and Quantitative Methods in Archaeology. Proceedings of the 28th Conference, Ljubljana, April 2000*, edited by Z. Stančič and T. Veljanovski, 13-17. Oxford: Archaeopress.

Schaich, M. 2002. "Computer Supported Excavation Documentation in a Network of Electronic Surveying." *Photogrammetry, CAD and Databases Workshop 5 Archäologie und Computer. 2000*. CD-ROM. Wien: Phoibos Verlag.

Smeets, M., and G. Avern (*in press*). "Recording Excavations with a Metrology Tracking System." In *16th International Conference on Cultural Heritage and New Technologies. 2012. Vienna, Austria*.

Vander Ginst, V., and M. Smeets. 2011. *Archeo-rapport 64: Het archeologische vooronderzoek in het Groot Begijnhof te Tienen, Kessel-Lo*. Kessel-Lo: Studiebureau Archeologie.

From the Slope of Enlightenment to the Plateau of Productivity: Developing Linked Data at the ADS

Michael Charno, Stuart Jeffrey
Archaeology Data Service, United Kingdom

Ceri Binding, Doug Tudhope
University of Glamorgan, United Kingdom

Keith May
English Heritage, United Kingdom

Abstract:

The Archaeology Data Service (ADS) has a mandate to provide a digital repository for outputs from research funded by AHRC, NERC, English Heritage and other bodies. Archaeology has seen increasing use of the Web in recent years for data dissemination, and the ADS holds a wide range of datasets from archaeological excavations. However datasets and applications are currently fragmented and isolated. Different terminologies and data structures hinder search and comparison across datasets. Because of these impediments, archaeological data can be hard to reuse and re-examine in light of evolving research questions and interpretations. In an attempt to address this, the ADS have begun to ingest some of its excavation data into a triple store and expose it as linked data. This paper will briefly discuss the STAR and STELLAR projects which led up to the development of ADS linked data and will also outline the technologies used to develop it. In particular, it will discuss the more practical details of creating the triple store, populating it with excavation data in RDF format, and finally publishing it as linked data. Finally, using the ADS linked data as an example dataset, an overview of possible future directions will be outlined, in particular to explore how we can enrich the existing and forthcoming linked data with both archaeological and non-archaeological data.

Keywords:

Linked Open Data, Semantic Web, Excavation Data, CIDOC CRM

1. Introduction¹

The Archaeology Data Service (ADS) is hosted by the Department of Archaeology at the University of York in the UK, and is funded partly by the UK Arts and Humanities Research Council and partly by project-based funding drawn from academia, the European Union, National and Local Government archaeology services and commercial archaeology. The core objective of the ADS is to support research, learning and teaching with high quality and dependable digital resources. It does this by preserving digital data for the long term, and by promoting and disseminating a broad range of data in archaeology. The ADS promotes good practice in the use of digital data in archaeology, provides technical advice to the research community, and supports the deployment of digital technologies.

Corresponding author: michael.charno@york.ac.uk

As a digital archive, it follows the Open Archival Information Systems (OAIS) Reference Model and was awarded the Data Seal of Approval in 2010.

Alongside its core responsibilities, the ADS has made great efforts to engage in research into relevant new technologies. After investigating the potential benefits, the ADS leapt aboard the technological bandwagon that is linked data. The dissemination of archaeological data has always been a key function of the ADS and linked data is a new and potentially promising delivery mechanism which could be used in addition to the existing methods employed by the ADS. Additionally, linked data opens up new levels of interoperability across datasets, something which the ADS have pursued in other projects. These have included a co-ordinating role in ARENA (Archaeological Records of Europe

Networked Access) to develop a cross-searchable European portal, completed in 2004 [<http://ads.ahds.ac.uk/arena/>], and subsequently taken forward as ARENA2 [<http://archaeologydataservice.ac.uk/research/arena2>] within the ESFRI DARIAH infrastructure project [<http://www.dariah.eu/>]. The ADS are also currently a leading partner within CARARE [<http://www.carare.eu/>], which is a best practice network funded by the European Commission's ICT Policy Support Programme. It brings together heritage agencies and organisations, archaeological museums and research institutions and specialist digital archives from all over Europe to establish a service that will make digital content for Europe's unique archaeological monuments and historic sites interoperable with Europeana.

Adding linked data to the existing dissemination techniques will hopefully provide novel ways of interrogating the data held by the ADS, and potentially connect otherwise unrelated datasets to the ADS holdings. The potential of linked data is undeniable, however linked data endpoints, particularly related to archaeology, are currently few and far between. It is hoped that the work described here will be useful to others, which may result in the creation of other linked data endpoints, thereby enhancing the data of all.

2. STAR & STELLAR Projects

The origins of the ADS linked data endpoint began with the Semantic Technologies for Archaeological Resources (STAR) project which was a collaboration between English Heritage (EH), the ADS and the University of Glamorgan [<http://hypermedia.research.glam.ac.uk/kos/star/>]. The project investigated the potential of semantic terminology tools for widening and improving access to digital archaeology resources. It also aimed to develop new methods for enhancing linkages between archived database resources and associated grey literature, exploiting the potential of a high level core ontology, and the application of multi-concept query generalisation techniques to cross-domain research in archaeology (May et al. 2010; Tudhope et al. 2011a). In addition to other outputs from the project, the development of a CRM-EH [<http://hypermedia.research.glam.ac.uk/resources/crm/>] extension to the CIDOC CRM [<http://www.cidoc-crm.org>] core ontology

specifically for the archaeology domain would be a core component to the follow on project, Semantic Technologies Enhancing Links and linked data for Archaeological Resources (STELLAR).

The STELLAR project [<http://hypermedia.research.glam.ac.uk/kos/stellar/>] attempted to develop an enhanced mapping tool for non-specialist users to map and extract archaeological datasets using the Resource Description Framework (RDF) representation aligned to the CIDOC CRM-EH ontology. One of the outcomes of STELLAR was the development of templates to generate RDF from traditional database tables based on CRM-EH concepts and relationships. For example, there were templates created for Contexts, Finds and Investigations, which allowed the data to be organised in ways familiar to most archaeologists. The STELLAR templates were then used in conjunction with newly created STELLAR tools, such as STELLAR.Console, a command line utility, and STELLAR.Web, a simpler, web based application. The STELLAR tools allowed for the easy creation of RDF based on the STELLAR templates, and alignment of the data to the CRM-EH. This made some of the most challenging aspects of linked data creation accessible to the whole archaeological community (Wright 2011). Additionally, the templates themselves are extensible and the addition of new templates is possible, as is the inclusion of new templates into the STELLAR tools (Tudhope et al. 2011b). The STELLAR tools created the building blocks of the ADS triple store, which forms the foundation of the ADS linked data resource.

3. RDF Creation

To publish linked data, the ADS needed to first create data in RDF format that would be suitable to populate a triple store. Linked data is typically created using the graph data structure in RDF format, which is constructed using a three-part form called a triple. Triples consist of subjects and objects, connected by a predicate. The predicate describes the relationship between the subject and object, thereby creating meaning between pieces of data which, when aligned to appropriate domain ontologies, allow the structure and relationships of the triples to be defined across the entire data set. Others choosing to align their datasets to the same ontology then allows data interoperability. For

more information on RDF and triples, see Powers (2003) and the W3C specification (1999). Since RDF data is textual in nature, we needed to select archives that either had database or spreadsheet data, and in particular focused on excavation data. We also wanted to use datasets from both academic and commercial projects, ensuring both these types of projects hosted by the ADS were suitable for publication as linked data. With these considerations in mind, data from sites in the Channel Tunnel Rail Link (CTRL) project [doi:10.5284/1000120], two Aggregates Levy Survey Fund (ALSF) project sites [doi:10.5284/1000365 and doi:10.5284/1000392], and data from the Excavations at St Peter's Church, Barton-upon-Humber [doi:10.5284/1000389] were selected. An initial manual examination of the data held in the project databases was also carried out to confirm that the data would be compatible with the existing STELLAR templates.

The standard ADS archiving procedure for preserving and disseminating database data is to break down the relational database into its smallest common denominator, the table. The resulting digital object is a Comma Separated Values (CSV) file for each table. Tables were selected which contained data matching pre-defined STELLAR templates. For this exercise, the data being converted into triples was archaeological contexts, events, finds, samples and groups information. Not all databases contained every element of this information, although all contained at least context and event information.

The process of preparing the database tables for the STELLAR templates effectively involved mapping field headings from the tables to match fields expected by the STELLAR templates. This required an ADS Digital Archivist to identify the relevant field in the CSV file and then rename them to match the templates terminology. For example, when applying a table to the context template, the name of the field containing the context info was changed to "context_id". Some other minor tinkering was required, such as concatenating start date and end date fields to create a date range field, but nothing which required significant reworking by the curator. With this prep work done to the database tables, the data was ready for the STELLAR tools to build RDF based on the CRM-EH.

Before the RDF creation a URI prefix (namespace prefix in the STELLAR tools) had to be decided upon, which would establish the final URI for each node in the linked data. A proper URI prefix is essential for linked data to work, and the STELLAR tools allow users to define their own URI prefix, making it meaningful even outside of STELLAR or the ADS. In the case of the ADS linked data, the prefix was the data subdomain within the main ADS domain, which was <http://data.archaeologydataservice.ac.uk>. This was chosen according to linked data best practice and follows on from other established linked data providers, such as the Ordnance Survey (OS) and the UK Government's data.gov.uk. Additionally we needed contexts beyond the URI prefix to provide uniqueness, and for this we used the Digital Object Identifier (DOI) of a collection. The basic structure of the URLs were:

- [http://data.archaeologydataservice.ac.uk/\[COLLECTION_DOI\]/\[CRMEH_NODE\]](http://data.archaeologydataservice.ac.uk/[COLLECTION_DOI]/[CRMEH_NODE])

which for an actual piece of data would look like this:

- http://data.archaeologydataservice.ac.uk/10.5284/1000362/EHE0061_0

Any further additions to the URI were archive dependent. For instance, CTRL sites had the same DOI, so an extra level of uniqueness was added to these, which was the site name. These URIs are the location of the node, and how it will be referenced by all other nodes in the linked data realm, so careful consideration should be given to how it is constructed.

4. The Triple Store

A container, such as a graph database, is typically needed to host the newly created RDF triples. One type of graph database is a triple-store, which is like a traditional Relational Database Management System (RDBMS), but instead of holding data in a strict relational manner the triple-store allows for flexibly defined relationships (or edges) between nodes of data. And similarly to a RDBMS, the triple-store enables the querying of data contained in it, usually via a SPARQL Protocol and RDF Query Language (SPARQL) interface. The

triple-store chosen by the ADS was AllegroGraph from Franz, Inc (<http://www.franz.com/agraph/allegrograph>). This system was chosen because it was free to use for up to 50 million triples, which allowed us to effectively test the feasibility of a graph database system with no financial risk. There were other products available, such as D2R [<http://d2rq.org/d2r-server>], but AllegroGraph was chosen since it was seen as more flexible and scalable. After 50 million triples, Franz, Inc. requires users to buy a developer edition license good for up to 500 million triples, and then after that it goes to the enterprise edition, which allows unlimited triples.

AllegroGraph, at the time of development, only ran on the 64 bit Linux architecture, so an old desktop machine was re-purposed as a 64 bit Ubuntu Server. This was only a temporary measure until a dedicated server for our machine room could be sourced, but the under-powered desktop machine was more than powerful enough for this dedicated purpose. Configuration of AllegroGraph was straightforward and sufficiently documented by Franz, Inc., allowing us to get the instance up and running very quickly. At the time of writing, AllegroGraph supports virtual machines, but that setup is not recommended for production services. Additionally they intend to support Windows, Mac and Solaris in the future, but this was still in development at the time of our implementation.

AllegroGraph offers a number of ways of connecting to the triple store, with the two primary methods being AG Web View, a web interface for managing and updating the triple store, as well as Gruff, a desktop based application which can connect to local instances of AllegroGraph. AG Web View was the primary interface used by the ADS to populate and manage the triple store. It supports uploading triples in a number of RDF serialisation formats (RDF/XML, Turtle, TriX) as well as the ability to upload individual triples manually. Many other advanced features are available via AG Web View, such as SPARQL querying and general repository management. Gruff also offers most of the capabilities of AG Web View, in addition to the ability to visualize the data. Gruff is capable of working with other triple stores or standalone RDF data as well, which makes it a very good tool for anyone using RDF or semantic data in general.

The RDF discussed in the previous section was ingested into AllegroGraph and produced 644,114 triples. These could then be interrogated using the SPARQL query interface in AG Web View, but this was only possible on the local network, since AG Web View was not publicly accessible due to University of York network restrictions. To enable access to the outside world, we set up a reverse proxy in Apache to forward all traffic to the “/sparql/*” context of the data subdomain (<http://data.archaeologydataservice.ac.uk>) to AG Web View. This allowed external access to our triple store, and the ability to send HTTP requests containing SPARQL queries or similar commands. Third party SPARQL clients, such as Sesame, can also connect to our SPARQL endpoint because of this arrangement. Unfortunately much of AG Web View’s functionality is not enabled for usage outside of the University of York network, which initially seems to be a combination of the reverse proxy and hard coded Ajax commands in the AG Web View application, but this has not been investigated further since the lost functionality of AG Web View outside of the University of York network can be replicated by the available client applications such as Sesame. A simple SPARQL query interface was also set up to allow external users to run and test SPARQL queries against the triple store from the web directly. This tool can be accessed at <http://data.archaeologydataservice.ac.uk/query> and has a number of pre-defined queries for testing purposes.

5. Linked Data

The triple store alone does not create linked data, and another application layer is needed to expose the triple store as linked data. The application chosen to do this was Pubby (<http://www4.wiwiss.fu-berlin.de/pubby>), a linked data frontend for SPARQL endpoints developed by Richard Cyganiak and Chris Bizer at Freie Universität Berlin. Pubby is a Java web application, so it can be deployed in any Java web container. We used an instance of Tomcat on our dedicated ‘semantic’ server to deploy the Pubby WAR, but any Java container could have been used (Glassfish, JBoss, Jetty, etc). Pubby required some configuration, as the SPARQL endpoint (AllegroGraph) and the web base (<http://data.archaeologydataservice.ac.uk>) need to be defined. Additionally, the Pubby interface was customised to look more like an ADS website instead of the default



Figure 1. A sample screenshot of the ADS linked data front-end.

Pubby style.

Once Pubby has been setup and configured to connect to the AllegroGraph SPARQL endpoint, Pubby ‘publishes’ the linked data. Pubby also provides content negotiation from HTTP requests for nodes in our triple store, so depending on whether a machine or human is making the request, the correct content can be provided. Typically, a human uses a web browser, in which case Pubby provides an HTML representation of the data, and a machine will use a standard HTTP request with no user agent header defined, in which case an XML representation of the data will be provided, although this can be overridden in either case. The URL of the node is dynamically changed in this process as well, appending “page” to the namespace prefix for HTML representations and “data” for XML representations.

Actual URL of a node in the Cobham dataset: `http://data.archaeologydataservice.ac.uk/10.5284/1000230/cobham/`

HTML representation of node: `http://data.archaeologydataservice.ac.uk/page/10.5284/1000230/cobham/`

XML representation of node: `http://data.archaeologydataservice.ac.uk/data/10.5284/1000230/cobham/`

Effectively, Pubby provides URL access to

the node from the data in the triple store, since the data in the triple store is not actually outward facing (other than via the SPARQL endpoint). So when a URL such as the one above is requested, it is processed by Pubby and delegated to AllegroGraph. Pubby also displays all the related links for the node requested, so if an archaeological context is selected from the ADS triple store, all related links to that context will also be listed (Fig. 1).

Likewise, since links are bi-directional by default, all contexts related to a particular CRM-EH element will also be defined. For example, when selecting a node related to CRM-EH element “cut”, all contexts which contain that element will be listed as well.

6. New Possibilities

Linked data provides many new possibilities with regards to the interrogation of data. Traditionally, data stored in RDBMS’ is only accessible via the Structured Query Language (SQL), which requires data to be structured in a predefined manner. One key difference with linked data is that the data is not ‘structured’ by the system, but rather the structure is defined by the data. The data is therefore more flexible both structurally and in queryability. This is because the data is not slotted into a predefined table structure, but is held in an unstructured repository. An analogy would be putting like data types into a series of labelled buckets with predefined links between them for a traditional RDBMS, whereas linked data is like putting all the data, regardless of its type, into a big barrel with all the relationships defined within the data. In the series of buckets model, getting any data out of it requires the user to know how every bucket is related, and new links between buckets cannot always be easily made. In the barrel model, all data exists in the same container, and all relationships between the data are defined within the data. This inherent structure within the data is flexible, and can be changed at any time without the requirement for new buckets or containers.

Graph databases follow the NoSQL model, and are seen to be more scalable and flexible (Cattell 2010). Although the data does not need to be structured from the outset like in a RDBMS, the relationships between data still need to be known

before the data can be effectively interrogated. Just because data is stored in a triple-store does not automatically mean that it is immediately discoverable. Writing SPARQL queries requires a basic understanding of the relationships or predicates existent in the triple-store, and without this knowledge, querying is very much like stabbing in the dark.

This intrinsic structuring and the openly accessible nature of linked data makes it truly interoperable, something many projects have attempted via differing approaches with varying degrees of success. Typically, attempts at interoperability have suffered by the requirement for data to become homogenized in one way or another. Linked data does not enforce a shared structure on the data or require the need for data to be mapped to a common standard, both of which can take time and further limit the usefulness of data. However, that data within the triple-store needs to be 'known' to a degree, and in particular which relationship schema(s) applies to the data. Linked data can be mapped to domain ontologies, which enable data from different sources to be shared without losing the original meaning of the data (Wright 2011).

7. Future Directions

Linked data is often seen as a way to turn the classic web into a 'web of data'. To make this happen however, we need other data to create a 'web' with, and thankfully there are already some mature efforts under way to develop these data sources. The biggest hurdle to get over for archaeology is the urgent need to convert domain ontologies which define common terms of the archaeological process, such as period definitions, into Simple Knowledge Organization System (SKOS) formats. Having a SKOSified version of period terms which linked data stores could link to would allow for multiple period definitions to be associated with a feature or artefact node. For example, imagine an artefact that has been identified as Romano-British by the original investigator, but later analysis has raised the possibility of the artefact being later in date. Within a linked data model this is simple to represent these dual interpretations without significant modification to the data, as new triples can be added as and when they are necessary. So the artefact, represented by a node, can have additional

nodes and edges added describing different or new meanings and interpretations related to the artefact. In this example, the artefact could have a new period specific node associated with it, from another ontology or thesauri, without modifying the existing data. Additionally, the new interpretation can be described explicitly in the data, with a new series of nodes and edges describing that abstract meaning, knowledge, or thought process behind the new interpretation. Similarly, new triples can be added to provide additional meaning to existing interpretations, such as including date ranges for the period terms, enabling greater interoperability capabilities without the need for data mappings or data conversions. By adding on multiple layers of interpretation and meaning to data, a much richer and more flexible set of data is possible, which follows more closely to the way data is structured and organized naturally by humans beings. Linked data enables the possibility of a knowledge organization scheme which is both flexible and powerful.

To fully realize the power of linked data though, more than the links within an isolated triple-store are necessary, and this can only be accomplished through linking of multiple external linked data repositories. There are a number of services in existence that publish linked data which have archaeological relevance, and therefore can be utilized by the ADS, in addition to hopefully being enriched by the ADS linked data as well. The ADS has already linked to OS data, although somewhat superficially at this point. Triples were manually ingested into the triple store for OS place names of where the excavations took place, as well as the "England" node added to the root node of the ADS linked data. These links are for the site nodes Dublin Core coverage properties, and enrich the data with modern place name references. Administrative boundaries and place-names change over time as well, but this can be represented in linked data like mentioned above. The Digital Exposure of English Place-names (DEEP) project plans to provide this information as linked data (Dunn 2011), which could then be linked to from the ADS linked data to further enrich it. Looking forward, the other obvious candidates for inclusion would be the Portable Antiquity Scheme (PAS) data, as well as the British Museum (BM) data. The initial thinking is that the ADS excavation data could link to related finds in the PAS via common OS or Geonames nodes or via

a yet undefined common vocabulary. Likewise, data in the BM repository could be associated with related excavation data in a number of ways, either similarly to the PAS suggestion above or with other interpretative data held in the BM. Similarly, more general concepts or organisations defined by DBpedia could also be linked to from our data, which could provide greater contextual information.

As stated previously, truly useful linked data in archaeology requires published thesauri, so that data can always be qualified and precise. In England, EH publishes a number of thesauri related to archaeological monuments, objects, and events amongst others (English Heritage 1999). Ideally, these thesauri would be published as linked data, which would enable the ADS and anyone else publishing linked data to consistently and accurately describe their data. Otherwise, data interoperability becomes more difficult and the data less useful in general. The data can still be described in a comment field as a thesauri term, but only as linked data can the meaning be accurately confirmed, and therefore provide interoperability with other nodes. This would be a great asset to the community creating and using linked data, and hopefully EH will pursue this in the near future.

A similar contribution from the ADS could be the publications of the SKOSified Roman Amphora [doi:10.5284/1000021] archive data. The Roman Amphora project provides an online resource for the study of Roman Amphorae, and in doing so created a comprehensive catalogue of Roman amphora types, fabrics, and related information. Work was done by Leif Isaksen to get this data into RDF [<http://romanamphorae.freebase.com/>], which could be extended to provide an authoritative resource on Roman amphora as linked data. This would then enable other projects to openly reference the ADS linked data and add further reliability to their own data and metadata.

Beyond making data held by the ADS accessible as linked data, it may also be beneficial to manage the internal ADS preservation process as linked data. Currently the ADS manage collection level data in its bespoke Collection Management System (CMS), which is only for internal use. It uses a traditional RDBMS and Java web application front end. Given the benefits of linked data with regards

to flexibility and scalability, it would potentially be beneficial to convert the RDBMS data into a graph database, and then into linked data. This would give greater flexibility to the data than currently exists, but new tools and interfaces would be necessary to make the management and usability appropriate for digital archivists, which could require considerable work. A further advantage of converting the CMS to linked data would be that data held within the CMS could also enrich the ADS excavation linked data by providing greater contextual information related to the general project.

While linked data has certainly been victim to the ups and downs of the Hype Cycle (Gartner 2012), it does seem to be finally arriving at the Slope of Enlightenment. Isaksen et al. (2010), Martinez et al. (2010) and Wright (2011) have convincingly espoused the values of linked data in archaeology. More people are getting involved in developments around this subject, but many more participants are needed to ensure that it reaches a critical mass of both data and users. With the barrier for creating linked data lowering due to the greater availability of both tools and knowledge, it is becoming more pervasive, and a future 'web of data' is appearing to become more feasible. In the same way that the classic web began, linked data began in the realm of specialists and research institutions, the question is now whether linked data will become the democratising tool that the classic web became as it matured, and allow non-technical users and data holders alike access to its potential.

Acknowledgements

The STAR and STELLAR projects were supported by the Arts and Humanities Research Council [grant numbers AH/D001528/1, AH/H037357/1].

References

- Cattel, R. 2010. "Scalable SQL and NoSQL data stores." *SIGMOD Rec* 39 (4): 12-27. doi:10.1145/1978915.1978919.
- Dunn, S. 2011. "Updated project description. The Digitisation of England's Placenames." Accessed 15 June 2012. <http://englishplacenames.cerch.kcl.ac.uk/?p=27>.
- English Heritage. 1999. "National Monuments Record

- Thesauri.” Accessed 12 June 2012. <http://thesaurus.english-heritage.org.uk/>.
- Gartner Inc. 2012. “Hype Cycle Research Methodology.” Accessed 15 June 2012. <http://www.gartner.com/technology/research/methodologies/hype-cycle.jsp>.
- Isaksen, L., K. Martinez, N. Gibbons, G. Earl, and S. Keay. 2010. “Linking Archaeological Data.” In *Proceedings of Computer Applications and Quantitative Methods in Archaeology (CAA2009), Williamsburg*, edited by B. Frischer, J. Webb Crawford and D. Koller. Oxford: Archaeopress.
- Martinez, K., and L. Isaksen. 2010. “The semantic web approach to increasing access to cultural heritage.” In *Revisualizing Visual Culture*, edited by C. Bailey and H. Gardiner. London: Ashgate.
- May, K., C. Binding, and D. Tudhope. 2010. “Following a STAR? Shedding more light on Semantic Technologies for Archaeological Resources.” In *Proceedings of Computer Applications and Quantitative Methods in Archaeology (CAA2009), Williamsburg*, edited by B. Frischer, J. Webb Crawford and D. Koller, 227-233. Oxford: Archaeopress.
- Powers, S. 2003. *Practical RDF*. Sebastapol: O’Reilly.
- Tudhope, D., K. May, C. Binding, and A. Vlachidis. 2011a. “Connecting Archaeological Data and Grey Literature via Semantic Cross Search.” *Internet Archaeology* 30. Accessed 18 March 2012. http://intarch.ac.uk/journal/issue30/tudhope_index.html.
- Tudhope, D., C. Binding, S. Jeffrey, K. May, and A. Vlachidis. 2011b. “A STELLAR Role for Knowledge Organization Systems in Digital Archaeology.” *Bulletin of the American Society for Information Science and Technology* 37 (4): 15-18.
- World Wide Web Consortium. 1999. “Resource Description Framework (RDF) Model and Syntax Specification.” Accessed 15 June 2012. <http://www.w3.org/TR/1999/REC-rdf-19990222/>.
- Wright, H. 2011. “Seeing Triple: Archaeology, Field Drawing and the Semantic Web.” PhD Diss. University of York.

Reflections on the Rocky Road to E-Archaeology

Geoff Carver and Matthias Lang

Georg-August-Universität Göttingen, Germany

Abstract:

Bridging the gap between humanities and science (or, according to other classificatory schema, between social and natural – or between “soft” and “hard” – sciences), archaeology should be an ideal discipline for digital applications. That the reality has proven otherwise offers an opportunity for examining the way various disciplines adopt – or adapt to – technological change. The diversity of potential data sources available to archaeologists makes it easy to understand how archaeology could benefit from the use of new technologies, especially since archaeology has long been criticised for producing too much data. More important is the fact that much data has always been difficult to access, stored either in remote archives or published in obscure journals of limited distribution or limited edition monographs. Computing has simply added more items to the menu – more data from more data sources – without necessarily addressing continued problems of data accessibility. That there are other aspects worth considering will be shown by examining the example provided by the ArcheoInf project. Developed by researchers at the Universities of Göttingen, Dortmund and Bochum, ArcheoInf was intended to provide German Classical Archaeologists with a long-term repository for primary excavation data. Until recently, it has been fairly common for each archaeological project (excavation, regional survey to discover new sites, etc.) to develop its own database. Since most of these used their own data structure and terminology, it is often difficult for data to be shared between one project and another, even in those rare cases when the databases themselves are accessible. So it was intended that ArcheoInf would integrate individual site databases into a single, easily-accessed database, and thereby provide a service similar to others which have already been successful elsewhere (notably the UK and Ireland). Development of ArcheoInf has been hindered by an unexpected problem in the form of a lack of willingness on the part of prospective users to submit data that can be imported into our database, not even for testing purposes (i.e. to see if the import and search processes work). Thus we face a vicious circle: we can't test the system because we don't have any data, and it is difficult trying to attract prospective users if we can't show the system in operation. So technical problems aside, development has been delayed because of the need to create “synthetic data” which simulates results. So the question is: why should something so useful and so obviously necessary be so difficult to implement? Why aren't archaeologists interested in benefiting from “digital humanities”? The presentation considers a number of answers to these questions, and a number of ways the problems they represent might be overcome.

Keywords:

Databases, Multidisciplinary Cooperation, Institutional Infrastructure, Metaphysics

1. ArcheoInf

This paper considers some of the problems encountered during the *ArcheoInf* project. These problems reflect the social contexts within which archaeology and archaeological computing are performed.

Funded by the Deutsche Forschungs-

gemeinschaft (DFG), *ArcheoInf* is a collaborative project involving the following institutions:

- Archäologisches Institut und Sammlung der Gipsabgüsse der Georg-August-Universität Göttingen,
- Institut für Archäologische Wissenschaften der Ruhr-Universität Bochum,
- Fachbereich Vermessung und Geoinformatik der

Corresponding author: gicarver@t-online.de

Hochschule Bochum,

- Universitätsbibliothek Bochum,
- Lehrstuhl für Softwaretechnologie in der Fakultät für Informatik der Technischen Universität Dortmund,
- Universitätsbibliothek Dortmund.

*ArcheoInf*² (cf. Carver, Lang, and Türk 2012) is intended to provide a long-term data repository combining primary archaeological data recorded during surveys or excavations undertaken by German university departments of Classical Archaeology. As such it might be compared to the *ADS* in the UK³, *DANS* in the Netherlands⁴, *IANUS* in Germany or *tDAR* in the USA.⁵ The major difference being that, whereas *ADS* stores documents, *ArcheoInf* aggregates data from any number of databases in a new database, thereby enabling inter-site queries (the *STAR* project⁶ may provide a closer analogy).

We have been facing unforeseen consequences of the fact that, when developing *ArcheoInf*, we believed the utility of our project was self-evident; we did not take into account the fact that we are archaeologists who feel comfortable with both computers and archaeology. We also did not expect that commitment from potential participants would be difficult to secure after initial expressions of interest; nor, given the lack of real data, the delays caused by the need to create synthetic data (cf. May, Binding, and Tudhope 2011, 114) both to get the system running and to show what the results look like.

A related reluctance for anyone to allow “their” data to be published reflects a proprietary interest. Although the present study considers an example from Germany, Adam Rabinowitz (personal communication) noted that this problem is also common in the US, thus suggesting that it

might be a variant on the widespread problem of unpublished data that Brian Fagan (1996) labelled “Archaeology’s Dirty Secret” (cf. Richards, Winters, and Charno 2011, 146).

The use of modern computer technology risks making this problem worse (cf. Berggren and Hodder 2003, 425, Carver 2006); although the media used for collecting, storing and/or disseminating data have changed, as has (to some degree) the institutional landscape, new aims have yet to be widely recognised. Since archaeologists have a duty to both the public and their professional peers to make data available, “such concerns cannot justify the long-term secreting of archaeological information” (cf. Hodder 2000a, 8). And yet... even members of the Çatalhöyük team have discussed the impact making data openly accessible might have on academic careers: students who used to work for low wages in exchange for access to material for a thesis or dissertation can no longer take that for granted (cf. Hodder 1999, 2000a, 8, Hamilton 2000: 126, Rabinowitz and Sedikova 2011, 139). It has also been suggested (Moreno Escobar 2012) that researchers might think of *their* data in terms of a capital investment: the only thing that is theirs, which they can take from one low-wage job to another.

The present study frames these problems in terms of metaphysical divisions between archaeologists and IT-technicians (i.e. computer “geeks” or “nerds”), which is in turn a manifestation of the ongoing conflict over the aims of the different sub-disciplines of archaeology (finds, faunal, geoarchaeology, etc.). Any number of archaeologists have criticised what they perceive to be a “fragmentation” most evident in the divide separating theoretical and “field” archaeology extending to the level of terminology (compare Kuhn 1996, 202-204):

Despite attempts to break down these barriers... each subsystem realm is developing its own vocabulary which is fast becoming incomprehensible to specialists in other fields. Much of the literature, teaching and research in archaeology is divided along these lines. There is a need for integration, which would be the logical

2 www.archeoinf.de

3 <http://archaeologydataservice.ac.uk/>

4 <http://www.ianus-fdz.de/>

5 <http://www.tdar.org/>

6 <http://hypermedia.research.glam.ac.uk/kos/star/>

result of a symbolic and 'contextual' archaeology (Hodder 1992, 27).

Such disciplinary "fragmentation" is not new, and it can be argued that – *pace* Clarke (1973, 18) – archaeology was never "one discipline."

Hodder's call for a "symbolic and 'contextual' archaeology" echoes other calls for revolution (derived, perhaps, from Thomas Kuhn's concept of "incommensurability."⁷ The emphasis on "context" and "reflexivity" are in some ways reminiscent of David Clarke's focus on metaphysics; but instead of "fragmentation" Clarke (1973, 11) argued that "the highly cellular structure of traditional archaeology" *inhibits* revolutionary change "by confining each disconformity to a localized compartment."

Whether such "fragmentation" promotes or inhibits change is ultimately less relevant than the problem that "fragmentation" is occurring despite the use of databases intended to break down "barriers" between various fields of information (the Members of the Ardnamurchan Transitions Project [2012: 117] noted the divisions between those who are or are not computer literate perpetuates "traditional hierarchies"; cf. Hamilton 2000, 123).

This paper is about trying to understand what went wrong in our attempts to *bridge* disciplinary domains, to act as an *interface* between computer "nerds" and archaeologists.

2. Preliminary Considerations of Archaeological Metaphysics

Despite having a bad reputation with most scientists (cf. Stephenson 2011, 95-96), we all use metaphysics under different labels. Explicit references to metaphysics have been rare in archaeology; exceptions include Clarke (1973) and Melzter (1979).

Aristotle's metaphysical causes provide a convenient example, both because they allow maintaining terminological and schematic consistency with later scientific explanation, and because they provide a framework – or a

metalanguage – for explicitly labelling parts of a process: materials, aims, and means.

Aristotle (2007, 89-92 [V.i-iii; 1013a-1014a]) identified four causes: material, formal, efficient and final. As an example, imagine someone wants to make a statue of Aristotle. The "material cause" is the material something is made of: marble, wood and bronze all need to be worked differently in order to make a statue, and the results will differ accordingly. The "formal cause" is the set of processes employed for transforming the raw material into a finished product. The "efficient cause" is the force that changes something: the sculptor (Michelangelo or Picasso?) applying skills and knowledge to transform these materials in order to achieve a goal, which is the statue or "final cause."

A more archaeological example would include the different types of tools or buildings that can be made from a given material (i.e. stone, bronze or iron). Similarly, the results of archaeological research will be different if the documentation is recorded on paper than if it is recorded digitally (i.e. "the medium is the message"):

Each recording medium (text, digital and still photography, video and audio) forces an archaeologist to take a different angle from which to observe and discuss the subject (Stevanovic 2000, 238).

Ideally, the most efficient action would use the best material for a given task, but we are almost always forced to compromise (Feyerabend 1993: xi; cf. Lévi-Strauss [1966, 17] on "bricolage").

Metaphysics helps construct a narrative, as shown by comparing Aristotle's metaphysics with the literary critic Kenneth Burke's "dramatistic pentad" (1969, xv):

five terms that constitute the grammar of action: the act itself; the agent performing it; the means or agency used; the scene, or context; and the purpose of the action (Lyne 1993, 146 [original emphasis]).

The five terms of Burke's pentad answer common questions any narrative seeks to answer: *what* action took place, *who* did it, *how* did they do

⁷ Although associated with Kuhn, Hacking (2002: 96) noted that credit should also be associated with Feyerabend; cf. Kuhn 1998: 132, Feyerabend 1993: 166, etc.

it, *where*, and *why*? Burke represents an advance on Aristotle because of his inclusion of location (where/scene), which, in archaeology, is important for questions of provenience and/or provenance (whereas in most sciences, theoretically “where” a given experiment takes place should not play a role in the results, so long as all the other conditions remain constant, archaeology is largely concerned with what Petrie [1904, 50] labelled the “main evidences of position”). It should be obvious why this concept could be equated with metadata in a computing environment, and “reflexive methods” (cf. Hodder 1997, 1999, 2000b, etc.)

Archaeology is further complicated by the fact that it has a double narrative (cf. Binford 2001: 46): we need to consider both the metaphysics of the events in the past *and* those of the archaeological investigation. In that sense archaeology is like a detective novel (cf. Frank 1989), in that one narrative arc tells of the detection (by Sherlock Holmes, Miss Marple, etc.), the other the crime (cf. McLuhan 2002, 106; Blaise 2000, 223).

Francis Bacon’s “idols” (2000, 39-42; cf. Jardine 2000, xix-xx) provide more nuanced extensions to Aristotle’s “material cause.” His Idols of the *Tribe*, *Cave*, *Marketplace* and *Theatre* refer, respectively, to constraints imposed by biological limitations (we need telescopes, microscopes and X-rays to extend our range of vision; cf. Freud 1930, 47-50, McLuhan 2003, 48-49, Tucker 2005, 7), personal experience, language/jargon, and institutions (or tradition).

Our problems with *ArcheoInf* also reflect the need to reconcile two parallel metaphysical structures: the assumptions, experience, language, traditions and goals of archaeology on the one hand and those of computing on the other. Hodder’s criticism of “fragmentation” raises the additional problem that interdisciplinary specialisation, jargon and “traditional” division of labour lead to different – and limited – aims, and overlooks the fact that archaeology has never “unified.” Archaeology has *always* been divided by on-site division of labour (photographer, excavator, draughtsperson, surveyor, finds specialist, etc.; cf. Collis 2001, 39-42, 44-45 [Fig. 2.14] for a historical overview), and further divided by specialisations (faunal [further subdivided into expertise on micro/macroufauna],

flora, dendrochronology, geoarchaeology, region and/or period, etc.): Wordsworth (1994) contrasted the “unlettered ploughboy” who found an artifact, the antiquary who studies it, and the “bard” whose interpretation brings it alive.

The limited results of limited aims are often “fragmented” – poorly integrated – since, if you’re too specialised, it is easy to lose track of the final cause: why we do something, or what anyone’s place is within the archaeological process. The potential consequences of such divisions are illustrated by the differences between how farmers and archaeologists have adopted technology in the almost 200 years since Wordsworth wrote his poem. While some of today’s “unlettered ploughboys” drive tractors equipped with GPS and GIS, many archaeologists still work with trowel, paper, and pieces of string.

Since computers were supposed to rectify this

– *The electronic ages seems to be abolishing the fragmented and specialist form of work called jobs and restoring us to the non-specialized and highly involved form of human dedication called roles (McLuhan 2003, 50) –*

the best solution to problems of “fragmentation” might be better database management. Archaeologists have always recognised that “we have to have a kind of integrity most fields don’t need. I need your data, and you need mine” (Flannery 1982, 276). now that we have to think of our goals in terms of publishing databases, spread sheets, etc., online, where “cellular isolation is no longer possible even were it desirable” (Clarke 1973, 11; cf. Backhouse 2006, 51), recognition of such “fragmentation” should lead to an emphasis on reflexive methods and metadata.

This raises questions of what we want to publish, and how we make our data available to other researchers. “Before future LEAP-style publications can be produced,” for example,

a shift in the way archaeologists prepare, create and think about data must occur. Traditionally the archive has been something of an afterthought, pieced together after the fieldwork is finished. If the

archive becomes part of the publication, however, more care from the outset of an archaeological project must be taken with regards to data creation. This means that full metadata and documentation of the data must be created. Putting the raw data alongside the publication may force data producers to take more care in creating their data. Archaeologists can sometimes create data knowing that few outside the project will ever see it. This new exposure will hopefully force archaeologists to break the uneven data management habits that have historically afflicted the discipline (Richards, Winters, and Charno 2011, 146).

The need to change the “material” by adding “full metadata” reflects new aims (“final cause”) and altered *effective* cause (addressing “uneven data management habits”).

3. Feedback

One of the factors contributing to our problems with *ArcheoInf* is something we call *Technofrust*: “frustration with technology.” *Technofrust* derives in part from of a sense of disappointment: a consequence of hype having led to unrealistic expectations of what computers can do. This disappointment must, however, be balanced against the sometimes limited expectations of people unfamiliar with archaeological computing (cf. Varytimiadis, Kakargias, and Stergiopoulou 2011: 150). A lack of documentation, backup or update strategy (outdated file formats on obsolete recording media) for the databases collected for *ArcheoInf* indicated that they had been put together by archaeologists lacking expert knowledge who had never intended to create anything other than limited, short-term tools for creating lists (such as finds inventories) or doing fairly superficial analysis.

A common response from prospective users was that a lot of software is unnecessarily complex. Software often seems to have been designed by computer “nerds” who want to impress other nerds with just how clever they are. Many “real” people, however, just want something that works and is easy to use; they do not want anything complex with a lot of extras and special effects that they can’t imagine ever needing. This suggests that programmers should not force users to confront

a lot of unnecessary gimmicks. We believe that a simple, user-friendly but robust design will attract users and win their confidence.

Experience has also shown that – once they have mastered the basics – computer users will find that a good system eventually becomes indispensable, and they will want to do more; it therefore helps if everything is built on a robust foundation which can be expanded as necessary. There is nothing more frustrating than investing time, effort and expense in adopting a given system, training staff, etc., only to reach a dead end (the software is obsolete, can no longer be read by updated versions, etc.).

Another problem is that, while our documentation and storage media have changed, the institutional infrastructure essentially has not. Any number of people (Anonymous 2008, 1; Lynch 2008, 28, etc.) have argued that – as goals change – funding needs to be linked to the production of data ready for archiving *and* for the costs of long-term archival storage (cf. Howe et al. 2008, 47-48; Lynch 2008, 28 etc.).

But even then, when the infrastructure is in place, we still have to educate individual researchers about *why* they need to do that extra work. Archaeologists do more than just excavate and/or publish journal articles and monographs; some now expect to make our data available for use in PowerPoint presentations or online, etc., but many still seem to perceive computer usage as something obligatory. Thus “data exchange with repositories is predominantly a one-way process: it is compulsory to deliver finds documentation to the repositories, but there is very little demand for digital data coming from the repositories” (Verhagen, Sueur, and Wansleeben 2011, 155), and it is to be expected that potential users will be reluctant to invest time and energy in designing or learning to use a better database, and then investing more time and effort entering data which will not be of any immediate and/or obvious use. This is reflected by the fact that those who lead rescue excavations are often only obliged to submit data to a state service for storage in a certain format – after which it essentially disappears: in many cases the excavators might never see it again (cf. Hamilton 2000, 123-126; May 2012, 168; Yamin 2012, 136 etc.) – and rarely perform anything more than the most superficial

analysis. A GIS might be used to produce maps for presentation, or a program might draw and print a Harris Matrix, but the Matrix itself will not be analysed further (Chadwick 1998).

We believe that archaeologists have not yet created the demand – have not yet reached a critical mass – without which there is no incentive to do more than just invest the minimum of effort required in the writing of obligatory reports:

a difficult step is convincing members to exchange the data with others. The more members (of more different disciplines) join, the better the global research (Eckkrammer, Feldbacher, and Eckkrammer 2011, 84).

This “more the merrier” attitude mirrors our observation that a certain amount of data must be available to show that the investment in time and effort will yield worthwhile results. Thus the need to create “synthetic” data (a problem also encountered by May, Binding, and Tudhope [2011]) not only to test *ArcheoInf* for functionality, but also to show that the project is not a waste of time.

We need to find a balance between “customer is always right” and a computer expert providing informed guidance. Being aware of – if not explicitly modelling parallel metaphysical assumptions – helps make these differences explicit, and thus helps with the translation Kuhn (1996, 202) argued was necessary for overcoming “incommensurability.”

4. The German Question

This process of fragmentation is accentuated when, despite claims that “archaeology is one discipline” (Clarke 1973, 18), the domain is already strongly divided, as is the case in Germany.

Germany provides a good example partly because, as Ian Hodder noted (1999, 9, cf. Taylor 1948, 15-16), “it is widely recognized that German archaeologists dig with a different method,” partly because the history of German *Archäologie* may already be somewhat familiar to an international audience, but largely because – unlike French or Scandinavian archaeologies – German *Archäologie*

provides a clear contrast to the Anglo-American model included in many introductory textbooks and histories (i.e. Bahn [1996] and Trigger [1989, 2006]), etc.

British definitions often emphasise the fact that, historically, *prehistoric* archaeology has been the *purest* form of British archaeology:

Archaeology is that branch of science which is concerned with past phases of human culture; in practice it is concerned more, but not exclusively, with early and prehistoric phases than with those illustrated by written documents (Crawford 1960, 15).

In this equally extreme German example, *Archäologie* is defined as the study of the material heritage of ancient cultures of the *Mediterranean* area:

so lässt sich Archäologie bestimmen als die Wissenschaft vom materiellen Erbe der antiken Kulturen des Mittelmeerraumes (Niemeyer 1968, 9; cited by Häuber and Schütz 2004, 15-16).

German *Archäologie* – which can be traced back to Winkelmann and has its roots in art-history, and is thus perceived as being more of a “humanity” than a “social science” – is and has been associated with what Anglo-American archaeologists consider to be the *sub-discipline* of *Classical* Archaeology, while prehistoric archaeology is a separate discipline with separate university departments and institutes, societies, etc., known either as *Ur- und Frühgeschichte* or *Vor- und Frühgeschichte*.

The significance of these disciplinary distinctions can be illustrated by attempting to translate Binford’s famous adage that “archaeology is anthropology or it is nothing” – into German. In their overviews of archaeological theory, Bernbeck (1997, 37), Veit (1998, 122) and Kümmel (1998, 122) do not even try.

Among other things, German *Anthropologie* refers to what Anglo-Americans know as *physical* anthropology. And although Binford might be right about US-American archaeology (where archaeology has traditionally been taught as a sub-

discipline of anthropology), in Britain archaeology might be thought of more in terms of being a prehistoric *extension* to history, whereas early Canadian archaeology was often done by geologists and published in geological journals (Mackie 1995, 181).

We generally ignore the fact that the English word *archaeology* does not have quite the same meaning (it does not cover the same disciplinary domain) as the German word *Archäologie*, or the French *archéologie*, etc. (cf. Deutscher 2011 for a recent overview of the relations between language and thought), etc. That may be fine in everyday discourse, where we can all agree that – whatever language we’re speaking – we’re all archaeologists, and we can work out our differences in the bar later on... but what happens when we *leave* the realm of the everyday, and *enter* the computerised world of cyberspace?

This problem of defining our “domain” provides a useful example because it should already be somewhat familiar to anyone acquainted with ontologies from building databases and using GIS or the CIDOC-CRM: we recognise the benefits of having clear definitions of artefact classes and monument types, etc. (cf. Clarke 1978, 23; Baines and Brophy 2006, 236). Yet such divisions extend all the way down to at the level of lists used to describe soil colour (compare Murray, Roskams, and Spence 1994, subsection 2; Finnern et al. 1994, 106-107 and ROB 1998, 4.2.1.3.1 - 4).

More significantly, in contrast to the UK – where Single Context Planning is more or less standard (as it is in Ireland, France, Slovenia, etc.) – German archaeologists use 16 different sets of documentation systems and guidelines (the Federal Republic of Germany is divided into 16 States, the *Bundesländer*, and the *Länder* are responsible for culture; cf. Carver 2004). Among other things, different laws regarding private excavation firms means that units have remained small relative to their American or British counterparts. Such limited mobility results in another type of “fragmentation.” This fragmented landscape has also inhibited the development of what Pitt-Rivers (1892, 24) called “a permanent Corps of efficient workmen”: the “circuit diggers” found in the UK, or “shovel bums” in the US.

Single Context Planning was developed with the intention that the person who excavates a given context is the person best-suited to do the relevant documentation (“interpretation at the trowel’s edge”; cf. Spence 1993, 25-26; Berggren and Hodder 2003, 424):

the practical excavator should appreciate more than any other archaeologist the degree to which his practice is controlled by his theoretical expectations, and these should accordingly be appropriate.... Thus with a more explicit theoretical awareness the practical excavator may contribute to a qualitative increase in understanding rather than simply a quantitative increase in data. In any case, practical men who believe themselves to be quite exempt from any intellectual influences are... usually the unwitting slaves of some defunct theorist (Clarke 1973, 18).

This assumes that excavators have enough “practical experience” and are sufficiently “competent,” an assumption that overlooks the roles played by training (Clark 1992, 18), “skill” (Lucas 2001, 39), “sharpness of the eyes” (Droop 1915, 17; cf. Franken 1984, 17), etc., the reality that many of the people with such “practical experience” often supervise and no longer excavate (cf. Chadwick 1998; McAdam 1992; Edgeworth 2011, 45).

Then again, “What is experience?” (Lönn 2012, 69). Another one of Bacon’s “idols.”

In contrast, McAnany and Hodder (2009: 5-6) refer to methodologies employed in “colonial contexts” where “the stratigraphic section allowed control, either of unskilled ‘native’ labourers or unskilled workmen, the unemployed or prison inmates – and later students – in Europe and the United States” with the “different social context” associated with use of the Harris Matrix: “large-scale urban excavations in Britain and the emergence of contract professional archaeology”), etc. A distinct division of labour is still the norm on German excavations, with clear distinctions between academically-trained archaeologists, technicians (*Techniker*), photographers and surveyors, and the often unskilled “*Hilfskräfte*” (cf. Planck 1999, 16/VLA 2006, 9) who do most of the digging.

Has anything *really* changed since Wordsworth's time?

Clarke's critique (1973, 11) of "the highly cellular structure of traditional archaeology" suggests reflexive methods could not evolve within such a "fragmented" environment. *Hilfskräfte* and *Technikers* do not see the results of their contributions, and archaeologists – often having little practical training or experience – are unable to critique methodology. Thus, since no one has a comprehensive overview of the archaeological process from excavation to published report, the "system" lacks a feedback loop that could evaluate whether what comes out at one end is a faithful reflection of what came in the other (raising the risk of "garbage in, garbage out").

"Publicity" (i.e. "the free exchange of scientific knowledge" [King et al. 2009, 85]) was another aspect of the Baconian ideal: "the exposure to the criticism of others of the evidence and reasoning on which the findings claimed are based" was viewed as being "essential for valid scientific work" (Quinton 1980: 31). But divided between 16 states, reflecting the on-site division of labour, etc., archaeological publications directed towards a small audience focus discussion on minutiae; it becomes difficult for archaeologists to talk to computer "nerds," or for any fragment to attain the "critical mass" necessary for the feedback loop that Bacon argued leads to good science.

Since it is unclear whether Bacon's influence on British antiquarianism relative to the more "humanistic" bent of German-language *Archäologie* was significant or not, a more important issue related to Bacon's issue of "publicity" is "who are we publishing for"? In addition to the "specialist reports" Hodder criticised, we have all that "grey literature": mandatory reports which are rarely accessible (cf. Flannery 1982; Snow et al. 2006).

So in that sense a project like *ArcheoInf* – by making data available – faces any number of obstacles (Bacon's "idols of the theatre").

5. DAI

A development which recently came to our attention is that the German Archaeological

Institute (Deutsche Archäologische Institut/DAI, with 15 departments and divisions worldwide) is funding research into a new project – *IANUS* – which is intended to perform the same tasks as *ADS*, *DANS* or *tDAR* (cf. Dally, Fischer, and Schäfer 2013). Although *IANUS* is only now launching a 3-year feasibility study, it is our belief that some of our prospective "customers" may have decided to wait for *IANUS*, rather than invest time and effort with *ArcheoInf*.

In this sense *IANUS* represents our competition, and competition often leads to better results. In this case, however, with such a small – and fragmented – market, it might be argued that we are limiting the usefulness of the results through "fragmentation."

This reveals something of the institutional landscape which, one might think, is not particularly *rational*. Funding from the same source (DFG) for two parallel – competing – projects may not appear to make sense to those of us used to working with limited resources, but in a way it does: the goal of a granting agency such as the DFG is *not* to determine whether one project is better than another, or even necessarily to ensure that one project reaches a successful completion. Rather, the DFG aims to fund worthy research. In addition, we have to consider the fact that funding is often available for developing systems, but not necessarily for their maintenance. For example,

The main reason for the lack of attention to facilitating data exchange... has been insufficient funding. While the system itself was developed with substantial government support, its maintenance budget is very limited (Verhagen, Sueur, and Wansleben 2011, 154).

This reinforces the idea that it is worth considering not only the cultural but also the institutional contexts within which research is performed.

6. Conclusions

This paper was intended to underline the need to contextualise not just the process of data production in the field ("reflexive methods"), but

the whole process of adopting technology within our own discipline. Although archaeologists study the process of adopting technologies by archaeological cultures (i.e. the traditional division of prehistory into the Stone, Bronze and Iron Ages), archaeology itself is not yet “reflexive” enough to consider how technology affects archaeology. Attention has been directed towards the adoption of “reflexive methods” and metadata, but could be strengthened by explicitly considering metaphysics, since this taps into a wider literature (i.e. commentaries on Aristotle and Francis Bacon) and recognition that progress occurs when we knock idols down (Kuhnian revolutions are exercises in iconoclasm).

This “metaphysical” approach runs the risk of oversystematising what is only a model (cf. Sabloff, Binford, and McAnany 1987, 205), a terminological framework which provides convenient labels for parts of a process. There is no need, for example, to classify a certain archaeological tradition and/or institution as being a “material,” “formal” or “efficient” cause, because – depending on the scale or focus of causal analysis – something like the strongly delineated social hierarchy in German-language *Archäologie* can and does play all these roles. Within the framework of systems analysis it might even be argued that (maintenance of) a tradition and/or institution is itself a “final cause”: *this* is the tradition we start with, *this* is the tradition in which we work and which shapes how we see and interpret the world, *and* this is the tradition we seek to *maintain* (i.e. cultural evolution is governed by “system-maintaining and system-destroying variables” [Harris 2001: 240]; for a commentary on the political dimensions of the role homeostasis plays in systems analysis, cf. Shanks and Tilley 1987, 52-53; both Bacon and Kuhn addressed the setting up of “idols” or the role of “normal science”).

The German example is meant to suggest that, where the potential audience for archaeological publications is small and divided, it would be very difficult to attain that “critical mass” which Bacon argued was necessary for good science. German archaeology – with its roots going back to Winckelmann and a more “art-historical” approach – also differs significantly from the British antiquarian model, which emphasised its links to

Bacon (cf. Van Riper 1993, 34-35).⁸

In this case, the focus has been on institutions, contrasting the clearly “fragmented” German landscape with that found in English-language archaeology, while stressing the need to consider media (whether archaeology is documented on paper or digitally [cf. Backhouse 2006, 51]) and aims or goals (so-called “gray literature,” *traditional* publications [journal reports and monographs], online). In order to rectify the situation, it is not enough to merely change aims or media; we also have to ask whether our methods, theories and/or resources can achieve them.

The whole point, though, is that we have limited aims. We only aim to write a report, or finish a database, or hand our archives or samples over to some state or national repository. And we do these things in part to fulfil contractual obligations, earn a university degree, get renewed funding or excavation permits, maintain “traditional archaeology,” etc., but not necessarily in order to do something as abstract as “advance science” if that means having to share “our” data. The aim of sharing data online not only increases the need for metadata and places more emphasis on issues of data quality (cf. Backhouse 2006, 54-55), but underlines the need to consider – and perhaps challenge – underlying metaphysical assumptions.

References

- Anonymous. 2008. “Community cleverness required.” *Nature* 455 (7209): 1.
- Aristotle. 2007. *The Metaphysics*. Translated by John H. McMahon. Mineola, NY: Dover Philosophical Classics.
- Backhouse, P. 2006. “Drowning in data? Digital data in a British contracting unit.” In *Digital Archaeology: Bridging Method and Theory*, edited by T. L. Evans and P. T. Daly, 50-58. London: Routledge.
- Bacon, F. 2000. “The New Organon.” In *The New Organon*, edited by L. Jardine and M. Silverthorne, 26-238. Cambridge: Cambridge University Press.

⁸ For a recent examination of Bacon’s context within British science in general and the Royal Society in particular, cf. Goldstein 2011.

Reflections on the Rocky Road to E-Archaeology
Geoff Carver and Matthias Lang

- Bahn, P. G. 1996. *The Cambridge Illustrated History of Archaeology*. Cambridge: Cambridge University Press.
- Baines, A., and K. Brophy. 2006. "What's another word for thesaurus? Data standards and classifying the past." In *Digital Archaeology: Bridging Method and Theory*, edited by T. L. Evans and P. T. Daly, 236-50. London: Routledge.
- Berggren, A., and I. Hodder. 2003. "Social Practice, Method, and Some Problems of Field Archaeology." *American Antiquity* 68 (3): 421-34.
- Bernbeck, R. 1997. *Theorien in der Archäologie*. Tübingen: A. Francke.
- Binford, L. R. 2001. *Constructing Frames of Reference: An Analytical Method for Archaeological Theory Building Using Hunter-Gatherer and Environmental Data Sets*. Berkeley: University of California Press.
- Blaise, C. 2000. *Time Lord: the Remarkable Canadian who Missed his Train and Changed the World*. New York: Pantheon.
- Burke, K. D. 1969. *A Grammar of Motives*. Berkeley: University of California Press.
- Carver, G. 2004. "Method and theory in German archaeology: preliminary findings." In *Digging in the Dirt: Excavation in a New Millennium*, edited by Geoff Carver, 135-39. Oxford: John and Erica Hedges.
- Carver, G. 2006. "Reflexive Digits: The human factor in archaeological documentation." In *Workshop 10: Archäologie und Computer 7.-10. November 2005*, edited by W. Börner and S. UhlirzWien: Magistrat der Stadt Wien.
- Carver, G., M. Lang, and H. Türk. *In press*. "Recent Developments in the *ArcheoInf* Project – Towards an Ontology of Archaeology." In *Fusion of Cultures. Proceedings of the 38th International Conference on Computer Applications and Quantitative Methods in Archaeology (CAA), Granada, April 6-9, 2010*, edited by Fco. Javier Melero Javier Melero and Pedro Cano.
- Chadwick, A. 1998. "Archaeology at the edge of chaos: Further towards reflexive excavation methodologies." *Assemblage* 3. www.assemblage.group.shef.ac.uk/3/3chad.htm
- Clark, P. R. 1992. "Contrasts in the Recording and Interpretation of 'Rural' and 'Urban' Stratification." In *Interpretation of Stratigraphy: a Review of the Art; Proceedings of a Conference held on 18th June, 1992 at the City of Lincoln Archaeology Unit*, edited by K. Steane, 17-19. Lincoln: City of Lincoln Archaeology Unit.
- Clarke, D. L. 1973. "Archaeology: the loss of innocence." *Antiquity* 47 (185): 6-18.
- Clarke, D. L. 1978. *Analytical Archaeology*. 2nd ed. London: Methuen & Co.
- Collis, J. 2001. *Digging up the Past; An Introduction to Archaeological Excavation*. Stroud: Sutton.
- Crawford, O. G. S. 1960. *Archaeology in the Field*. London: Phoenix House.
- Dally, O., F. Fischer, and F. Falko Schäfer. 2013. "The new Research Data Centre for Archaeology and Historical Studies in Germany." Paper presented at *Computer Applications and Quantitative Methods in Archaeology (CAA), Southampton, March 26-29, 2012*.
- Deutscher, G. 2011. *Through the Language Glass: Why the World Looks Different in Other Languages*. London: Arrow.
- Droop, J. P. 1915. *Archaeological Excavation*. Cambridge: Cambridge University Press.
- Eckkrammer, F., R. Feldbacher, and T. Eckkrammer. 2011. "Cidoc CRM in Data Management and Data Sharing. Data Sharing between Different Databases." In *On the Road to Reconstructing the Past. CAA 2008: Computer Applications and Quantitative Methods in Archaeology. Proceedings of the 36th Conference, Budapest, April 2-6, 2008*, edited by E. Jerem, F. Redö, and Vajk Szeverényi, 80-85. Budapest: Archaeolingua.
- Edgeworth, M. 2011. "Excavation as a ground of archaeological knowledge." *Archaeological Dialogues* 18 (1): 44-46.
- Fagan, B. 1996. "Archaeology's Dirty Secret." *Archaeological ethics*, edited by Karen D Vitelli, 247-51. Walnut Creek: AltaMira Press.
- Feyerabend, P. 1993. *Against Method*. 3rd ed. London: Verso.

- Finnern, H., W. Grotenthaler, D. Kühn, W. Pälchen, Walter-Götz Schraps, and H. Sponagel, editors. 1994. *Bodenkundliche Kartieranleitung. 4th ed.* Hannover: E. Schweizerbart'sche Verlagsbuchhandlung.
- Flannery, K. V. 1982. "The Golden Marshalltown: A Parable for the Archeology of the 1980s." *American Anthropologist* 84 (2): 265-78.
- Frank, L. 1989. "Reading the Gravel Page: Lyell, Darwin, and Conan Doyle." *Nineteenth-Century Literature* 44 (3): 364-87.
- Franken, H. J. 1984. "Lithology and Stratigraphy of Archaeological Sites." *Stratigraphica Archaeologica* 1: 16-23.
- Freud, S. 1930. *Das Unbehagen in der Kultur*. Wien: Internationaler Psychoanalytischer Verlag.
- Goldstein, R. Newberger. 2011. "What's in a Name? Rivalries and the Birth of Modern Science." In *Seeing Further: The Story of Science & the Royal Society*, edited by Bill Bryson, 108-29. London: HarperPress.
- Hacking, I. 2002. *Historical Ontology*. Cambridge, Mass.: Harvard University Press.
- Hamilton, C. 2000. "Faultlines: the Construction of Archaeological Knowledge at Çatalhöyük." In *Towards reflexive method in archaeology: the example at Çatalhöyük*, edited by Ian Hodder, 119-27. Cambridge: McDonald Institute for Archaeological Research.
- Harris, M. 2001. *The Rise of Anthropological Theory: A History of Theories of Culture. Updated ed.* Walnut Creek, CA: AltaMira Press.
- Häuber, C., and F. Xaver Schütz. 2004. *Einführung in Archäologische Informationssysteme (AIS): Ein Methodenspektrum für Schule, Studium und Beruf mit Beispielen auf CD*. Mainz am Rhein: Philipp von Zabern.
- Hodder, I. 1992. "Symbols in action." *Theory and Practice in Archaeology*, edited by I. Hodder, 24-44. London: Routledge.
- Hodder, I. 1997. "'Always momentary, fluid and flexible': towards a reflexive excavation methodology." *Antiquity* 71: 691-700.
- Hodder, I. 1999. *The Archaeological Process: an Introduction*. Oxford: Blackwell.
- Hodder, I. 2000a. "Developing a Reflexive Method in Archaeology." *Towards reflexive method in archaeology: the example at Çatalhöyük*, edited by Ian Hodder, 3-14. Cambridge: McDonald Institute for Archaeological Research.
- Hodder, I. 2000b. *Towards reflexive method in archaeology the example at Çatalhöyük*. Oxford: Oxbow Books.
- Howe, D., M. Costanzo, P. Fey, T. Gojobori, L. Hannick, W. Hide, D.P. Hill, R. Kania, M. Schaeffer, S. St Pierre, S. Twigger, O. White, and S. Yon Rhee. 2008. "The future of biocuration." *Nature* 455 (7209): 47-50.
- Jardine, L. 2000. "Introduction." *The New Organon*. Francis Bacon, vii-xxviii. Cambridge: Cambridge University Press.
- King, R.D., J. Rowland, S. G. Oliver, M. Young, W. Aubrey, E. Byrne, M. Liakata, M. Markham, P. Pir, L. N. Soldatova, A. Sparkes, K. E. Whelan, and A. Clare. 2009. "The Automation of Science." *Science* 324 (5923): 85-89.
- Kuhn, T. S. 1996. *The Structure of Scientific Revolutions. 3rd ed.* Chicago: University of Chicago Press.
- Kuhn, T. S. 1998. "The Natural and the Human Sciences." In *Introductory Readings in the Philosophy of Science. 3rd. ed.*, edited by E D Klemke, Robert Hollinger, David W. Rudge, and A David Kline, 128-34. Amherst: Prometheus Books.
- Kümmel, C. 1998. "Marxistische Perspektive in der gegenwärtigen englischsprachigen Archäologie." In *Theorie in der Archäologie: Zur englischsprachigen Diskussion*, edited by M. K H Eggert, and U. Veit, 115-81. Münster: Waxmann.
- Lucas, G. 2001. "Destruction and the Rhetoric of Excavation." *Norwegian Archaeological Review* 34 (1): 35-46.
- Lynch, C. 2008. "How do your data grow?" *Nature* 455 (7209): 28-29.
- Lyne, J. 1993. "Angels in the Architecture: A Burkean Inventional Perspective on "Spandrels." In *Understanding*

Reflections on the Rocky Road to E-Archaeology
Geoff Carver and Matthias Lang

- Scientific Prose*, edited by Jack Selzer, 144-57. Madison: University of Wisconsin Press.
- Lévi-Strauss, C. 1966. *The Savage Mind*. Chicago: University of Chicago Press.
- Lönn, M. 2012. "An Archaeology of many steps." In *Reconsidering Fieldwork: Exploring On Site Relationships Between Theory and Practice*, edited by P. Richardson, H. L. Cobb, O. J. T. Harris, and C. Jones, 67-84. New York: Springer.
- Mackie, Q. 1995. "Prehistory in a multicultural state: a commentary on the development of Canadian archaeology." In *Theory in archaeology: a world perspective*, edited by P. J. Ucko, 178-96. London: Routledge.
- May, K., C. Binding, and D. Tudhope. 2011. "A STAR is Born: Some Emerging Semantic Technologies for Archaeological Resources." In *On the Road to Reconstructing the Past. CAA 2008: Computer Applications and Quantitative Methods in Archaeology. Proceedings of the 36th Conference, Budapest, April 2-6, 2008*, edited by E. Jerem, F. Redö, and V. Szeverényi, 111-16. Budapest: Archaeolingua.
- May, S. 2012. "Commentary." In *Reconsidering Fieldwork: Exploring On Site Relationships Between Theory and Practice*, edited by P. Richardson, H. L. Cobb, O. J. T. Harris, and C. Jones, 167-80. New York: Springer.
- McAdam, E. 1992. "Discussion Session: Feedback Mechanisms from Post-Excavation to Excavation." In *Interpretation of Stratigraphy: a Review of the Art; Proceedings of a Conference held on 18th June, 1992 at the City of Lincoln Archaeology Unit*, edited by K. Steane. Lincoln: City of Lincoln Archaeology Unit.
- McAnany, P. A., and I. Hodder. 2009. "Thinking about stratigraphic sequence in social terms." *Archaeological Dialogues* 16 (1): 1-22.
- McLuhan, M. 2002. *The Mechanical Bride: Folklore of Industrial Man*. Corte Madera: Gingko Press.
- McLuhan, M. 2003. "Cybernetics and Human Culture (1964)." In *Understanding Me: Lectures and Interviews*, edited by S. McLuhan, and D. Staines, 57-75. Cambridge: MIT Press.
- Meltzer, D. J. 1979. "Paradigms and the Nature of Change in American Archaeology." *American Antiquity* 44 (4): 644-57.
- Members of the Ardnamurchan Transitions Project. 2012. "The Struggle Within: Challenging the Subject/Object Relationship on a Shoestring." In *Reconsidering Fieldwork: Exploring On Site Relationships Between Theory and Practice*, edited by P. Richardson, H. L. Cobb, O. J. T. Harris, and C. Jones, 113-30. New York: Springer.
- Moreno Escobar, M. del Carmen. 2013. "ARCA: creating and integrating archaeological databases." Paper presented at *Computer Applications and Quantitative Methods in Archaeology (CAA), Southampton, March 26-29, 2012*.
- Murray, J., S. Roskams, and C. Spence. 1994. "Recording deposits." In *Archaeological Site Manual. 3rd ed.*, edited by Craig Spence, 3.1.2. London: Museum of London Archaeological Service.
- Niemeyer, H. G. 1968. *Einführung in die Archäologie*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Petrie, W. M. Flinders. 1904. *Methods and Aims in Archaeology*. London: Macmillan and Co.
- Pitt-Rivers, A. Lane-Fox. 1892. *Excavations in Cranborne Chase, near Rushmore, on the borders of Dorset and Wilts., Vol. III: Excavations in Bokerly and Wansdyke, Dorset and Wilts. 1888-1891*. London: Privately printed [Harrison and Sons].
- Planck, D. 1999. "Archäologische Ausgrabungen und Prospektionen: Durchführung und Dokumentation." *Archäologisches Nachrichtenblatt* 4 (1): 12-45.
- Quinton, A. 1980. *Francis Bacon*. Oxford: Oxford University Press.
- Rabinowitz, A., and L. Sedikova. 2011. "On Whose Authority? Interpretation, Narrative, and Fragmentation in Digital Publishing." In *On the Road to Reconstructing the Past. CAA 2008: Computer Applications and Quantitative Methods in Archaeology. Proceedings of the 36th Conference, Budapest, April 2-6, 2008*, edited by E. Jerem, F. Redö, and V. Szeverényi, 134-40. Budapest: Archaeolingua.
- Richards, J. D., J. Winters, and M. Charno. 2011. "Making the LEAP: Linking Electronic Archives and Publications."

- In *On the Road to Reconstructing the Past. CAA 2008: Computer Applications and Quantitative Methods in Archaeology. Proceedings of the 36th Conference, Budapest, April 2-6, 2008*, edited by E. Jerem, F. Redö, and V. Szeverényi, 141-46. Budapest: Archaeolingua.
- ROB. 1998. *Handboek van ROB-specificaties*. Amersfoort: Rijksdienst voor het Oudheidkundig Bodemonderzoek.
- Sabloff, J. A., L. R. Binford, and P. A. McAnany. 1987. "Understanding the Archaeological Record." *Antiquity* 61 (232): 203-9.
- Shanks, M., and C. Y. Tilley. 1987. *Re-constructing Archaeology: Theory and Practice. New Studies in Archaeology*. Cambridge: Cambridge University Press.
- Snow, D. R., M. Gahegan, C. L. Giles, K. G. Hirth, G. R. Milner, P. Mitra, and J. Z. Wang. 2006. "Cybertools and Archaeology." *Science* 311 210: 958-59.
- Spence, C. 1993. "Recording the archaeology of London: the development and implementation of the DUA recording system." In *Practices of Archaeological Stratigraphy*, edited by E. C. Harris, M. R. III Brown, and G. J. Brown, 23-46. London: Academic Press.
- Stephenson, N. 2011. "Atoms of Cognition: Metaphysics in the Royal Society, 1715-2010." In *Seeing Further: The Story of Science & the Royal Society*, edited by Bill Bryson, 84-105. London: HarperPress.
- Stevanovic, M. 2000. "Visualizing and Vocalizing the Archaeological Archival Record: Narrative vs Image." *Towards reflexive method in archaeology: the example at Çatalhöyük*, edited by I. Hodder, 235-38. Cambridge: McDonald Institute for Archaeological Research.
- Taylor, W. W. 1948. *A Study of Archeology.*, Vol. 50 No. 3, Part 2. Memoir Series of the American Anthropological Association.
- Trigger, B. G. 1989. *A History of Archaeological Thought. 1st ed.* Cambridge: Cambridge University Press.
- Trigger, B. G. 2006. *A History of Archaeological Thought. 2nd ed.* Cambridge: Cambridge University Press.
- Tucker, J. 2005. *Nature Exposed: Photography as Eyewitness in Victorian Science*. Baltimore: Johns Hopkins University Press.
- Van Riper, A. Bowdoin. 1993. *Men Among the Mammoths: Victorian Science and the Discovery of Human Prehistory*. Chicago: University of Chicago Press.
- Varytimiadis, S., A. Kakargias, and E. Stergiopoulou. 2011. "The D.E.A.D. Project: An Application for the Digital Recording of Archaeological Data." In *On the Road to Reconstructing the Past. CAA 2008: Computer Applications and Quantitative Methods in Archaeology. Proceedings of the 36th Conference, Budapest, April 2-6, 2008*, edited by E. Jerem, F. Redö, and V. Szeverényi, 147-51. Budapest: Archaeolingua.
- Veit, U. 1998. "Zwischen Tradition und Revolution: Theoretische Ansätze in der britischen Archäologie." *Theorie in der Archäologie: Zur englischsprachigen Diskussion*, edited by M. K. H. Eggert, and U. Veit, 15-65. Münster: Waxmann.
- Verhagen, P., C. Sueur, and M. Wansleben. 2011. "Setting a Standard for the Exchange of Archaeological Data in the Netherlands." In *On the Road to Reconstructing the Past. CAA 2008: Computer Applications and Quantitative Methods in Archaeology. Proceedings of the 36th Conference, Budapest, April 2-6, 2008*, edited by E. Jerem, F. Redö, and V. Szeverényi, 152-55. Budapest: Archaeolingua.
- VLA. 2006. *Archäologische Ausgrabungen und Prospektionen: Durchführung und Dokumentation*. Version 03.04.06., revised edition. Verband der Landesarchäologen in der Bundesrepublik Deutschland.
- Wordsworth, W. 1994. "Roman Antiquities discovered at Bishopstone, Herefordshire." In *The Collected Poems of William Wordsworth.*, 275. Ware: Wordsworth.
- Yamin, R. 2012. "Through Many Eyes: A Non-hierarchical Approach to Interpreting a Site in New Brunswick, New Jersey." In *Reconsidering Fieldwork: Exploring On Site Relationships Between Theory and Practice*, edited by P. Richardson, H. L. Cobb, O. J. T. Harris and C. Jones, 131-46. New York: Springer.

Least-cost Networks

Irmela Herzog

The Rhineland Commission for Archaeological Monuments and Sites, Bonn

Abstract:

Many archaeologists applying GIS software for least-cost studies do not fully understand the functions employed to create least-cost networks, and so they are not aware of the prerequisites and drawbacks of the model implemented by the software. This paper discusses several models for connecting dots on the map resulting in least-cost networks: all pair shortest paths, least-cost network to the builder, network connecting the nearest neighbours, triangulation network, minimum spanning tree, central point network, and main travel route with subsidiary paths. Most of the models are applied to a study region in the Bergisches Land, Germany. This region provides less attractive natural conditions than the area close to the Rhine, and for this reason, only very few settlements existed in this region prior to the year 1000 AD. Detailed historical sources and old maps are available to check whether the models agree with the historical reality.

Keywords:

Least-Cost Paths, Networks, Triangulation, Minimum Spanning Tree, Steiner Tree

1. Introduction

Many archaeological least-cost path (LCP) studies deal with networks, i.e. the aim is to reconstruct a network of routes (e.g. van Leusen 2002, 16·12–16·18; Bell et al. 2002). This paper discusses a set of models each minimizing costs in some sense but resulting in a different network topology. Unfortunately, most push-button GIS software packages do not support different LCP network models, and hardly any information can be found in the documentation of the software concerning the models and algorithms used for calculating least-cost networks. Consequently, most archaeologists applying this software are not aware that different models exist and that the software might not be appropriate for the model they have in mind.

The first part of this paper deals with models where all nodes of the network are known. In this case, the aim is to reconstruct links between these nodes so that the network costs are minimized according to some criteria. The second part introduces methods for reconstructing paths when only some important locations are given or a main travel route is known.

A hilly study area east of Cologne, Germany, serves as a test landscape. This area was sparsely populated before Medieval times because the other regions of the Rhineland offered more attractive natural conditions for farming. From the 9th century onward, nearly all the fertile regions of the Rhineland were occupied by farms, and at the same time the climatic conditions for agriculture improved, so that the population increased in the study area (Landschaftsverband Westfalen Lippe, and Landschaftsverband Rheinland 2007, on CD: 282). Therefore, historical data can be used to test the network models. For several parishes in the study area, a list of place names and the year when they were first mentioned was published (Pampus 1998). Nearly all of the place names could be located on maps created in the late 19th century, and this allowed georeferencing the centres of the settlements, most of which were still very small at the end of the 19th century. The parishes with listed place names cover about 448 km² of the study area, and within this area, 513 settlements were georeferenced that were first mentioned before 1500 AD. However, only 15 of these place names were mentioned in documents issued before 1150 AD.

In addition, a publication on Medieval trade routes is used in this study (Nicke 2001). The trade routes described by Nicke were tracked on historical maps from about 1845. Unfortunately, older maps do not cover the study area in total. So the trade routes shown in Fig. 1 rely on the assumption that the Medieval route layout is well preserved on the

Corresponding author: i.herzog@lvr.de

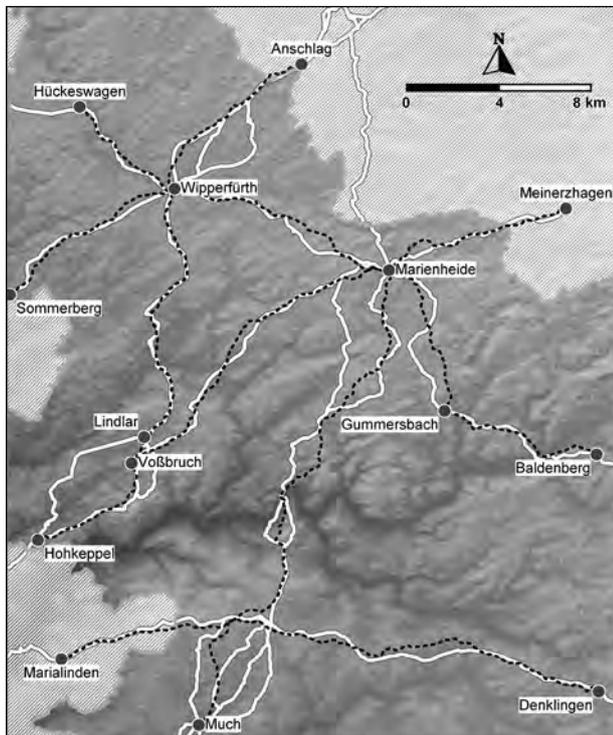


Figure 1. Hilly study region east of Cologne, Germany. The background map is the ASTER DEM with a resolution of about 30 m, ASTER GDEM is a product of METI and NASA. White lines indicate the ancient trade routes, sometimes with alternative route sections. The black dotted lines are the LCPs agreeing best with the known trade routes. For the areas covered by a white pattern, Pampus (1998) does not provide any information on place names and the year when they were first mentioned.

maps from about 1845. For each trade route two or more locations on the route were selected and LCPs were calculated with different cost components and cost weights. LCPs avoiding wet soils and with a critical slope of 13 % fitted best to the trade routes. Slope dependent costs were modelled by a quadratic function (Herzog *in press*) and for wet soils a multiplier of 5 was used. The reconstructed paths and the trade routes do not coincide perfectly, and this could be due to (i) inaccuracies and resolution of the elevation model (ASTER data provided by NASA and METI), (ii) natural and man-made landscape modifications (e.g. quarries, meandering of rivers, dam construction for water reservoirs), (iii) additional cost components not considered here or (iv) inaccuracies in locating the trade routes. However, the cost model found is quite appropriate for most trade routes in the study area, and so it is

used for all least-cost calculations presented in this paper.

The direction of the paths is not considered in this study, it is assumed that the same route is taken on the way from A to B and back. This requires an isotropic cost function, which can be generated easily by averaging the costs of movement in both directions (Herzog *in print*). This isotropic cost function is applied to effective slope derived from the digital elevation model as described by Yu et al. (2003). The software used for this paper is part of a set of programs for least-cost calculations created by the author. Least-cost path concepts implemented in the software were already presented in more detail elsewhere (Herzog *in print*; Herzog 2013). The programs generate MapInfo Interchange Format (MIF) files that can be imported easily into MapInfo 8.0. The GIS MapInfo 8.0 with the plugin Vertical Mapper was used to create the maps in this contribution. In future it is planned to improve the user-interface of the software so that it can be made available to other people interested in archaeological least-cost analysis.

2. All Pair Shortest Paths

Within an all pair shortest paths network, each pair of contemporary sites in a region is connected by an LCP. The multitude of paths created by Indians in California may serve as an example. Earle (1991, 11) refers to a historic report describing this network layout, which most likely was a result of the fact that hardly any effort was needed to create a path in fairly flat terrain, and so, whenever desirable a new path was established. On the same page he presents ethnographic evidence that in hunter and gatherer societies, paths are ephemeral, redundant and without centrality if the terrain is fairly flat.

The term *least-cost networks to the user* is used by Waugh (2000, 615) for this network type, because the traveller can move as quickly and directly as possible between any pair of sites in the network. According to this model, whenever a new site was founded within the region considered, this site established shortest path connections to all other sites that existed previously. The number of connections required has a quadratic growth rate, and this is one of the reasons why in general such networks consist of a limited number of sites only.

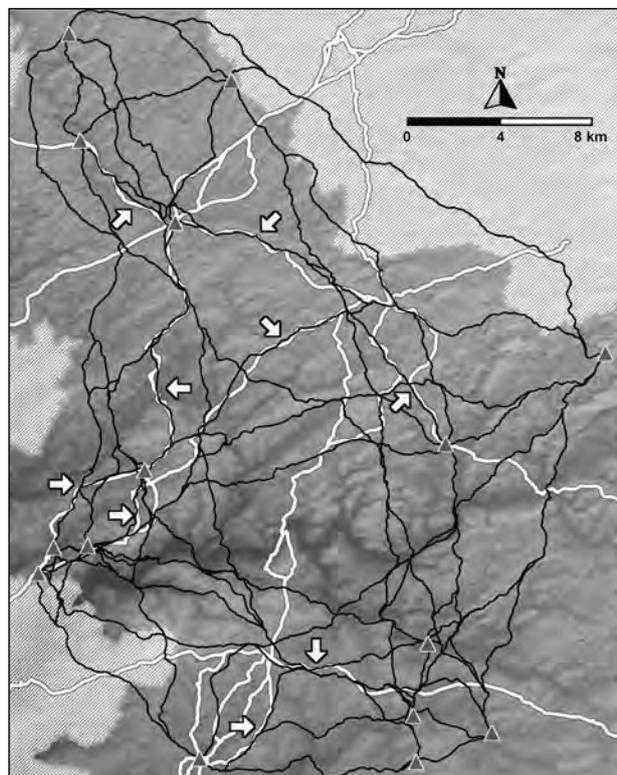


Figure 2. All pair shortest paths network for the 15 settlements in the study area first mentioned before 1150 AD. The settlements are marked by triangles. The ancient trade routes are depicted as in Fig. 1, the black lines form the all pair shortest paths network consisting of 106 LCPs. Nine arrows indicate the path segments where the LCPs run along the trade routes.

Such networks are highly redundant and without centrality. The relative chronology of the paths cannot be deduced from the path layout. Effective algorithms solving this all-pairs shortest paths problem have been published (Cormen et al. 2001, 620–42).

Due to the quadratic growth rate of the number of connections, only the all pair shortest paths network for the 15 settlements mentioned before 1150 AD was generated. This network consists of 106 LCPs, and nine path segments of this network coincide with the ancient trade routes (Fig. 2).

3. Least-cost Network to the Builder

A society living on crops grown in fields near the villages will not tolerate innumerable paths crossing these fields. Instead the route network should cover

only a minimum amount of the valuable land used for agrarian purposes. Similarly, if some effort is required for creating paths, connecting all point pairs including those at large distances by LCPs is not effective.

A *least-cost network to the builder* (Waugh 2000, 615) connects all sites but minimizes the total costs of route construction. With respect to modern geography, Waugh notes that such networks are found in areas with sparse population and in places where road construction costs are high. In such a network, long detours are often necessary to reach a neighbouring location.

A configuration with extra intermediate points is optimal in terms of total path costs. Such intermediate points are known as Steiner point. In a landscape with constant friction, *Steiner trees* (Smith et al. 2007, 341) are optimal networks to the builder; however, calculation complexity is high. Due to this fact an implementation of this network layout was not attempted, instead readers are referred to the work of Verhagen et al. (2011) on this topic. All site locations must be known for creating this network. Consequently, a chronological development of this network is not part of the model.

For a given set of site locations, the sum of costs for all paths in the least-cost network to the user may be regarded as an upper bound of the total path costs in the network connecting the sites, whereas the total cost of paths in the least-cost network to the builder serves as a lower bound. Depending on the landscape, the technological skills and the travelling requirements of the society considered, an ancient route network can be ranked somewhere in the range between these extremes. Models for networks that are a compromise between the two extremes are presented in the next sections.

4. LCPs to the Nearest Neighbours

The model underlying proximal point analysis is based on the assumption that establishing a route to the neighbouring village is more important than the connection to a settlement at a larger distance.

According to Terrel and Welsch (1997), Terrell introduced a simple graph-theoretic technique to model likely interaction spheres, a technique

he called proximal point analysis in 1974. This technique was applied by Spennemann (2003) in his attempt to reconstruct the road network on the island of Tongatapu, Kingdom of Tonga. The resulting network connects each settlement to its three nearest neighbours by straight lines.

White (2012) extends this approach by using LCPs instead of straight-line connections and by considering a varying number of neighbours. After analysing the results he comes to the conclusion that up to five neighbours are appropriate for his study area. A disadvantage of this approach is that the resulting network may consist of several connected components, especially if sites are clustered. The examples presented by White exhibit this effect, although the connected components are not necessarily separated by large distances. Within a connected component of a network taking the n nearest neighbours into account, long detours are sometimes necessary to reach a close neighbour with rank $n+1$. These drawbacks are avoided by triangulation networks.

5. Least-cost Triangulation Network

Least-cost triangulation networks connect each site with its immediate neighbours by LCPs that form a triangulation. A triangulation LCP network is based on a heuristic approach with the aim of ensuring good connection to neighbouring settlements in all directions; settlements at a distance can only be reached via intermediate stops at neighbouring settlements. With straight-line distances, there are several candidate triangulations for a given set of points. A Delaunay triangulation avoids long thin triangles and acute angles and therefore is preferred in many applications (e.g. Worboys and Duckham 2004, 190–92; Smith 2007, 113–15).

Most GIS software packages support the creation of a straight-line Delaunay triangulation for a given set of irregularly spaced points. On average, for each location the number of neighbours found by the Delaunay triangulation is six. In network analysis, the nodal degree is the number of paths that are connected to a particular seed location in the network (Conolly and Lake 2006, 241). The nodal degree is often considered as a measure of importance or accessibility of a location

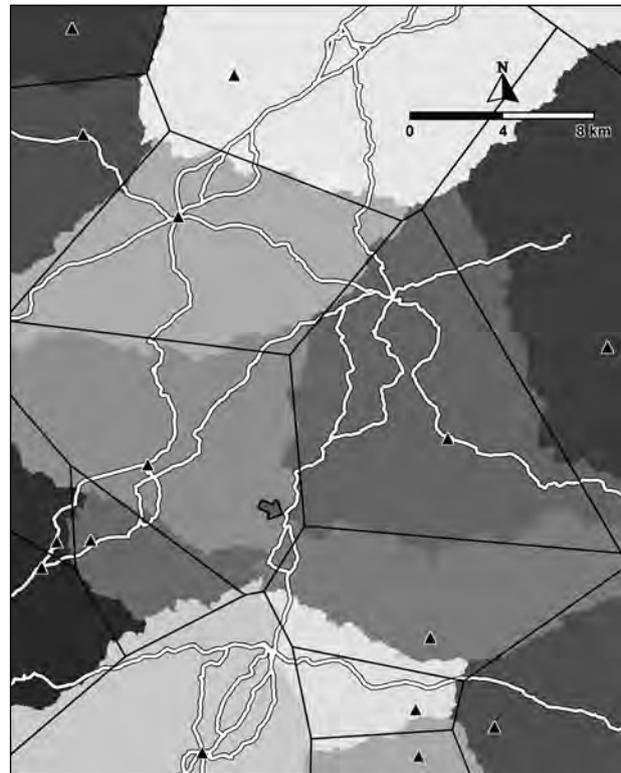


Figure 3. Comparison of least-cost Voronoi patches and straight-line Voronoi polygons for the 15 settlements in the study area first mentioned before 1150 AD. The arrow points to a section of the ancient road Zeitstraße that coincides fairly well with the boundaries between the least-cost patches.

(e.g. Rodrigue et al. 2009, 29; Llobera et al. 2011). Consequently, the LCP triangulation network model is appropriate only for seed locations of similar importance.

If most interaction takes place between neighbours of roughly the same importance, this network probably is an appropriate compromise between the two extremes – least-cost network to the builder and to the user. Such networks are with hardly any centrality. In a straight-line Delaunay triangulation, adding a new point can change many triangles in the network; however, when adding a new site to an existing road network, such a radical change is highly unlikely.

The following algorithm was used to create least-cost triangulation networks for a given set of seed points:

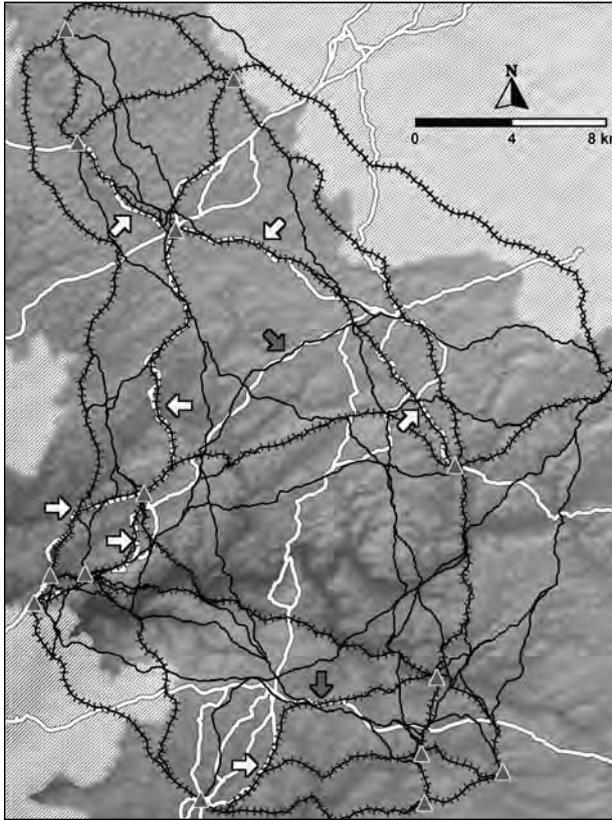


Figure 4. Same as Fig. 2, in addition the least-cost triangulation network for the 15 settlements is marked by small (blue) bars perpendicular to the route. The colour of two of the arrows shown in Fig. 2 is changed to dark grey because the corresponding path segments are not part of the least-cost triangulation network.

- Calculate least-cost buffers for the seed cells.
- Assign to each cell the number of the least-cost nearest seed cell, resulting in least-cost Voronoi patches (Fig. 3).
- Neighbours are defined: The seeds of two Voronoi patches are neighbours if the common boundary exceeds a predefined length.
- The quick and dirty least-cost triangulation connects neighbours via LCPs that traverse the common boundary (see below).

The basis of this algorithm are two facts: (1) A cumulative cost surface with several seed points allows to identify maximum cost boundaries that delimit least-cost Voronoi patches (van Leusen

2002, 16·17–16·18; Smith et al. 2007, 116); (2) when considering straight-line distances, a Delaunay triangulation results from connecting all centres of neighbouring Voronoi patches (e.g. Worboys and Duckham 2004, 191, 206). Consequently, a least-cost triangulation can be derived from least-cost Voronoi patches. If the triangulation includes LCPs that are very long and costly in comparison to the average LCPs in the triangulation, it might be useful to delete these LCPs after choosing an appropriate threshold, e.g. by outlier detection (Smith et al. 2007, 181–83).

Figure 3 shows the least-cost Voronoi patches for the 15 settlements mentioned before 1150 AD. The standard Voronoi tessellation is indicated by black straight lines. The two Voronoi diagrams differ in some places so that the adjacent neighbours of a site's territory do not always agree. A section of the Zeitstraße, which could not be reconstructed successfully by the cost model applied in this paper, agrees well with the boundary of a Voronoi patch. Several publications note that Medieval routes sometimes coincide with a boundary (e.g. Gissel 1986). Although the first reconstruction attempt for the section of the Zeitstraße was not successful using the cost model, applying the cost model for generating least-cost Voronoi boundaries results in a fairly good match.

The least-cost triangulation network for the 15 settlements mentioned before 1150 AD consists of 29 LCPs (Fig. 4). Although the number of links is considerably lower than that of the all pair shortest path network, seven out of the nine path segments coinciding with the ancient route are still part of the reconstructed network.

To speed up calculations, a quick and dirty triangulation was implemented that forces the connecting path between two neighbours across the common border. This results in the correct LCP in most but not in all cases. A proper implementation of the least-cost triangulation is planned in near future. Figure 5 shows the result of this quick triangulation for the 34 settlements in the study area mentioned before 1250 AD. Although there are more paths than in Figure 4, the number of segments where ancient trade routes and triangulation paths coincide does not increase.

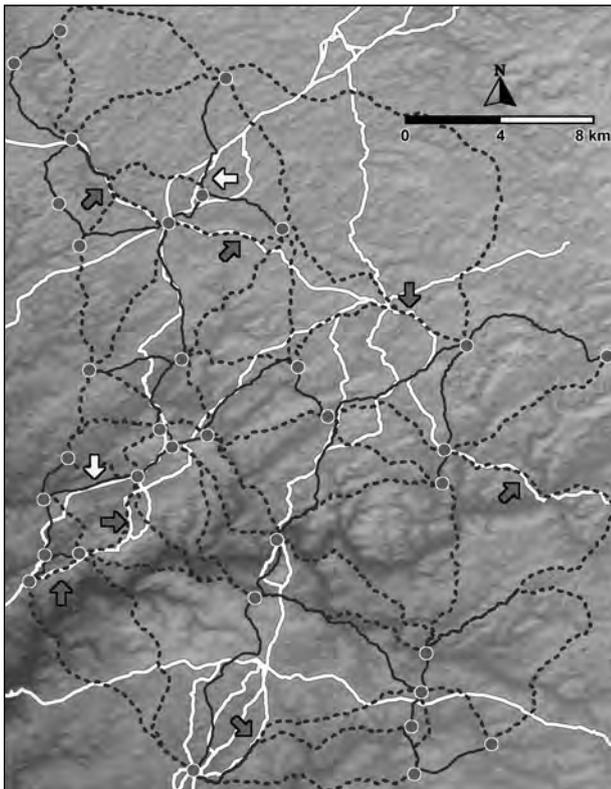


Figure 5. Dot symbols indicate 34 settlements mentioned before 1250 AD. The 33 LCPs forming the least-cost MST are indicated by solid lines. The quick least-cost triangulation consists both of the solid and dotted (blue) lines. Dark grey arrows mark the path segments where the least-cost triangulation and the trade routes agree quite well. Light (yellow) arrows mark those LCP segments that form the least-cost MST and coincide with a trade route.

6. Minimum Spanning Tree

A minimum spanning tree (MST) consists of edges that are part of the Delaunay triangulation (Cormen et al. 2001, 561–79; Smith et al. 2007, 339). Only $n-1$ connecting lines are in the MST if n is the number of seed locations, and therefore the total costs of all LCPs forming the network are significantly lower than in a least-cost triangulation network. Prim's algorithm (Cormen et al. 2001, 570–73) can be applied to identify the least-cost MST in the least-cost triangulation.

The total costs of all LCPs forming the network often exceeds that of the least-cost network to the builder. However, no additional Steiner points

need to be constructed, and so generating a least-cost MST is easier. The Steiner ratio can be used to compare the efficiency of the MST with the optimal Steiner tree (Ganley 2004). This key figure is the largest possible ratio between the total length of a MST and the total length of the optimal Steiner tree. With straight-line distances, the ratio is conjectured to be about 1.15. So the total length of the worst-case MST exceeds that of the optimal Steiner tree by 15%.

Due to the low number of connections, long detours are often necessary in a MST to reach a neighbouring location. Moreover, if a link is broken, network connectivity is lost. Smith et al. (2007, 339) point out that real-world networks normally are implemented with a higher level of connectivity to avoid this problem.

The solid (blue) lines in Figure 5 form the least-cost MST for the 34 settlements in the study area mentioned before 1250 AD. The calculation is based on the quick triangulation. Only two segments of the least-cost MST coincide with the known trade routes.

Spennemann (2003) also applies two variants of the MST that take the importance of the nodes into consideration. In his example, the proxy for importance is the frequency of mentioning a settlement in historic sources. Spennemann uses the terms weighted MST and linked weighted MST in this context, but does not provide any references or a description of the methods. The results shown in the publication indicate that the number of edges of the weighted MST is higher than that of the MST, and with the linked weighted MST still more edges are added. In the example presented by Spennemann, the nodal degree of important sites within the linked weighted MST is higher than that of less important sites. This result indicates that the linked weighted MST approach justifies further investigation. Moreover, in many archaeological least-cost networks, the sites considered differ with respect to their importance, and only very few approaches are known that include importance into the calculation.

7. Least-cost Basin-Clustering

Rodrigue et al. (2009, 18–19) describe hub-and-spoke networks, where the hub serves as the

centre of efficient distribution. Least-cost basin clustering can be applied to identify the hubs and the distribution links in such hub-and-spoke networks. Basin clustering was proposed by Hader and Hamprecht (2003), and a few years ago this approach was applied to Early Iron Age settlements and cemeteries from the Main triangle, Germany (Herzog 2009). At that time, only a straight-line version of the algorithm was implemented by the author, but the need was felt to account for the travel costs in the hilly study area. Finally, after long research into least-cost methods, first results of the least-cost version are shown in the present paper (Fig. 6).

The following algorithm was used to create the least-cost basin clustering for a given set of seed points:

- Calculate the least-cost Kernel Density Estimation (KDE) to derive a density value for each site (Herzog and Yépez in print). The user can control the impact of the sites by choosing the bandwidth parameter, corresponding to the cost distance where the impact of the site ceases. The scale of the clustering depends on the size of the bandwidth parameter. It is possible to refine standard least-cost KDE so that the importance of a site is reflected by the height of the kernel.
- Generate a least-cost triangulation to identify the relevant neighbours of each site.
- Establish for each site a single link to a neighbour with higher density. Hader and Hamprecht (2003) propose choosing the neighbour with highest gain, i.e. the neighbour with a higher density where the density difference versus path costs ratio is highest.
- Determine the clusters, i.e. linked groups of sites
- Identify the hubs, i.e. the site of maximum density within each cluster. These are considered as cluster centres.

Figure 6 shows the result of a least-cost basin clustering based on a quick triangulation for the 34 settlements in the study area mentioned before 1250 AD. The basin clustering network consists of six clusters and 27 LCPs. Four segments of these LCPs

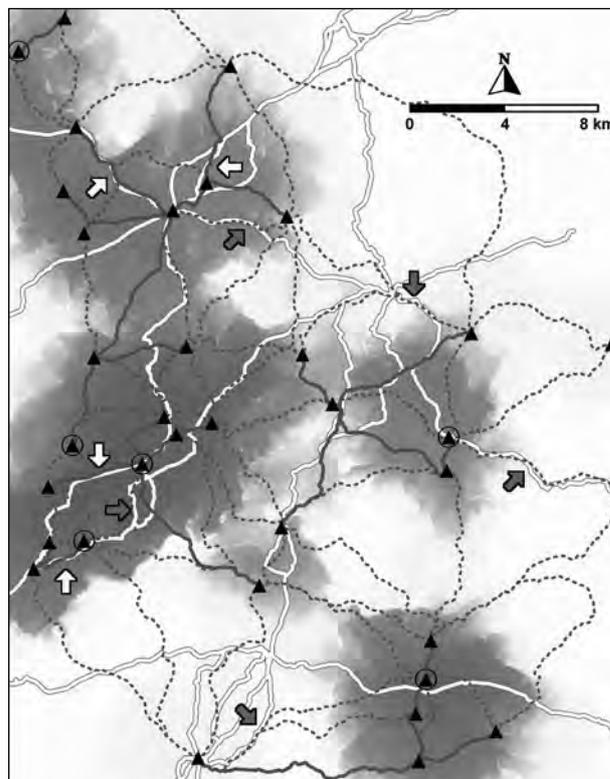


Figure 6. Black triangles indicate settlements mentioned before 1250 AD. The quick least-cost triangulation consists of the solid and dotted (green) lines. The basin clustering of the settlements is based on the least-cost KDE with a spread parameter of 8 km for level dry ground. The least-cost KDE distribution is shown in the background. The cluster centres are marked by black rings. The LCPs forming the basin clustering are indicated by solid lines. Dark grey arrows mark the path segments where the least-cost triangulation and the trade routes agree quite well. Four light (yellow) arrows mark those LCP segments that are part of the basin clustering and coincide with a trade route.

coincide quite well with the ancient trade routes. Two of the cluster centres were probably important central places in Medieval times (Lindlar and Gummersbach). But the settlement in the study area that first received town privileges (Wipperfürth) is not a cluster centre. So the result is far from satisfactory with respect to the cluster centres. This may be due to one of the following reasons: (1) An inappropriate bandwidth parameter was chosen; (2) The model is not appropriate for the Medieval settlements of the study area; (3) The outcomes of this approach suffer from the edge effect because important sites at the border of the study area are

not recognized by the method. However, the latter could be fixed by assigning a higher weight to these sites by increasing the height of the kernel.

8. Central Point Networks

Central point networks spread from a single point, thus exhibiting a high degree of centrality. Van Leusen (2002, 16-12) modelled such a central point network by selecting appropriate end locations and calculating multiple LCPs connecting the central point Wroxeter with these end locations. However, if probable end locations are not known, it is still possible to model central point networks. A method for generating such central point networks was proposed by Fábrega Álvarez and Parceró Oubiña (2007). They determine probable paths from an origin outward on the basis of the cost-surface: Destination points of these paths are the locations with maximum distance versus cost ratio, i.e. where large distances can be covered at low costs. This procedure is carried out for several origins, and the authors check if a connected route network results.

Another approach to model central point networks was published by Llobera together with Fábrega Álvarez and Parceró Oubiña (2011). A drain procedure is applied to the accumulated cost surface to identify potential paths to the origin. However, the drawbacks of a drain approach are well-known (Smith et al. 2007, 145-46), and it is for this reason that in my view, the initial method of Fábrega Álvarez and Parceró Oubiña is to be preferred.

A historic example of a central point network is given by Hindle (2002, 31-35). He analyses the Gough map of about 1360 covering most of England. He comes to the conclusion that the map reflects the centralisation of the government; a road system radiating from London is evident, despite some inexplicable omissions.

Based on the network structure, it does not seem possible to reconstruct the chronological sequence of the routes in such a network.

The following algorithm based on the ideas of Fábrega Álvarez and Parceró Oubiña (2007) is proposed to create a central point network for a given seed point:

- Create the least-cost buffer for the seed point with a user-selected radius (i.e. maximum cost value).
- Identify the location with maximum straight-line distance on the border of the buffer.
- The LCP to this location is the first path in the central point network.
- Successively identify the next best locations on the border of the buffer. Make sure that distinct paths are chosen by keeping a minimum straight-line distance to previously selected paths and by ensuring that the LCPs chosen represent local optima with respect to the neighbouring cells on the border.

Progression on the first paths selected in the process described above is best, and for this reason it seems more likely that these paths correspond to ancient routes than the other paths in the reconstructed network. However, this applies only if the first paths selected cover several directions. An example for an application of the central point network approach given above is depicted in Figure 7. The centre of the network is Wipperfürth, the oldest town in the study area. Three arrows indicate sections where the reconstructed paths agree well with the known trade routes. These sections are part of the three paths selected first. Both Wipperfürth and Wiehl were first mentioned in 1131. But only one section of the central point network for Wiehl consisting of ten LCPs coincides with an ancient trade route (not shown).

9. Dendritic Networks

The central feature of a network could also be a line rather than a point, e.g. a river or a main travel route. The least-cost subsidiary paths together with the central line form a dendritic network. The algorithm proposed for generating central point networks can be readily extended for creating the subsidiary paths of a given main travel route. Instead of one seed point, all raster cells traversed by the main travel route form the seeds of the least-cost buffer. So this model assumes that two types of roads existed at a certain period of time, roads of different importance. Several road classes are known from Roman times (Heinz 2003, 22-26): main

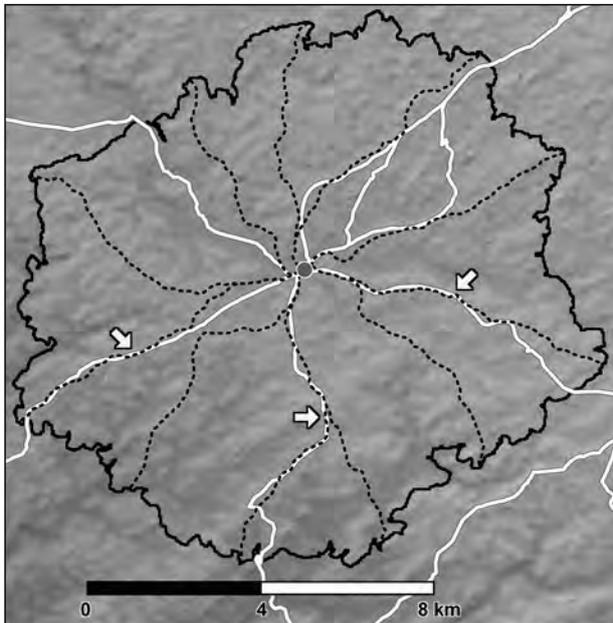


Figure 7. Central point network (black dotted lines) consisting of 11 paths for the oldest town (Wipperfürth) in the study area. The cost buffer boundary depicted by a black solid line delimits all locations that can be reached by an effort less or equal to covering 9 km on level dry ground. The minimum straight-line distance between any of the reconstructed paths on this boundary was set to 3 km. White lines indicate the known trade routes, sometimes with alternative route sections. Three arrows mark the path segments where the reconstructed network and the trade routes agree quite well.

routes (viae publicae), local routes (viae vicinales) as well as private and military roads.

Bell et al. (2002) describe a central line network in Italy: The Sangro river valley probably was the principal long-distance artery, supplemented by local communications between sites in the same general area of the valley. However, the focus of the study by Bell et al. is only on the local communication routes, and these were reconstructed by generating LCPs connecting sites of a given period.

It is quite obvious that the central line must have existed before the subsidiary paths were created, however, the chronological sequence of the subsidiary paths cannot be deduced from the path layout. As with the central point network, progression on the first paths selected is best. If progression ease was the only criteria for creating subsidiary paths in the past, then we could assume

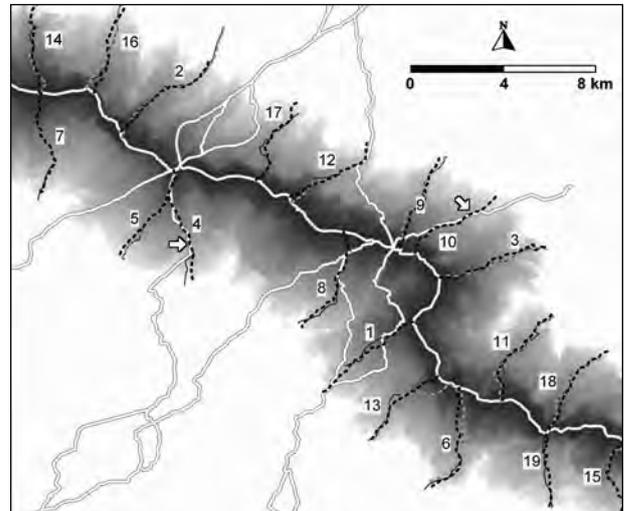


Figure 8. Reconstructed subsidiary paths (black dotted lines) for the old trade route Eisenstraße. White lines indicate the known trade routes including the Eisenstraße. The least-cost buffer for the Eisenstraße with a limit corresponding to covering 6 km on dry level ground is shown by decreasing colour saturation. The numbers indicate the sequence of selecting the paths, i.e. the reconstructed subsidiary path no. 1 covers the longest distance. Two arrows mark the path sections where the reconstructed network and the trade routes agree quite well. Light (orange) dotted lines are digitised paths from the historic map issued in 1840, dark (blue) solid lines indicate roads on the 1890 map. Only paths coinciding roughly with the reconstructed paths were digitised.

that the longest paths were created first. However, in the example presented in Figure 8 the longest reconstructed subsidiary path does not agree with an ancient trade route or any other path on a 19th century map. Out of the 19 reconstructed paths, paths number 4 and 10 agree in some sections with the known trade routes. Nearly all reconstructed paths including path no. 19 coincide with roads recorded on the 19th century maps.

10. Conclusions and Future Work

Many factors influence the outcome of LCP calculations besides the cost model, including resolution and quality of the topographic data, landscape changes since the period considered, as well as the least-cost algorithm and implementation. When generating least-cost networks the situation is even more complicated, because additional issues should be considered:

Method	No of nodes	No of paths	Coinciding route sections	Depicted in Fig.
All pair shortest paths network	15	106	9	2
Least-cost triangulation network	15	29	7	4
Least-cost triangulation network	34	81	9	5
Least-cost minimum spanning tree	34	33	2	5
Least-cost basin clustering	34	27	4	6

Table 1. Least-cost network methods implemented by the author and results for the trade routes and Medieval settlements in the study area. The focus is on networks connecting a given set of dots. The ratio of the number of coinciding route sections versus the total number of paths in the network is highest for the triangulation network based on settlements mentioned before 1150 AD.

Did the network evolve, or were the paths constructed after contemporary sites had already been founded? Many studies and methods assume that roughly contemporary sites suddenly popped up out of nowhere. The question how the founders of the sites got to these places is hardly discussed at all. But often the previously existing road connections played an important role for the next period. An example is presented by Hindle (2002, 6, 31). According to Hindle, the Roman road system in England provided a basic network in Medieval times: On the Gough map of about 1360 covering most of England, almost 40 % of the routes shown are along the line of Roman roads. Least-cost Steiner-tree related methods suffer from the drawback that they do not answer the question concerning the evolution of the network. Triangulations are not quite as problematic although adding a dot may change many triangles in the network; however, this happens only rarely.

Are all the dots on the map of identical importance? During Medieval times, probably more paths led to a market town than to a single farmstead surrounded by fields. This example shows that the importance of a dot in terms of attracting people plays an important role. The triangulation approach assumes that all dots are of equal importance. According to my knowledge, the MST variants applied by Spennemann (2003) and basin clustering are the only methods taking the importance of the dots into account.

What is the purpose of the road network considered? The cost factors governing the layout of specific route types and the topology of the networks vary depending on the purpose of the routes. Several road classes are known from Roman times (Heinz 2003, 22–26), and White (2012) presents a classification of four path types

in desert areas, ranging from local paths connecting neighbouring settlements to interregional travel paths that often bypass settlement areas. Main routes may evolve from paths connecting neighbouring settlements but may also result from the need for a long-distance connection. Hindle (2002) discusses in some detail English Medieval roads with different purposes: church and corpse roads (pp. 10–12), pilgrim routes (pp. 12–14), monastic routes (p. 14), and drove roads (pp. 14–16). According to Hindle, drove roads tended to avoid villages.

How important was movement and how costly was it to create paths? If paths are used rarely or if the costs of path construction are high, network layouts similar to optimal networks to the builder are to be expected. However, sometimes a compromise towards a more user-friendly network is needed. Helbing et al. (1997) suggest a method for constructing networks in the range between builder- and user-friendliness.

Jiménez and Chapman (2002) connect locations by networks called Beta-skeletons. This concept includes both Gabriel networks (see also Smith et al. 2007, 339–341) and Relative Neighbourhood networks. The parameter beta allows the user to identify neighbours on different scales. If beta is set to 2, the resulting graph is a subset of the Delaunay triangulation and a superset of the MST. Implementing and testing this method on a least-cost basis is a future task.

Due to the large variety of factors influencing the outcome of least-cost networks calculations, it is necessary to include ground truthing and checks for equifinality in archaeological studies trying to reconstruct path networks of past times. The examples from the study area in the Bergisches

Land show that most of the models discussed in this paper lead to some good reconstruction results (Table 1). So a network model reproducing some of the ancient path sections may still not be the best choice. However, any least-cost network created without validation against archaeological reality is most probably not very helpful.

Acknowledgements

Special thanks are extended to Axel Posluschny, Andrew Bevan, Tom Brughmans, Karsten Lambers, Klaus Kleefeld, Jens Andresen, and Petra Dittmar for fruitful discussions and/or providing useful references.

References

- Bell, T., A. Wilson, and A. Wickham. 2002. "Tracking the Samnites: Landscape and Communications Routes in the Sangro Valley, Italy." *American Journal of Archaeology* 106:169–86.
- Conolly, J., and M. Lake. 2006. *Geographical Information Systems in Archaeology*. Cambridge: Cambridge University Press.
- Cormen, T.H., C.E. Leiserson, R.L. Rivest, and C. Stein. 2001. *Introduction to Algorithms. Second Edition*. Cambridge, London: The MIT Press.
- Earle, T. 1991. "Paths and road in evolutionary perspective." In *Ancient Road Networks and Settlement Hierarchies in the New World* edited by C. D. Throbbold, 10–16. Cambridge: Cambridge University Press.
- Fábrega Á.P., and C.P. Oubiña. 2007. "Proposals for an archaeological analysis of pathways and movement." *Archeologia e Calcolatori* 18: 121–40.
- Ganley, J.L. 2004. "Steiner ratio." In *Dictionary of Algorithms and Data Structures* edited by P. E. Black. U.S. National Institute of Standards and Technology. Accessed June, 2012. <http://xlinux.nist.gov/dads/HTML/steinerratio.html>.
- Gissel, S. 1986. "Verkehrsnetzänderungen und Wüstungserscheinungen im spätmittelalterlichen Dänemark." In *Siedlungsforschung. Archäologie – Geschichte – Geographie* 4, edited by K. Fehn, K. Brandt, D. Denecke, F. Irsigler, W. Janssen, W. Krings, J. Lüning, M. Müller-Wille, H.-J. Nitz, G. Overbeck and W. Schich., 63–80. Bonn: Verlag Siedlungsforschung.
- Hader, S., and F.A. Hamprecht. 2003. "Efficient Density Clustering Using Basin Spanning Trees." In *Between Data Science and Applied Data Analysis. Studies in Classification, Data Analysis, and Knowledge Organization*, edited by M. Schader, W. Gaul, and M. Vichi, 39–48. Berlin, Heidelberg, New York: Springer.
- Heinz, W. 2003. *Reisewege der Antike*. Stuttgart: Theiss Verlag.
- Helbing, D., J. Keltsch, and P. Molnár. 1997. "Modelling the evolution of human trail systems." *Nature* 388: 47–50.
- Herzog, I. 2009. "Analyse von Siedlungsterritorien auf der Basis mathematischer Modelle." In *Kulturraum und Territorialität: Archäologische Theorien, Methoden, Fallbeispiele. Kolloquium des DFG-SPP 1171 Esslingen 17.-18. Januar 2007*, edited by D. Krausse and O. Nakoinz, 71–86. Rahden/Westf.: Verlag Marie Leidorf.
- Herzog, I. 2013. "The Potential and Limits of Optimal Path Analysis." In *Computational Approaches to Archaeological Spaces*, edited by A. Bevan and M. Lake. Walnut Creek: Left Coast Press.
- Herzog, I. *In press*. "Theory and Practice of Cost Functions." In *Fusion of Cultures. Proceeding of the XXXVIII Conference on Computer Applications and Quantitative Methods in Archaeology, CAA 2010*, edited by F. Contreras, M. Farjas and F.J. Melero. Granada.
- Herzog, I., and A. Yépez. *In press*. "Least-Cost Kernel Density Estimation and Interpolation-Based Density Analysis Applied to Survey Data." In *Fusion of Cultures. Proceeding of the XXXVIII Conference on Computer Applications and Quantitative Methods in Archaeology, CAA 2010*, edited by F. Contreras, M. Farjas and F.J. Melero. Granada.
- Hindle, P. 2002. *Medieval Roads and Tracks*. Buckinghamshire: Shire Publications.
- Jiménez, D., and D. Chapman. 2002. "An application of proximity graphs in Archaeological spatial analysis." In *Contemporary Themes in Archaeological Computing*, edited by D. Wheatley, G. Earl and S. Poppy, 90–99. Oxford: University of Southampton Department of

Archaeology, Monograph 3.

Landschaftsverband Westfalen Lippe, and Landschaftsverband Rheinland, ed. 2007. *Erhaltende Kulturlandschaftsentwicklung in Nordrhein-Westfalen*. Münster, Köln (detailed version on CD).

Llobera, M., P. Fábrega-Álvarez, and C. Parcero-Oubiña. 2011. "Order in movement: a GIS approach to accessibility." *Journal of Archaeological Science* 38: 843–51.

Nicke, H. 2001. *Vergessene Wege. Das historische Fernwegenetz zwischen Rhein, Weser, Hellweg und Westerwald, seine Schutzanlagen und Knotenpunkte*. Nümbrecht: Martina Galunder-Verlag.

Pampus, K. 1998. *Urkundliche Erstnennungen oberbergischer Orte*. Gummersbach: Bergischer Geschichtsverein.

Rodrigue, J.-P., C. Comtois, and B. Slack. 2009. *The Geography of Transport Systems*. London, New York: Routledge.

Smith, M.J. de, M.F. Goodchild, and P.A. Longley. 2007. *Geospatial Analysis. A Comprehensive Guide to Principles, Techniques and Software Tools*. Leicester: Matador.

Spennemann, D.H.R. 2003. "The road to urbanisation. Post-dicting the evolution of the road network on Tongatapu, Kingdom of Tonga." In *Archäologische Perspektiven. Analysen und Interpretationen im Wandel. Festschrift für Jens Lüning zum 65. Geburtstag*, edited by

J. Eckert, U. Eisenhauer, and A. Zimmermann, 163–78. Rahden/Westf.: Verlag Marie Leidorf.

Terrel, J.E., and R.L. Welsch. 1997. "Lapita and the temporal geography of prehistory." *Antiquity* 71(273): 548–72.

Van Leusen, M. 2002. *Pattern to Process. Methodological Investigations into the Formation and Interpretation of Spatial Patterns in Archaeological Landscapes*. PhD Diss. University of Groningen. Accessed June 2012. <http://irs.ub.rug.nl/ppn/239009177>.

Verhagen, P., S. Polla, and I. Frommer. 2011. "Finding Byzantine Junctions with Steiner Trees." Paper presented at the workshop "Computational approaches to movement in archaeology", Berlin, Germany, January 6.

Waugh, D. 2000. *Geography. An Integrated Approach*. Cheltenham: Nelson Thornes.

Worboys, M., and M. Duckham. 2004. *GIS. A Computing Perspective. Second Edition*. Boca Raton: CRC Press.

White, D.A. 2012. "Prehistoric Trail Networks of the Western Papaguería. A Multifaceted Least Cost Graph Theory Analysis." In *Least Cost Analysis of Social Landscapes*, edited by Sarah L. Surface-Evans and Devin A. White, 188–206. Salt Lake City: The University of Utah Press.

Yu, C., J. Lee, and M.J. Munro-Stasiuk. 2003. "Extensions to least-cost path algorithms for roadway planning." *International Journal of Geographical Information Science* 17(4): 361–76.

Linking Roman Coins: Current Work at the American Numismatic Society

Ethan Gruber

American Numismatic Society

Gilles Bransbourg, Sebastian Heath

New York University

Andrew Meadows

American Numismatic Society

Abstract:

This paper details a current project at the American Numismatic Society: Online Coins of the Roman Empire (OCRE), an open-access catalogue of Roman Imperial coin types. The paper discusses linked data approaches applied to numismatic collections, the application architecture of the project, and the practical functionality of the project's web interface, illustrating its value to researchers.

Keywords:

Linked data, Information Architecture, Numismatics

1. Introduction

This paper describes a series of related initiatives to publish Roman coins—the basic material of the discipline of Roman numismatics that works to place coins in their historical and economic context—on the Internet using practices and methods that draw on the “Linked Data” approach to access and re-use of internet-based resources. The components of the work described here are: Nomisma.org, a digital resource that establishes stable URIs for numismatic concepts, with a current focus on both Greek and Roman coins; Numishare, a software platform for the management of numismatic data that supports linked data approaches; Online Coins of the Roman Empire (OCRE), a project recording and publishing all the known types of coins issued by the Roman Empire, accessible at <http://numismatics.org/ocre/>. While all of these projects are publicly available, they continue to be under active development. In combination they demonstrate the power of stable links between high-quality structured data.

In its capitalised form - that is as a defined set of practices - “Linked Data” is “a term used to describe a recommended best practice for exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web using URIs

and RDF.”² In this paper we focus on three aspects of Linked Data and describe how we implement them within the field of Roman coinage. Those three aspects are: stable URIs for each component of the study of coins, meaning URIs for descriptions of coin types, and for the vocabulary terms we use to describe those coins; availability of machine parsable data via those URIs; and rich re-use and interlinking between separate datasets that allows both cross-project efficient and discovery.

Coinage, as one of the ancient world's few examples of a mass-produced medium, is an excellent disciplinary laboratory—so to speak—in which to implement the concepts of Linked Data. Variations such as obverse and reverse designs and inscriptions, together with materials, denominations and dimensions offer a categorisation system that has long been employed by numismatists. It has achieved its canonical form for the coinage of the Roman Empire in the type-corpus Roman Imperial Coinage. This series of ten volumes, originally published between 1923 and 1994 and in the process of being updated, attempts to provide a unique identification number for every recorded variety of Roman Imperial coin, and has come to serve as a standard cataloguing resource for archaeologists and museum curators.

Corresponding author: ewg4xuva@gmail.com

² “Linked Data – Connect Distributed Data across the Web,” [Linkeddata.org](http://linkeddata.org), accessed October 31, 2012, <http://linkeddata.org/>.

Roman Imperial Coinage (RIC) is at one level an internally Linked Data product in its own right: it employs permutations of stable numismatic identifiers to assert a series of permanent identifiers (type numbers). Furthermore, the complex of information that makes up a coin type, as well as the coin type itself are inherently linkable to other categories of information about the ancient (or indeed modern) world: emperors, cities, findspots, magisterial offices all have significance beyond the world of numismatics. All of these attributes can be assigned stable URIs that facilitate cross-project and machine readable representations of this information through Nomisma.org.

2. Nomisma.org

Nomisma.org - begun in 2010 and hosted by the American Numismatic Society - is an ongoing project to establish stable URIs for numismatic concepts. It is currently available at the URI <http://nomisma.org>. What do we mean by numismatic concepts and why do such concepts need a separate namespace for their definitions? Numismatics has developed since the 16th century as an increasingly formal discipline encompassing the study of coinage and other physical monetary instruments used since the 7th century BC around the world. It is an important feature of numismatics that the majority of coins - here including all metal monetary instruments produced by a political entity to an official weight and with an established and distinctive visual appearance - can be described using a common vocabulary that is applicable to a vast range of ancient and modern examples. While the extremely broad scope of the discipline means that all general statements have exceptions, coins usually have an obverse and reverse, for which the colloquial English terms are “heads” and “tails”. Furthermore, there is only occasional ambiguity about how these two terms are used in the description of a particular coin. More specialised terms include “Legend”, meaning any words - whether abbreviated or not - on a coin, and “Axis” the angular relationship between the obverse and reverse. Most modern coins have an “Axis” of 6 on a scale of 12 because the reverse image is upside down in relation to the obverse. Additionally, it is important to know whether a visual or textual element - the portrait of a ruler or the phrase “E Pluribus Unum” - appears

on the obverse or the reverse. Because these terms have a specialized and well-established meaning, it is useful for the discipline to have a single resource which defines those meanings, both for use within the numismatics, and for the purpose of linking to other fields of study. Accordingly, Nomisma.org is inspired by the desire to promote interoperability, re-use and transparency within and beyond the very specialized field that it addresses.

An example Nomisma.org URI drawn from Roman numismatics is [http://nomisma.org/id/ric.1\(2\).aug.1a](http://nomisma.org/id/ric.1(2).aug.1a). From the perspective of the role of Nomisma within the discipline, the most important feature of that URI is that it is stable, relatively short, and semantically clear within its numismatic context. In particular, the opening characters of the final component of the path, “ric”, are recognizable as the de facto abbreviation for the series Roman Imperial Coinage, the standard reference work establishing identifiers for coins struck under the authority of the Roman Empire. Other segments of this identifier represent volumes (1(2) = volume 1, second edition), emperors (aug = Augustus), and reference numbers (no. 1a). This URI is itself too new to have seen widespread adoption, but it suggests a future in which multiple numismatic collections point that URI as an unambiguous identifier for coins of that type. This is a solution to the current problem whereby existing catalog entries—both online and in print—use various ad hoc abbreviation systems for RIC coins. These can be thought of as labels, and Nomisma does not mandate that collections use any particular sequence of characters for such human-readable indications of RIC type. Pointing to a URI is a sufficient and interoperable indication of shared identity across multiple collections.

It is a principle of Linked Data that there be machine parsable data available via the URIs that identify resources, with a strong preference for an RDF-based serialization. Nomisma.org has adopted XHTML+RDFa 1.1 as the archival representation of the information it represents about each numismatic concept it describes. Figure 1 shows that the RDFa markup is exposed to human readers in the form of labels on the attributes of that coin type. For example, the value “rome” is marked as being the mint. An RDFa distiller, such as that deployed by the W3, will produce the triple: ‘nm:ric.2.tr.432

nm:mint nm:rome' . Note that 'rome' is actually a reference to the resource <http://nomisma.org/id/rome> , this RDFa based resource in turn points to the Pleiades reference for Rome, which is a widely recognized identifier for the ancient capital of the Roman Empire. In this way the Nomisma.org ID facilitates bi-directional links to non-numismatic efforts to establish stable URIs for scholarly concepts.

Given our focus on cross-project reuse of vocabularies and data, this paper moves from describing our efforts to create stable URIs for numismatic concepts, to an application of such concepts within a further specialized field of numismatics, to a description of the software platform that supports specialized applications, and which is currently the basis for hosting and managing further numismatic data. This last point suggests that we are at a point of rapid progress in applying linked data concepts to many areas of the very wide field of numismatics.

3. Numishare: Technical Underpinnings

OCRE has three basic purposes. It is intended first as a typological catalogue of Roman Imperial Coinage that harnesses basic information from the Nomisma.org IDs to make the knowledge encapsulated in the RIC print volumes available to a broad audience, within the archaeological discipline, but also to a broader public. Using the principles of Linked Data it is also intended to provide links from numismatic descriptions to other online resources describing the ancient world, such as, for example the Pleiades project ([\[pleiades.stoa.org/\]\(http://pleiades.stoa.org/\)\). Thirdly, OCRE is designed to be able to accumulate examples of multiple examples of typologically similar coins to facilitate quantitative and qualitative analysis of the material, including metrological, compositional and die-study. To date, the coins from the American Numismatic Society's collection have been added. This was possible due to the prior digitisation of the collection in the Society's MANTIS database \(<http://numismatics.org/search/>\), the curatorial staff of the ANS having manually added all RIC-related information. In due course, other collections will follow. OCRE has been built on the Numishare platform.](http://</p></div><div data-bbox=)

On a technical level, OCRE is built on a data model in which numismatic metadata is represented as XML. Apache Cocoon is the server application for serializing and delivering data. Cocoon is an open-source Java-based framework designed to effectively process XML. In addition to Cocoon, the stack includes other open-source Java-based applications which run in Apache Tomcat on the server: Apache Solr for faceted searching, eXist XML database, and Orbeon XForms for editing, managing, and publishing XML. These applications, including various XSLT stylesheets, Javascript files, and CSS stylesheets, comprise Numishare, which is likewise open-source and freely available through GitHub at <https://github.com/ewg118/numishare>. The seeds of Numishare were planted in 2007 with the digitisation and publication of the University of Virginia Art Museum Numismatic Collection, a project which was detailed in a paper presented at CAA 2009 in Williamsburg, Virginia. Numishare has evolved considerably since 2009, to the point where very few lines of code from the earliest era of the

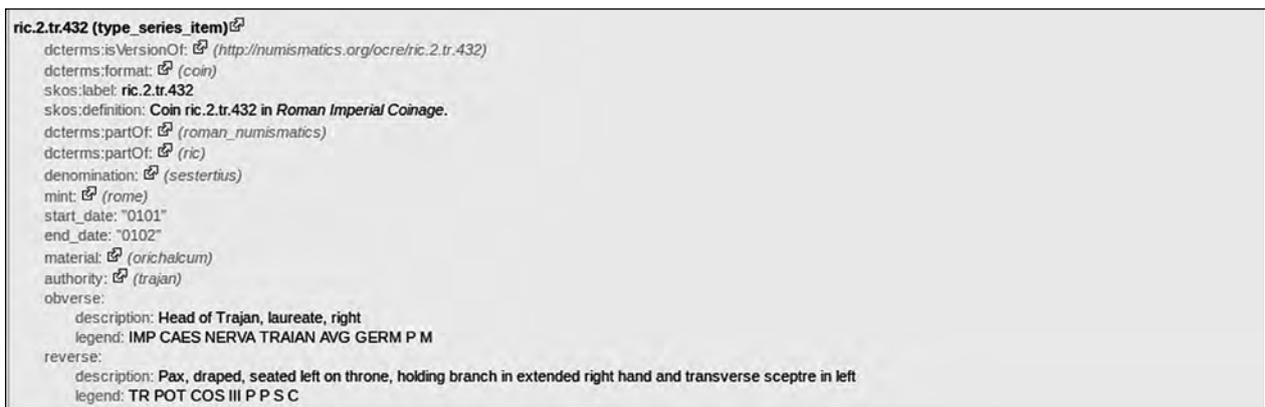


Figure 1. Nomisma coin type description.

project remain in the trunk of the repository. First and foremost, the XML data model was completely reinvented.

3.1 Numismatic Description Standard

The data model of the U. Va. Art Museum Numismatic Collection was an adaptation of Encoded Archival Description (EAD) to coins. EAD is an XML metadata standard used throughout the library and archival communities for encoding electronic finding aids.³ Thus, EAD is focused primarily on the description of written documents, though photographs, maps, and other predominantly non-textual objects can also be described. At the time, no standard for numismatic metadata had been authored, and the adaptation of EAD to coins was satisfactory given the needs of that project. Over time EAD was found to lack descriptors needed distinctly by numismatists and museum curators, as well as attributes to facilitate linked data interactions.

Independent of the development of EAD for coins and the U. Va. digitisation project, Sebastian Heath and Andrew Meadows of the American Numismatic Society had worked with a number of colleagues, at a series of workshops funded by the UK's AHRC, to develop a recommended list of numismatic and curatorial fields called the Numismatic Description (formerly Database) Standard (NUDS).⁴ NUDS contained many fields which were not included in EAD. Some of these were related to provenience and collection management: auction history, sale lot and price, owner, etc. Others were related to description of coins themselves: artists and engravers, symbols and countermarks, edge descriptions, and findspots. Many of these fields, like artist and engraver, can be expressed generally about the coin as a whole or more specifically about the obverse or reverse of the coin. NUDS fields were proposed with a database management system (DBMS) in mind, thus hierarchical complexities and repetition are difficult to represent easily or efficiently. Moreover, representing uncertainty

³ See "About EAD," The Library of Congress, last modified July 11, 2012, <http://www.loc.gov/ead/eadabout.html> for more information about the standard. The U.Va. collection is available at <http://coins.lib.virginia.edu>

⁴ Numismatic database fields detailed on Nomisma.org, accessed October 31, 2012, http://nomisma.org/nuds/numismatic_database_standard

on any level within the record is also difficult. The EAD and NUDS approaches had distinct advantages and disadvantages, and the American Numismatic Society has endeavored to address the differences between both in order to come to a solution that meets the requirements for all implementations of the model.

In summer 2011, work began on a data model which uses NUDS fields reconceived in an XML schema purpose-built for numismatics. By June 2012, the model, although still considered to be a draft, is stable, and several projects have been built on top of it, including OCRE.⁵ NUDS remains the label for the XML adaption of these fields, though the acronym now stands for "Numismatic Description Standard." The model influenced by other common XML standards found in the Library, Archive, and Museum world, like EAD, Encoded Archival Context – Corporate, Personal, and Family (EAC-CPF), VRA Core, Metadata Encoding and Transmission Standard (METS), Metadata Object Description Schema (MODS), and Text Encoding Initiative (TEI). Like several of these standards, NUDS uses the World Wide Web Consortium (W3C) xlink attributes for semantic linking, enabling links to URIs of Nomisma concepts, places on Geonames.org, and personal and corporate names defined by Virtual International Authority File (VIAF). A NUDS/XML record is fundamentally "linked data" because it is an electronic record which links to other resources on the web, although it is not linked data in the way that many of us may conceive of it: RDF in a triple store with a SPARQL endpoint for querying. Like traditional linked data architectures, querying for machine-readable data is supported in Numishare and will be discussed at a further point in this paper.

3.2 NUDS/XML: A synopsis

The document root of a NUDS/XML document is <nuds>, defined by the namespace <http://nomisma.org/nuds>. It requires one vital attribute for denoting the type of object represented by the electronic document. Objects are either physical or conceptual. A record for a physical object is self-explanatory: the XML document describes coin,

⁵ NUDS/XML data model draft detailed at <http://wiki.numismatics.org/nuds:nuds>, last edited October, 2012.

medal, token, paper note, or other numismatic object which physically exists in a collection. A conceptual object, however, is an intellectual construct. In numismatic terms, it is a coin type. Since OCRE is a corpus of coin types, this paper will focus primarily on describing the aspects of NUDS which define conceptual records.

Regardless of the physical or conceptual nature of the object being described, all NUDS documents must contain a NUDS Header. Like headers in TEI and EAD, the NUDS Header contains metadata about the electronic record itself: unique identifier, publication information, rights statements, and revision history. Within the document root, below the NUDS Header is Descriptive Metadata about the object itself. This includes, title, subjects, and physical, typological, undertype, findspot, and reference descriptions, in addition to administrative history for encoding provenience, ownership and accession information, and other other fields necessary for collections management. A physical object may utilize any or all of these categories of descriptors, but a coin type record would not contain descriptors for collections management, physical state, nor undertypes, which are only specific to physical objects. Thus, the Typological Description node is the only required Descriptive Metadata section.

The Typological Description (<typeDesc>) section of the NUDS document contains a mix of elements which utilize the W3C xlink:role and xlink:href attributes for semantic linking and other elements which simply contain free text. In Figure 2, the XML fragment shows that the <typeDesc> contains an object type, denomination, manufacture method, and material, each which link to concepts defined by Nomisma URIs. Below these elements are structured data: obverse, reverse, geographic, and authority nodes. The obverse and reverse may contain identical elements, including legend, type, and personal or corporate names defined by xlink:role. Roles may be authority, issuer, mint, region, deity, portrait, or others, as defined in Nomisma. The type element nested within obverse and reverse may contain repeatable description elements, with language defined by the xml:lang attribute. Numishare supports rendering documents in multiple languages, as long as labels denoted

by xml:lang exist either within the NUDS/XML record itself or within the RDF representation of the Nomisma URI.

As a practical example, suppose that a NUDS record contains both German and English type descriptions. A language parameter for German (“de”), may be passed to Numishare to display the appropriate type description, and lookups can be performed upon all Nomisma URIs within the document, parsing the RDF for each Nomisma ID and extracting the German SKOS preferred label for serialization into HTML. This move toward internationalization, greatly aided by Nomisma, will make it easier to develop and maintain multilingual interfaces for a single collection, as well as make it possible to query across multiple collections, regardless of the native language of the XML documents. When a NUDS record is published in Numishare and indexed into Solr, the URIs for Nomisma, the Pleiades Gazetteer of Ancient Places, Geonames, and VIAF IDs are stored in the index, enabling queries of the URIs directly (e. g., query all silver coins by the ID <http://nomisma.org/id/ar>, regardless of local encoding practice for materials). Furthermore, indexing of URIs for Pleiades places has facilitated integration with the PELAGIOS (Pelagios: Enable Linked Ancient Geodata In Open Systems) project, allowing Roman Imperial coins from Ephesus to be cross-searchable with other archaeological finds from the ancient city or classical texts which reference it⁶.

Below the Descriptive Metadata section within the NUDS document is the section for Digital Representations. In physical record types, one can embed a fragment of METS to link to digital images of a coin. Within OCRE, on the other hand, references to coins of that particular coin type can be inserted. The American Numismatic Society may have several physical specimens of a particular coin type, and other private or museum collections may also have specimens. Linking to physical coins which are represented online as stable URIs enables OCRE to ingest measurement data, findspots, and

⁶ In summation, PELAGIOS is a triplestore (and associated APIs) which aggregates RDF of various ancient collections online that are connected together via references to ancient places defined by Pleiades URIs. Read more at “About PELAGIOS,” The PELAGIOS Project, accessed October 31, 2012, <http://pelagios-project.blogspot.com/p/about-pelagios.html>.

```
<typeDesc>
  <objectType xlink:href="http://nomisma.org/id/coin">Coin</objectType>
  <date normal="-0018">18 BC</date>
  <denomination xlink:href="http://nomisma.org/id/denarius">Denarius</denomination>
  <material xlink:href="http://nomisma.org/id/ar">Silver</material>
  <obverse>
    <legend>CAESARI AVGVSTO</legend>
    <type>
      <description xml:lang="en">Head of Augustus, laureate, right</description>
    </type>
    <persname xlink:role="portrait">Augustus</persname>
  </obverse>
  <reverse>
    <legend>MAR VLT</legend>
    <type>
      <description xml:lang="en">Hexastyle domed temple, aquila between two standards inside</description>
    </type>
    <persname xlink:role="deity">Mars</persname>
  </reverse>
  <geographic>
    <geoname xlink:role="mint" xlink:href="http://nomisma.org/id/colonia_patricia">Colonia Patricia</geoname>
    <geoname xlink:role="region" xlink:href="http://nomisma.org/id/spain">Spain</geoname>
  </geographic>
  <authority>
    <persname xlink:role="authority" xlink:href="http://nomisma.org/id/augustus">Augustus</persname>
  </authority>
</typeDesc>
```

Figure 2. Typological Description fragment.

URLs to images, if available. These data can be used for quantitative and geographic analyses.

3.3 Numishare

Having introduced the NUDS/XML data model, it is now possible to discuss the applications which compose Numishare—how they function and how a user interacts with the public interface. As mentioned previously, there are four open-source, Java-based applications that run in Apache Tomcat. Essentially, Numishare is the code which links these separate applications together into one consolidated software suite. XSLT is the predominant scripting language within Numishare, with Javascript and CSS used to enhance the user experience. Apache Cocoon is the primary web publishing platform upon which the Numishare public interface is built. The Cocoon sitemap defines a myriad of pipelines which represent the Model-View-Controller relationship in XML applications development. Output serialization is formed by the combination of the data model (in some cases, NUDS/XML) and a controller and/or view (typically an XSLT stylesheet). The output serialization of a NUDS/XML document can be HTML rendered to a user of the OCRE interface in his or her browser, but XSLT stylesheets can also generate KML for rendering in maps or simply output the XML document itself to a user interested in taking advantage of the raw, unserialized data. The NUDS documents are stored in an eXist XML database, and therefore Cocoon reads the model directly through eXist's REST interface.

Cocoon also handles interactions between the user and the Solr search index. According to the official website for Solr:

*“its major features include powerful full-text search, hit highlighting, faceted search, dynamic clustering, database integration, rich document (e.g., Word, PDF) handling, and geospatial search. Solr is highly scalable, providing distributed search and index replication, and it powers the search and navigation features of many of the world's largest internet sites.”*⁷

Netflix was among its earliest adopters in private industry, and it has since become the *de facto* standard search index throughout the library community in the United States. It serves as the index for Blacklight, VuFind, and various Fedora repository implementations, as well as countless other projects. Solr has been a part of Numishare's application stack since 2007. It was chosen as the search index as a means of improving the user experience in searching and browsing that were seen as shortcomings of Numishare's two main influences, the Princeton University Numismatic Collection and the Berlin Münzkabinett, which were in 2007 and continue to be among the most visible numismatic databases on the web.⁸

7 “Apache Solr,” Apache Software Foundation, accessed June 20, 2012, <http://lucene.apache.org/solr/>.

8 See “RBSC: Numismatic Collection,” Princeton University, accessed October 31, 2012, <http://www.princeton.edu/~rbsc/department/numismatics/> and “Münzkabinett der Staatlichen Museen Berlin,” accessed October 31, 2012, <http://www.smb.museum/ikmk/>.

The following is a brief description of the functionality of the search and browse interfaces in Numishare. Solr accepts queries following the Lucene query syntax submitted by the user through the interface, and Solr responds through its REST API with query results in the form of XML (although it also supports JSON output). Cocoon intertwines this XML model with an XSLT stylesheet, which renders the results in the form of human-readable HTML. Query results can also be serialized into several other formats. Results can be delivered in the form of an Atom feed, which allows users to subscribe to changes made to coins which fit their key area of interest (e. g., coins of Hadrian). More importantly, the Atom feed is the primary means by which data may be harvested from OCRE by programmatic means. The feed contains 100 hits per page, with links to next and previous pages, as well as links to alternates to the default HTML serialization of a coin type, such as RDF, KML, and NUDS/XML. Therefore, it is possible for a script to page through the entire collection and extract the machine-readable NUDS metadata. In addition to the Solr-based Atom feeds, Numishare also supports serializing Solr search results as KML and CSV. While Cocoon handles output for the public user interface, XForms applications processed by Orbeon, which also runs in Tomcat, form the foundation of the private administrative interface.

XForms is a W3C specification for next-generation web forms which adhere to the Model-View-Controller philosophy.⁹ Orbeon is used in the administrative back-end to edit complex XML metadata and interact with REST services. XForms applications query various APIs on the web for enhancing controlled vocabulary and manage Create, Read, Update, and Delete (CRUD) operations between the eXist database and Solr index. Specific details about the architecture of Numishare's back-end are too complex to be detailed within the limited space of this paper, but advantages are briefly enumerated below:

- XForms supports validation. For example, numeric fields such as diameter and weight must adhere to the `xs:decimal` datatype. In other cases, XForms validation warns the user that certain

⁹ The XForms Wikipedia article provides a sufficient description of the standard. "XForms," Wikipedia.org, last modified October 21, 2012, <http://en.wikipedia.org/wiki/XForms>.

fields may not be left blank. Invalid data cannot be sent into eXist.

- Authority control. XForms taps into auto-suggest for local controlled vocabulary delivered from the Solr index or interacts with REST APIs of Geonames, VIAF, and Nomisma.org to embed URIs directly into the `xlink:href` attribute within the NUDS/XML document.
- Link to images with Flickr's APIs.
- Easily link to coins in other collections through automated processing.
- A fuller description of XForms, as well as other use cases for similar applications are detailed in "XForms for Libraries, An Introduction," published in *Code4Lib* 11.¹⁰

Together, this suite of applications known as Numishare forms the basis for a powerful research tool for those interested in the study of Roman Imperial Coins. Several features of the public interface have already been discussed (KML and Atom feeds), but this paper will now turn toward a more focused discussion of the public interface of OCRE.

4. OCRE: A Powerful Research Tool for Roman Imperial Coinage

4.1 Searching and browsing

Despite numerous and significant advances in Numishare since its inception, especially with respect to collections management and publishing in the XForms-driven backend, the application in its core is designed to deliver a simple, yet sophisticated interface for navigating and interacting with coin collections. The search and browse interfaces are good places to begin a discussion about the features of Numishare, for these are what separates Numishare from most other collections. Solr supports faceted searching based on a handful of typological categories: authority, deity, denomination, manufacture, issuer, material, mint, portrait, and region. Full-text search is supported,

¹⁰ Ethan Gruber, Chris Fitzpatrick, Bill Parod, and Scott Prater, "XForms for Libraries, An Introduction," *Code4Lib* 11 (September 21, 2010), accessed October 31, 2012, <http://journal.code4lib.org/articles/3916>.

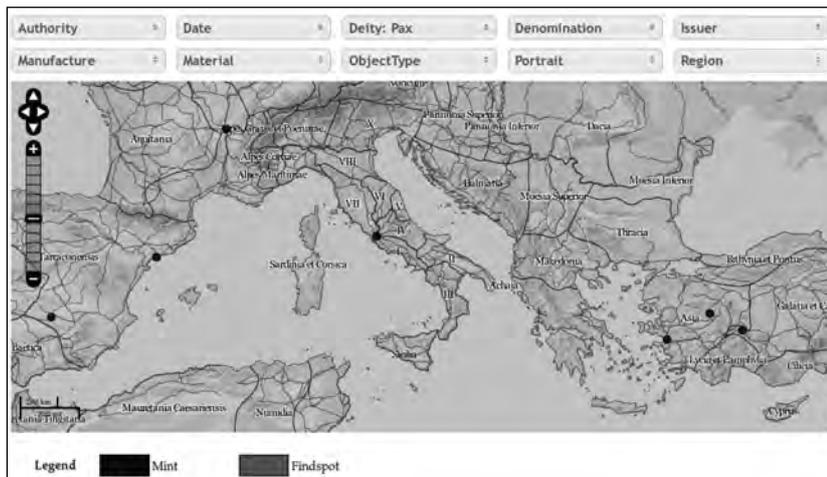


Figure 3. Preview of dynamic mapping interface.

as well as searches of other fields, like obverse or reverse legend or type, date ranges, among others. Search results can be sorted by a handful of fields, including all those listed above.

Importantly, Solr supports wild card searches, which are especially useful legends which are only partially legible. Such searches can aid in the identification of coins acquired in archaeological excavations which may be in a poor state of preservation. For example, one may search for “P?CI AVG*” within the reverse legend to yield several dozen matches for the legend “PACI AVGVST” or “PACI AVGVSTAE.” A “?” wild card represents zero or one unknown characters while “*” represents one or more. Results may be filtered further by facets. Given this partial legend query, suppose the metal of bronze is readily identifiable, and the obverse portrait, however worn, appears to exhibit the features of the emperor, Domitian. In OCRE, this yields one result (through the reign of Antoninus Pius, ending in A.D. 161). Such queries can be represented by Atom, and the metadata for coin types matching these criteria can be ingested into local databases by machine-readable processes, reducing the data entry workload of archaeologists or other recorders.

With the adaptation of Numishare to the American Numismatic Society’s collection, called Mantis, in early 2011, a new faceted geographic search interface was introduced, in addition to the traditional textual interface typically associated with Solr searches. Javascript was written to connect

faceted search with OpenLayers, an open-source Javascript library for mapping. These Javascript functions accept interactions from the user of the interface to query Solr, which replies with XML piped through a Cocoon+XSLT transformation into KML to rapidly update the OpenLayers map. Visualizing queries geographically is potentially immensely useful. One may map the distribution of coins over time and space (Figure 3 shows the mints which produced coins depicting the personification, Pax) or map the distribution of a

particular coin type over the whole of the Roman Empire, which would give an economic historian a glimpse of ancient trade networks.

The final interface constructed upon Solr is one that visualizes search results in the form of charts and graphs. A query of all coins which depict the deity, Victory, may be visualized as a bar graph depicting the total counts per Roman emperor. In this, Vespasian leads all other emperors.¹¹ Such visualisations may lead to research questions which may have otherwise never been asked.

4.2 Coin type records

Like most other aspects of Numishare, the coin or coin type record HTML serialization has evolved considerably since it was initially documented at CAA 2009, although some features of the HTML page have remained. The design of the object record page was conceived in 2007 to address the perceived shortcomings of record pages in other databases, namely the “dead-end” nature of those pages. Links from search results typically led to pages for coins which contained metadata, but no way to seamlessly traverse from those records to records of related coins. The 2007 design created links for typological attributes—like denomination, material, and mint—which would direct the user to the search results page for that attribute. One could navigate from a particular denarius of Augustus to other denarii

¹¹ Visualisation parameters are passed RESTfully to the URL. See http://numismatics.org/ocre/visualize?limit=32&type=areaspline&category=authority_facet&q=deity_facet%3A%22Victory%22 for a representation of this query.

and interpretation of stratigraphic contexts, even when the excavation lacks numismatic specialists.

OCRE will continue to evolve even after the publication of this paper. More NUDS records will be added into OCRE, terminating with Anastasius in the early sixth century. New features for quantitative analyses and visualisations will be added into Numishare. Physical coins in other collections will continue to be linked to coin types. At this moment, the Portable Antiquities Scheme, British Museum, OpenContext and the Münzkabinett of the Staatlichen Museen in Berlin have committed to sharing their data with the project. Since a relative small proportion of the American Numismatic Society's coins linked to OCRE have attested findspots, it is a primary goal moving forward to import data about archaeologically excavated coins into the project. Only then can the full scholarly potential of OCRE be realised. In some sense, OCRE, at the date of this publication, is merely a starting point in the development of one of the most comprehensive and sophisticated tools for the study of Roman coins on the web.

References

- Apache Software Foundation. 2012. "Apache Solr." Accessed June 20, 2012. <http://lucene.apache.org/solr/>.
- Gruber, E. 2009. "Encoded Archival Description for Numismatic Collections." Paper presented at the *Computer Applications and Quantitative Methods in Archaeology Conference*, Williamsburg, Virginia 23-27 March, 2009.
- Gruber, E. 2012. "Numismatic Description Standard: Draft." Last modified October, 2012. <http://wiki.numismatics.org/nuds:nuds>.
- Gruber, E, C. Fitzpatrick, B. Parod, and S. Prater. 2010. "XForms for Libraries, An Introduction." *Code4Lib* 11. <http://journal.code4lib.org/articles/3916>.
- The Library of Congress. 2012. "About EAD." Last modified July 11, 2012. <http://www.loc.gov/ead/eadabout.html>.
- Linkeddata.org. 2012. "Linked Data - Connect Distributed Data across the Web." Accessed June 20, 2012. <http://linkeddata.org/>.
- Nomisma.org. 2012. "Numismatic Description Standard." Accessed June 20, 2012. http://nomisma.org/nuds/numismatic_database_standard.
- PELAGIOS Project. 2012. "About PELAGIOS." Accessed October 31, 2012. <http://pelagios-project.blogspot.com/p/about-pelagios.html>.
- Wikipedia.org. 2012. "XForms." Last modified October 21, 2012. <http://en.wikipedia.org/wiki/XForms>.
- World Wide Web Consortium. 2012. "XML Linking Language (XLink) Version 1.1." Accessed June 20, 2012. <http://www.w3.org/TR/xlink11/>.

GeoDia: or, Navigating Archaeological Time and Space in an American College Classroom

Adam Rabinowitz

The University of Texas, Austin, USA

Abstract:

*GeoDia (<http://geodia.laits.utexas.edu>), a spatial timeline focusing on the archaeology of the ancient Mediterranean, is an attempt to use accessible, familiar online visualisation tools to help students understand synchronisms and material connections between cultures, while also serving as a study aid in support of the absorption of facts about time, space, and objects in the ancient world. Based on *timemap.js*, an open-source Javascript library that mashes up Google Maps and SIMILE timeline APIs, GeoDia provides an interactive interface that combines periodised archaeological sites, image resources, historical events, and attributes like culture and region. This paper presents the theoretical considerations involved in the creation of the structure of GeoDia and reviews survey data from several hundred students who have used GeoDia either during beta-testing or in the course of formal class assignments. It also addresses linked data issues and the reusability of the dataset independent of the interface.*

Keywords:

Geotemporal Visualisation, Pedagogy, Periodisation

1. Introduction

There is an inherent tension in archaeology between the complexity, uncertainty, and incompleteness of the evidence and the relatively clear and simple narratives that present the results of archaeological research to the public. This tension has been exacerbated by digital approaches to archaeology: on one side, field researchers struggle to represent and share complex datasets and the processes that generated them for the benefit of other researchers, while on the other, three-dimensional reconstructions and museum websites elide complexity and process to present a seamless, coherent vision of the past for the benefit of lay audiences. The archaeological computing community should see as one of its “grand challenges” the use of digital tools to help students and the public understand the imperfect and complicated process of interpretation that connects archaeological evidence to the stories we tell about the past.

These issues are relevant to the way many of us present archaeology in university classrooms, especially in terms of the spatial and temporal aspects of material culture. Archaeologists are well aware of the multiplicity of terms and concepts that are used to group material remains into

Corresponding author: arabinow@utexas.edu

chronological periods or geographic units and of the way in which space and time interact in these classifications. If an author describes a ceramic vessel as “Iron Age”, the reader must know where the vessel was found, where the author conducts research, and what disciplinary camp he belongs to in order to understand what that term implies about absolute date, style, and culture. This fact has led to much work on the modelling of periods and time intervals in ontologies like CIDOC-CRM and on the harmonization of temporal information using tools like the STAR Timeline application (Binding 2010; Hypermedia Research Unit, University of Glamorgan 2012). Cultural interactions in the ancient world present similar complexities. Not only has the concept of the “archaeological culture” been called into question in archaeological theory since the second half of the 20th century, but 21st-century developments in the digital world have encouraged a view of antiquity in terms of a web of social networks and trade connections that ignore modern concepts of geospatial and cultural boundaries (e.g. Malkin 2011; Scheidel and Meeks 2012). Yet when we teach archaeology to undergraduate students, we find ourselves falling back on temporal and cultural paradigms that go back to the 19th or even 18th century. These paradigms insist on the separate and unique nature of historical cultures and on the “biological metaphor” of birth,

adolescence, maturity, and senescence to explain their development.

2. Teaching the Past

At least on North American university campuses, introductory art and archaeology courses typically focus on a single culture, as do most of the available textbooks (e.g. Kleiner 2007; Pedley 2012), and courses and textbooks alike spend a disproportionate amount of time on the *floruit* of that particular culture (for example, Classical Athens or the early Roman Empire). Interactions with other cultures are mentioned primarily in terms of conflict with or assimilation to the culture being studied (“Hellenisation” and “Romanisation”). The lasting impression is of clearly delineated and distinct cultures that follow chronological arcs of growth, peak, and decline within fixed geographical spaces that often coincide with modern national borders. This paradigm is even more pronounced in textbooks that cover several cultures together. In the first volume of Marilyn Stokstad’s introductory Art History textbook, for example, the student encounters first ancient Near Eastern and Mesopotamian art, then Egyptian art, then Bronze Age Aegean art, then Greek and then Roman art (Stokstad and Cothren 2010). Although all of these cultures were in contact with each other to a greater or lesser degree throughout almost the entire timespan covered by these chapters, each is presented in isolation -- and, implicitly, in succession. As Daniel Rosenberg and Anthony Grafton have ably demonstrated in their recent book *Cartographies of Time*, the division of the past into a series of culturally and geographically distinct empires, in which preeminence passes from one to the next in teleological progression, has a pedigree stretching back to antiquity and has exerted a deep influence on representations of the chronology of the ancient world (Rosenberg and Grafton 2010).

Both students and the general public are conditioned to expect this approach to the past. As instructors, we also find it easier to adopt a clean, linear narrative in which dates and periods are unambiguous, cultures are compartmentalized, and certain historical moments are more important than others. Departing from this narrative goes against the grain for both the presenter and the audience and requires substantial additional effort to keep

the audience oriented in both time and space. I encountered this first-hand in a course on Classical Archaeology I taught at The University of Texas at Austin in 2007. This course was meant to introduce students to the archaeology of the Mediterranean world from the introduction of writing in Sumeria to the foundation of Constantinople. Rather than follow a traditional succession-of-empires paradigm, I tried to structure the course in such a way that the students could see connections and interactions across space and time. We therefore followed a linear timeline arranged as a series of horizontal slices: that is, we looked at snapshots of the entire Mediterranean world at particular moments in time while following diachronic developments at a specific subset of important individual sites such as Babylon and Athens.

The students, however, found this approach tremendously confusing. I offered strategies to help them organize the large quantity of disparate information involved, such as the creation of grids in which archaeological monuments could be associated with regions by period. I also made efforts to deploy online mapping tools to help them understand spatial relationships: for instance, they were given a group project in which they had to locate and provide information for a particular archaeological site using the neogeographical website Wikimapia (<http://wikimapia.org>; for an example of the results, see <http://wikimapia.org/#lat=37.6610827&lon=27.2968894&z=16&l=0&m=b&show=/3653339/Priene&search=priene>). While this assignment was successful in helping the students make sense of space in the past, there were few comparable online resources to help them make sense of time in the past, and even fewer that brought the two together. This should not be surprising: a number of scholars have noted that the “spatial turn” brought to humanities scholarship by GIS and neogeographical tools has not been accompanied by new and effective ways to represent space across time (Gregory 2007; Bodenhamer, Corrigan, and Harris 2010). In 2007, the only platform that offered the capacity to integrate space and time was provided by the TimeMap project, which allows the creation of maps integrated with timelines in a Java-based environment (Johnson 1999; Johnson 2008). TimeMap is an innovative solution to the problem of geotemporal visualisation, but it presents a high bar for entry for both website creators and users,

and has tended to be used for the illustration of authoritative narratives rather than for student-driven investigation of confusing, interrelated, and sometimes contradictory information. The same is true of SahulTime, another geotemporal visualisation interface that began to take shape around the same time (Coller 2009). Here an Adobe Flash platform is used primarily to show the changing shape of Australia and the surrounding archipelagoes as sea-level changed over time. Although the interface includes interactive elements, it too presents an authoritative linear narrative. GeoDia (<http://geodia.laits.utexas.edu>) grew from a desire for an interface that would provide a more user-centered encounter with the complexities of the spatial and temporal aspects of the ancient Mediterranean.

3. Building GeoDia

GeoDia also emerged from a series of lucky coincidences and collaborations. While I was looking for better ways to teach ancient Mediterranean archaeology, Nick Rabinowitz was exploring geotemporal visualisation as a graduate student in the School of Information at the University of California at Berkeley. He had already begun to experiment with a mashup of map and timeline APIs, and we decided to work together on the development of an interface for the display of information with both spatial and temporal attributes. The project was also fortunate to have the support of the Liberal Arts Instructional Technology Services of the University of Texas at Austin in the form of two generous grants; the time of developer Peter Keane, programmer Stuart Ross, and project manager Adrienne Witzel; and the DAsE (Digital Archives Services) content-management architecture created by Keane (<http://code.google.com/p/dase>).

The GeoDia interface is based on, and developed in concert with, Nick Rabinowitz's open-source timemap.js Javascript library (<http://code.google.com/p/timemap>), which integrates the Google Maps API with the API of MIT's SIMILE Timeline. This permits the loading of data



Figure 1. Screen capture of the sites tab in GeoDia, with Athens selected.

simultaneously onto both a map and timeline, with a separate data structure to manage cross-references and interactions between the two visual representations of a given record (Fig. 1). The result is an online interface in which user interaction with either the map or the timeline can be reflected in both visualisation elements; for example, clicking an item on the timeline may bring up an informational window on the map, and markers on the map may be hidden when the corresponding marker on the timeline is out of view. The central dataset consists of geolocated archaeological sites, which appear as continuous lines broken into the periods that are conventionally used by scholars who study those particular sites, and with which are associated images of objects and monuments produced or found there, arranged according to those periods.

To allow the user to create data views interactively through both searching and faceted browsing, we took a somewhat unorthodox approach to periods, cultures, and geographic divisions. In most online timelines, periods exist as independent concepts (that is, “the Archaic period” describes a fixed set of characteristics, applicable across a fixed date-range) and are assigned to conventionally-defined regions (“Greece”, “Egypt”, etc.). In our datamodel, by contrast, those broad period concepts are manifested primarily in the form of “site periods”, which describe the application of period concepts to the concrete periodisation of a specific site (Fig. 2). Archaeologists may use the same term to refer to different absolute dates at different sites, for example, or teams with different intellectual orientations may use different period

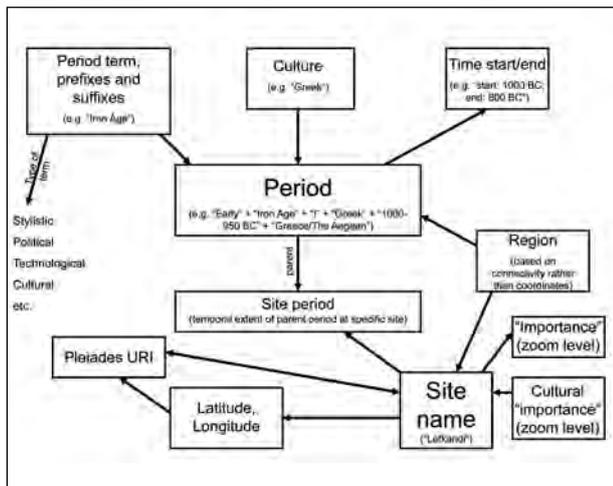


Figure 2. The GeoDia datamodel.

terms to indicate the same date range at closely related sites. By applying the periodisation used by the archaeological literature that discusses each individual site, GeoDia introduces students to some of the complexities of the use of the concept of “period” in archaeological research.

Similarly, “archaeological culture” is expressed as a function of time as much as it is of geographic area. Culture is thus incorporated into period definitions, so that Greek, Roman, and Etruscan cultures are each associated with their own distinct “Archaic” period. When a period is associated with a site, its related culture is instantiated at that geographic location during that time-range. This allows the interface to color-code “culture” in the timeline and on the map without insisting that a given site belongs only to one culture. On one level, GeoDia takes a traditional approach to the idea of archaeological culture by using the concept as one of its main browsing facets and by color-coding only the most familiar ancient cultures. This was necessary, however, to make the site accessible to students whose understanding of the past is already shaped by the conventional idea of culture, and who encounter ancient cultures as discrete units in the classroom. At the same time, we tried to represent at least some of the complexity of culture in the ancient world by showing graphically that the same site might be occupied by several different cultures over the course of its history. The interface also emphasizes that those cultures are not mutually exclusive: periods can be associated with two cultures at once, in which case they appear with zebra-stripping in the timeline. This permits us

to represent, for example, the “culture” of Roman Athens, where absorption into the Roman political system transformed but did not replace the Greek culture of the site’s inhabitants.

The description of regions in GeoDia also reflects the fluidity of the Mediterranean world across time. Approaches to the geographical representation of the ancient world tend to be framed either by modern national borders or by efforts to represent ancient administrative or military boundaries. The first is essentially arbitrary with respect to the past, and not useful for browsing -- a student who sees only sites in modern Greece when searching for “Greek sites” will miss large parts of the ancient Greek world in the western Mediterranean or along the Black Sea. The second is also problematic: not only is our knowledge of the past too fragmentary to create a continuous representation of shifts in political control across the entire Mediterranean over more than three millennia, but mapping those boundaries tends to obscure their permeability and the continuous interactions that took place across them. For GeoDia, we decided to define regions by points rather than by polygons -- that is, we defined them according to interconnectivity and shared material culture between individual sites, rather than according to linear boundaries and borders. The result, displayed at <http://geodia.laits.utexas.edu/regions.html>, is organized in part by natural barriers like mountain chains and by the connectivity provided by waterways, and in part by affinities in material culture over time. This makes for some counterintuitive divisions: Sicily, for example, is not grouped with Italy, but split between “the Aegean” and “North Africa”. But such choices make more sense when the historical record is considered. Not only was eastern Sicily settled by Greeks and western Sicily by Punic colonists in antiquity, but the same areas were later divided between Byzantine and Arab control.

The presentation of images by period and site was a fundamental part of the original concept for the interface, to provide users with a sense of the material culture of a particular site across time, offer students another way to contextualize the material they encountered in their textbooks, and permit the comparison of objects and monuments across time and between cultures. The interface therefore provides an image search as an additional point of

entry to the dataset of periodised sites. The user can search by keyword for specific images, and, by clicking on the link to the site with which an image is associated, see the image in its spatial and temporal context. In response to student feedback, the beta version of GeoDia included an additional dataset consisting of historical events that can be placed at a specific location in space within a specific calendar year. They can thus be displayed as points on the timeline, which distinguishes them from sites and makes it possible to see more of them at once. Events, like sites, are associated both with a region and with one or more cultures, and can be browsed or searched through a separate “events” tab in the interface.

GeoDia is much closer to a neogeographic platform than a historical GIS, in that it adds information to a representation of our landscape as it is now, rather than attempting to depict a landscape as it was in the past. The interface does not display ancient cities at certain moments in time; instead, it displays present archaeological sites that once were ancient cities, with the objects we attribute to those cities arranged according to the chronological groupings we use to describe them. This is in sharp contrast to projects that map the ancient world itself, most notably the Pleiades Project, which grew from the Barrington Atlas of the Greek and Roman World (Talbert 2000) to present an online historical gazetteer of the cities, sites, regions, and other spatial divisions of the ancient Mediterranean as they appeared in antiquity (<http://pleiades.stoa.org>). It also differs from more recent approaches to the temporal modelling of digital data, which focuses on the temporal dimensions of a thing *in the past* (an empire, a building, a human life: Fazal 2010). This is a deliberate choice: GeoDia is not intended to provide students with an illustration of “the way things really were” in the past, but to confront them with the way we, as archaeologists and historians, study and describe evidence for the past today. This approach has now been adopted by other projects, one of which has recently adapted the `timemap.js` library to display georeferenced historical maps of archaeological monuments in Greece (Koussoulakou et al. 2011). The distinction between present and past is less clear in GeoDia’s current dataset of historical events, which does not yet include primary sources for its items and which features only a few initial efforts to represent scholarly controversy about

dates. It is our intention, however, to emphasize knowledge sources and conflicting interpretations in future iterations of this dataset too.

In fact, the current version of GeoDia has fulfilled only one of the project’s two original goals. The first goal envisioned the creation of a spatial timeline of ancient Mediterranean archaeology and history, with authoritative content defined, reviewed, and formatted by an administrative group and presented to students as set information that could be browsed and searched. This was intended to be both a study tool, helping students to understand chronological terms, stylistic groupings, and geography, and a complement to the linear, simplified narrative provided by textbooks, revealing connections and synchronisms between cultures and regions. Essentially, it serves as a contextualizing engine for material culture: instead of presenting individual objects and sites in geographic and chronological isolation as episodes in a grand overarching story, it allows the user to see what scholars associate with a single moment in time across space and with a single space across time.

The second goal of the project, still unrealized, was to bring that complexity more fully to life by allowing students to create their own datasets. If the current interface is an authoritative presentation of curated data, the second goal envisioned an interface equivalent to the “My Maps” section of Google Maps, where the same geotemporal tools could be used to display user-generated content. The best parts of the latter could then be merged with the main dataset, allowing students to participate actively in the generation of authoritative content while expanding GeoDia’s coverage through crowdsourcing. On a pedagogical level, this would have two effects on students: it would allow them to actively construct their own narrative framework for the material, which they would be more likely to internalize than one received passively from a textbook; and it would force them to confront, in the process of creation, the complexities and ambiguities that are inherent in archaeology and ancient history but have been smoothed out of the authoritative dataset for practical reasons. The University of Virginia has successfully deployed such an approach to visualisation in connection with student research projects (Ferster 2011), although

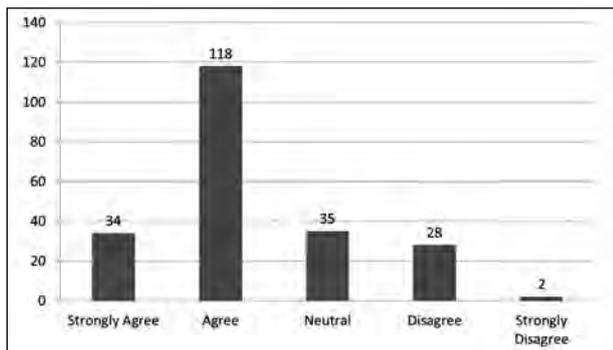


Figure 3. Graph of responses to 2009-2010 survey prompt: “While navigating the site, using the interactive timeline and map was easy and effective.”

there the platform is Adobe Flash, which was not suitable for our project because of limitations for data interoperability and export.

4. What Students Want

In the end, we lacked the time and resources to build the interface and infrastructure necessary to enable this sort of unmediated content creation in GeoDia. We hope to include it in a future phase of development. In the meantime, however, we have had the opportunity to test various iterations of the site with student users, both as passive consumers and as information producers. With the help of colleagues in the departments of Classics and Art History, we were able to solicit feedback from a wide range of students at two different stages of the website’s development. Project manager Adrienne Witzel conducted beta-tests and administered surveys during GeoDia’s first public iteration in 2009-2010, and I followed this in 2011 with surveys administered to students in Greek and Roman Archaeology classes who were assigned group projects involving the investigation and periodisation of archaeological sites not yet included in the GeoDia dataset. We wanted to collect information both about the ease-of-use and efficiency of the interface and about its effectiveness as a teaching and learning tool.

In the first round, surveys were administered to students in an introductory Art History class, an introductory Roman Archaeology class, an introductory ancient Mediterranean history class, and a Celtic Archaeology class. In 2009, students were given a simple GeoDia-based assignment

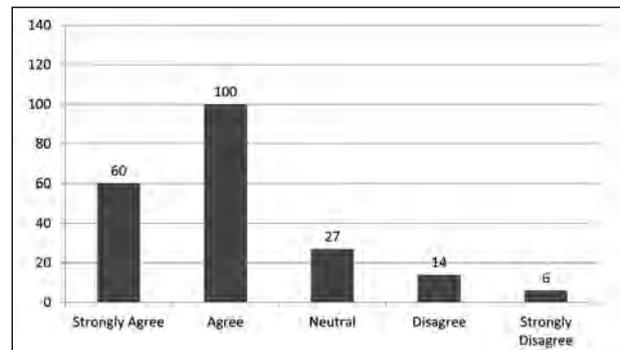


Figure 4. Graph of responses to 2009-2010 survey prompt: “The layout of the website allows me to effectively associate the artifact images with a site, culture, and time period.”

and a 36-question survey, while the 2010 students were given a 39-question survey that included new questions about a user guide that had been added to the site. Completed surveys were collected from a total of 237 students during these two semesters. I will focus here on quantitative results related to this central question: does the site offer an intuitive and comprehensible way for students with no background in archaeology or ancient history to look at space and time?

At least in terms of basic user experience, the answer seemed to be “yes”. The vast majority of the respondents routinely used the internet for their studies (119 of 237 said they “always” use online sources in their research), so they were already fairly familiar with online applications. When students were asked more specifically in the 2011 survey about the online resources they used, spatial tools were the most frequently used after UT Austin’s course-management platform, so it seems safe to assume a widespread familiarity with online maps within the student population. The students also seemed to be comfortable with the mashup of map and timeline in GeoDia: a majority found the interface easy to use (Fig. 3). A similar majority thought that the interface was successful in allowing them to gather temporal and cultural information about sites, and agreed that it helped them to associate images of objects and monuments with specific places and times (Fig. 4). Almost 140 of the 237 respondents answered “agree” or “strongly agree” to the statement “This site is a useful research tool that I will use in my studies”.

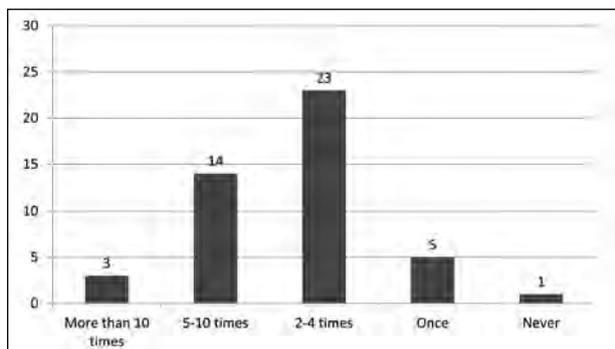


Figure 5. Graph of responses to 2011 survey prompt: “Over the course of the semester, I have visited the GeoDia website...”

These positive responses, however, were tempered by the short-answer survey prompts. Lengthier student comments across these various classes indicated a strong desire for the very sort of clean narrative framework the site had been designed to avoid. A large number of students wanted more textual information about the sites themselves, and an even larger number wanted historical events to be represented. The second demand was so pronounced that we modified the interface in its second iteration to include events. It is interesting to note the students’ strong desire for more information in the same format -- the short explanatory essay -- conventionally provided by online timelines and museum exhibits. This theme reappeared in the surveys administered in 2011, although with a closer connection to textbooks and course content.

In the second round, surveys were given to students in two courses: an introduction to Greek archaeology and an introduction to Roman archaeology. These students had been strongly encouraged to use the site to study course material, and had also been assigned group projects that involved gathering, for a given archaeological site, the same sort of information that GeoDia provides (periodisation, images, location, etc.). Responses were collected from 46 of those students, with some unexpected results. Although a similar percentage of these students found the interface easy to use, they did not seem to use it often in practice (Fig. 5). Even more strikingly, a distinct majority -- 26 out of 46 -- indicated that they “never” used the site “to understand course material or study for exams”. The most frequent reason respondents gave for

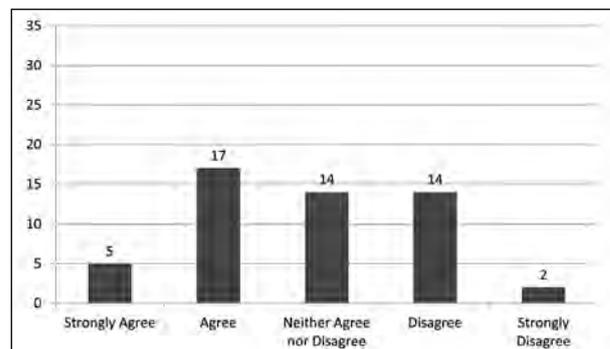


Figure 6. Graph of responses to 2011 survey prompt: “My work on the GeoDia group projects increased my understanding of the course material.”

not using the site to study was that the content in GeoDia did not overlap enough with the specific material presented in the textbook and course lectures. Since the tests were based on the textbooks and lectures, and since the dates and period terms in GeoDia sometimes did not precisely match those in the textbook, the students felt that the site would not help them recall the material on which they were tested. This contrasted sharply with my own ideas about the site: as an archaeologist, I had assumed that looking at a wide range of objects in their spatial and temporal context would give students a better visual sense of such context, making it easier for them to place both known and unknown objects in space and time.

There was a similar disjunction between the assumed pedagogical benefits of active engagement with issues of time and space, and how useful the students felt the group projects to be. Superficially, those projects appear to have failed: a majority of students did not think the group projects increased their understanding of the course material (Fig. 6), and an even larger majority did not think that the projects were a valuable part of the course. From a pedagogical perspective, however, the projects were more successful: significant majorities felt that the projects had increased their understanding of library research, chronology and periods in the ancient world, and the process of archaeological field research. Most surprisingly, after the largely negative assessment of the value of the projects, almost all the students surveyed felt the experience had made them more able to collect and evaluate evidence in general (Fig. 7).

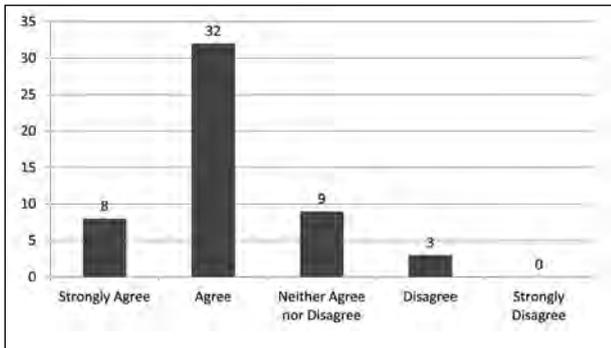


Figure 7. Graph of responses to 2011 survey prompt: “My work on the GeoDia group projects improved my ability to collect and evaluate information about academic subjects in general”.

These results suggest two basic conclusions. First, engaging students actively in the process of knowledge generation and in complicated questions of interpretation is a useful strategy, even when the students themselves are not enthusiastic about the experience. The development of interactive resources for history and archaeology should therefore consistently take into account the possibility of student and public engagement. “Citizen science” has already made its way into archaeology and Classics, from the Portable Antiquities Scheme (<http://finds.org.uk/>) to the Ancient Lives Project (<http://ancientlives.org/>). It remains to be seen where crowdsourcing approaches will be most successful in terms of scholarly advancement, but it is already clear that they have great potential for making students and lay audiences stakeholders in the exploration of the past. Second, it is impossible to ignore the desire of students to use an interface like GeoDia to find more information. If GeoDia is to be of lasting value, its future development must include better mechanisms to link its dataset to additional sources of information about the ancient world, and to expose its information in turn for reuse by other platforms. I think that Linked Data approaches are the most likely to make this possible.

5. Linking the Past, Linking the Future

Since we began to develop GeoDia, online interfaces for geotemporal visualisation have increased in both number and complexity. The most recent addition is Neatline, which provides tools to build such visualisations in an Omeka environment (<http://neatline.org>; Nowvickie 2011). Yet the more

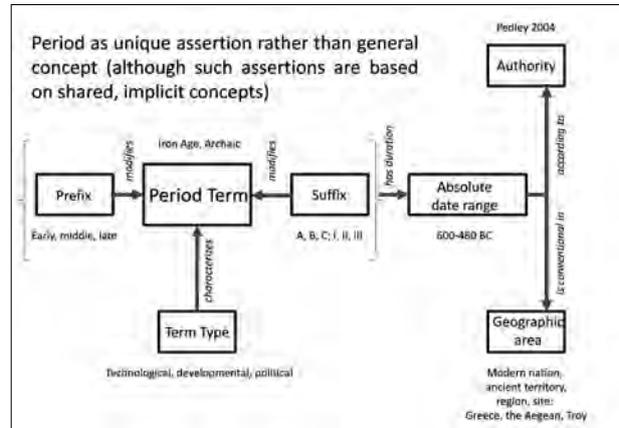


Figure 8. Model of a period as a unique assertion rather than a general concept.

complex these interfaces become, the more fragile they become and the more vulnerable they are to degradation over time (Nowvickie and Porter 2010). GeoDia is no exception. Not only does it depend on external APIs over which it has no control, but its data structure forces the DASE architecture, which was built to manage flat data, to behave like a relational database. Furthermore, as long as content must be added painstakingly by hand by one or two editors, its possibilities for expansion are limited. Yet the information it contains has been collected with care and is of value to both archaeologists and the public. At the very least, shared information about periods could prevent other datasets from having to reinvent this particular wheel. By the same token, data associated with URIs for ancient sites through Pleiades, now being aggregated by the Pelagios project (<http://pelagios-project.blogspot.com/>), could significantly enrich the experience of GeoDia users hungry for additional information. We have taken initial steps to link and preserve our data by associating GeoDia sites with Pleiades URIs, by exposing GeoDia query results as JSON, and by making it possible to export filtered results sets as KML files. The next phase, however, should bring better integration with existing linked data sources. The Flickr machine tags now being used to associate images with Pleiades places, for example (Elliott 2011), could allow the same images to enrich, and be enriched by, associations with site periods in GeoDia.

At the moment, however, there is no corresponding gazetteer of scholarly assertions about periods that would allow images to be located

in time. This is one area in which the GeoDia dataset could be used for the good of the broader community. We are currently participating in an initiative to build such a gazetteer, which would create new possibilities for the linking and visualisation of geotemporal information about the ancient world. Rather than focusing on periods as objective things, this gazetteer would compile, like GeoDia, assertions made by specific authorities about the temporal and geographic extents of period terms (Fig. 8). Such an approach will sidestep the complications inherent in attempts to standardize period schemata, and will make it possible to compare information across multiple specialized thesauri. Most importantly, it will make the multivocality of archaeological interpretation even more transparent to students and to the public.

This is crucial at a time when tools for visualisation are becoming ever more compelling. One of the problems with the visualisation of time is the lack of conventions for the representation of vagueness and uncertainty (which, of course, pervade the study of the ancient world). The display of temporal information with a standard timeline or time-slider requires firm dates at each end of a range. More complex and uncertain relationships can be represented, but the tools necessary to analyze and visualise these relationships are better developed for the spatial sphere. In forcing our fuzzy information into clean displays, we create a false impression of certainty. As archaeologists, it is our task to provide a counterbalance to interfaces like Chronozoom (<http://www.chronozoomproject.org>), which, although very attractive, presents only a single, unambiguous (and currently somewhat inaccurate) representation of the periodisation of the ancient Mediterranean world. The risk of these exciting and intuitive tools is that they obscure for the public the chain of assumptions, argumentation, and source-criticism that underlies our understanding of the past. New tools for data visualisation are creating the foundations for richer approaches to the ancient world, but it is important to remember that a visualisation is only as good as the data that go into it -- and in the case of archaeology, those data reflect our own assumptions and intellectual traditions as much as they do the past itself.

References

- Ancient World Mapping Center and Institute for the Study of the Ancient World. 2012. "Pleiades. A community built gazetteer and graph of ancient places." Accessed November 3, 2012. <http://pleiades.stoa.org/>.
- Binding, C. 2010. "Implementing Archaeological Time Periods Using CIDOC CRM and SKOS." In *The Semantic Web: Research and Applications. Proceedings of the 7th Extended Semantic Web Conference, ESWC 2010, Heraklion, Crete, Greece*, edited by L. Aroyo, G. Antoniou, E. Hyvönen, A. ten Teije, H. Stuckenschmidt, and L. Cabral, 273–287. Berlin and Heidelberg: Springer Verlag.
- Bodenhamer, D. J., J. Corrigan, and T. M. Harris, eds. 2010. *The Spatial Humanities: GIS and the Future of Humanities Scholarship*. Spatial Humanities. Bloomington: Indiana University Press.
- Coller, M. 2009. "SahulTime: Rethinking Archaeological Representation in the Digital Age." *Archaeologies* 5 (1): 110-123.
- Elliott, T. 2011. "Horothesia: Pleiades, Flickr, and the Ancient World Image Bank." *Horothesia*. Accessed June 25, 2011. <http://horothesia.blogspot.com/2011/12/pleiades-flickr-and-ancient-world-image.html>.
- Fazal, R. 2009. "Models for complex spatio-temporal relationships and their implementation using Open Source components." *CAA 2009 Online Proceedings*. Accessed September 9, 2012. http://www.caa2009.org/articles/Fazal_Contribution378_c%20%282%29.pdf.
- Ferster, B. 2011. "Teaching using visualization." *Discovery Exhibition*. Accessed September 9, 2012. <http://www.discoveryexhibition.org/pmwiki.php/Entries/2011Ferster>.
- Gregory, I. 2007. *Historical GIS: Technologies, Methodologies, and Scholarship*. Cambridge, New York: Cambridge University Press.
- Hypermedia Research Unit, University of Glamorgan. 2012. "Handling Time Periods in STAR, Hypermedia Research Unit, University of Glamorgan." Accessed June 23, 2011. <http://hypermedia.research.glam.ac.uk/kos/star/time-periods/>.
- Johnson, I. 1999. "Mapping the Fourth Dimension:

- The TimeMap Project." In *Archaeology in the Age of the Internet*, CAA97, edited by L. Dingwall, S. Exon, V. Gaffney, S. Laflin, and Martijn van Leusen, 82. Oxford: Archaeopress.
- Johnson, I. 2008. "Mapping the Fourth Dimension: a Ten Year Retrospective." *Archeologia e Calcolatori* (19): 31-44.
- Keane, P. "DASe." Accessed November 3, 2012. <http://code.google.com/p/dase/>.
- Kleiner, F. S. 2007. *A History of Roman Art*. Victoria, Belmont: Thomson/Wadsworth.
- Koussoulakou, A., M. Karantoni, Y. Mitzias, and K. Efkleidou. 2011. "The heritage of archaeological mapping in Greece and some tools for a digital approach." *e-Perimtron* (6:3): 152-160.
- Malkin, I. 2011. *A Small Greek World: Networks in the Ancient Mediterranean*. Greeks Overseas. New York: Oxford University Press.
- Nowviskie, B. 2010. "'Inventing the Map' in the Digital Humanities: a Young Lady's Primer." *The Journal of the Initiative for Digital Humanities, Media, and Culture* (2:1). Accessed November 3, 2012. <http://journals.tdl.org/paj/index.php/paj/article/view/11>.
- Nowviskie, B., and D. Porter. 2010. "The Graceful Degradation Survey. Managing Digital Humanities projects through times of transition and decline." *Digital Humanities 2010*. Accessed November 3, 2012. <http://dh2010.cch.kcl.ac.uk/academic-programme/abstracts/papers/html/ab-722.html>.
- Pedley, J.G. 2012. *Greek Art and Archaeology*. Boston: Prentice Hall.
- Pelagios Project. "Pelagios." Accessed November 3, 2012. <http://pelagios-project.blogspot.com/>.
- Rabinowitz, N. "Timemap." Accessed November 3, 2012. <http://code.google.com/p/timemap/>.
- Rosenberg, D., and A. Grafton. 2010. *Cartographies of Time*. New York: Princeton Architectural Press.
- Scheidel, W., and E. Meeks. 2012. "ORBIS: The Stanford Geospatial Network Model of the Roman World." Accessed June 23, 2011. <http://orbis.stanford.edu/>.
- Scholars' Lab. "Neatline". Accessed November 3, 2012. <http://neatline.org/>.
- Stokstad, M., and M.W. Cothren. 2010. *Art History, Volume 1, 4th edition*. Upper Saddle River: Prentice Hall.
- Talbert, R.J.A., ed. 2000. *Barrington Atlas of the Greek and Roman World*. Princeton: Princeton University Press.
- The British Museum. Accessed November 3, 2012. "Portable Antiquities Scheme." <http://finds.org.uk/>.
- The Outercurve Foundation. "Chronozoom Beta." Accessed November 3, 2012. <http://www.chronozoomproject.org>.
- University of Oxford. "Ancient Lives." Accessed November 3, 2012. <http://ancientlives.org/>.
- UT Liberal Arts Instructional Technology Services. "GeoDia." Accessed November 3, 2012. <http://geodia.laits.utexas.edu>.
- Wikimapia. "Priene." Accessed November 3, 2012. <http://wikimapia.org/#lat=37.6610827&lon=27.2968894&z=16&l=o&m=b&show=/3653339/Priene&search=priene>.

Dating Historical Rock Art on Marble Surfaces by Means of a Mathematical Model for Natural Erosion Processes

Paolo Emilio Bagnoli
University of Pisa, Italy

Abstract:

The present research was motivated by the need to date rock-art engravings discovered in the Apuan Alps mountain chain (northwestern Tuscany, Italy). The engravings, which were made on the horizontal surfaces of marble, are contour-line etchings of billhooks (Italian “pennato”), whose chronology and meanings are still unknown. In this paper we present a computer-aided mathematical simulation of the effects of natural erosion processes on an ancient linear etching engraved on a flat horizontal marble surface. The main goal is to demonstrate that such engravings may still be visible after centuries-long exposure to natural erosion. The mathematical core of the simulation program is the widely used Monte Carlo method, based on continuous repetition of microscopic stochastic events whose probability laws are known. Formulating such laws for the task at hand requires knowing the speeds of the two main mechanisms involved in the erosion of limestone: freeze-thaw cycles and chemical dissolution. By applying the procedure developed, it has been possible to observe the evolution over time of the engravings’ cross sections during a time span of about 2000 years and thereby evaluate the values of the depth and width of the grooves forming the engravings. Analysis of these values using an estimation of the average annual rainfall has enabled formulating an algorithm to calculate the absolute dating of the engravings. The results of a first experimental application of the present method on the so-called “Billhook Step” (Mount Gabberi, Apuan Alps) are presented and discussed, including the influence of both random and systematic errors.

Keawords:

Absolute Dating, Rock Art, Historical, Marble

1. Introduction

In the last fifteen years several rock-art sites have been discovered in the mountain chain of the Apuan Alps (the famous white marble mountains) along the northwestern coast of Tuscany. These rock-art sites bear figurative engravings depicting billhooks (Italian “pennato” or “roncola”, a sort of curved hand knife used for woodcutting) as their most often reoccurring subject. The billhook engravings, which were etched out life-size, though in contour only, and are extremely worn, are found mainly on horizontal flat marble surfaces at high positions overlooking the lower peaks. Recent studies (Bagnoli, Panicucci and Viegi 2005) have come to the conclusion that the chronology of these markings is highly uncertain because of the lack of comparisons with other sites elsewhere in Italy (with the exception of two billhook representations

in the southern part of Trentino (Riva del Garda and Arco, Italy).

The local literature contains various, contrasting hypotheses regarding interpretation of the markings and their chronology. Dating estimates range from the Celtic-Etruscan age, to the Roman period, and even attribution to the modern age (as some maintain that such figures would not be preserved long on limestone), while their significance has been hypothesized to relate to the cult of the god Silvanus (Citton and Pastorelli 1995), (Citton and Pastorelli 2001), (Sani 2006), or to some generic relationship with ‘pastoral’ culture (Guidi 1992).

The research presented herein was prompted precisely by this uncertainty in the engravings’ dating. Its aim is to add further objective data to the discussion about the origins and meaning of these

Corresponding author: p.bagnoli@iet.unipi.it

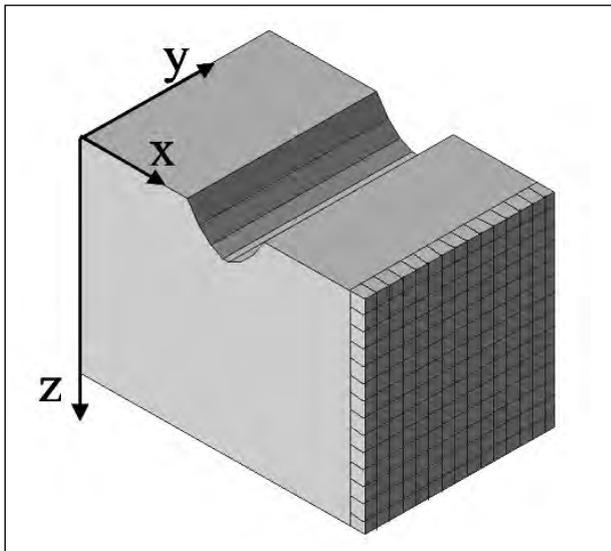


Figure 1. View of the solid used for the simulations and its coordinate system.

artifacts by analyzing the processes of erosion of the limestone of the Apuan Alps and the evolution over time of the cross section of engraved markings.

The time evolution of a small groove etched into stone is studied with the aid of computer simulations using the so-called Monte Carlo method. This mathematical simulation technique enables studying the macroscopic characteristics of a granular system (i.e. composed of a very large number of microscopic elements), by applying a continuous repetition of stochastic events at the microscopic scale, such as for instance, the removal of minute marble particles, whose probability laws are presumed to be known.

Using the above method, which requires as input data the average erosion speed due to the natural physical-chemical processes affecting marble (in particular, freeze-thaw cycles and chemical dissolution of calcium carbonate into soluble calcium-bicarbonate) and the overall average annual rainfall, it is possible to obtain the evolution over time of a small etched groove over a time span of 0 – 2000 years, and therefore to determine the mathematical trends of the variations in the two main dimensions of the groove: width and depth.

As will be shown later, knowing the trend laws and using accurate experimental measurements of the current values of the markings' dimensions,

has allowed formulating an algorithm that can determine the time elapsed since the marking was made and therefore provide an absolute dating for the rock art. In the following, the simulation procedure is explained in detail, and the results of a first application of the method to one of the most important Apuan rock-art sites, the "Billhook step" (Il "Ripiano dei Pennati", Mount Gabberi, located on the western side of the Apuan Alps), are presented and discussed in depth.

2. The Simulation Procedure

The marble block to which the simulation procedure was applied, and whose structure is depicted in Figure 1, is made up of a regular net of elementary cubic cells representing the smallest material clusters subject to erosion. Since the lateral dimensions ($L_x, L_y=15$ cm) of the block can be chosen arbitrarily, the single cell size depends upon the number of cells (n_x, n_y, n_z) along the x, y and z directions. The horizontal top surface of the block is exposed to the environment above, while the bottom must be considered unlimited(?)infinite). Cylinder-like boundary conditions exist on the lateral sides along the x and y directions. This means that, with x_i, y_i and z_i the coordinates of the centers, the cells placed on the lateral faces having $x_i=1$ or $x_i=n_x$ may be considered to be in contact with those on the opposite face and have x coordinates $x_i=n_x$ or $x_i=1$ respectively. On the top surface of the block a small parabola-shaped linear groove was etched, with initial width w_0 and depth h_0 , along the y direction in the center of the x side.

For the sake of clarity, we define the width of the engraving as the distance between the groove borders, measured on the horizontal plane of the un-etched marble surface and orthogonal to the groove direction, while the depth of the engraving is the maximum vertical distance of the groove from such plane.

The model geometry chosen, the cylindrical boundary conditions in the y direction, and the hypothesis that the erosion mechanisms may be deemed uniform along the y direction, enable minimizing the number of cells in the y direction and thereby minimize the computation time required for the simulation procedure ($n_x=1000; n_y=5$).

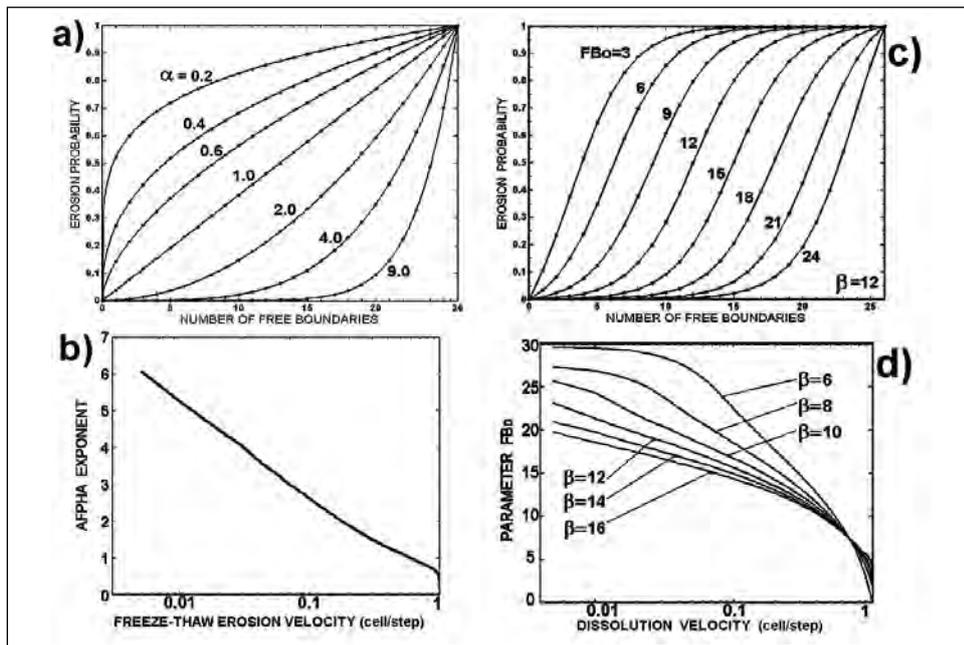


Figure 2. a) Plots of the P_a probability functions vs. the number of free boundaries, FB , for several values of α . b) Freeze-thaw erosion speed vs. exponent α . c) Plots of the P_b probability functions vs. the number of free boundaries, FB , for several FBo values. d) Chemical erosion speed curves vs. parameter FBo .

In a regular cubic network, each cell is directly bound to 26 surrounding cells, including those in the diagonal directions. If a cell has all 26 surrounding boundaries occupied, it means that this cell is located within the mass of the material. If some of the boundaries are unbound, it implies that the given cell is located close to the top surface.

From the mathematical perspective, such a solid body may be described by a pair of three-dimensional matrices for which the three element indices, ix , iy and iz , denote the position of the cells in three-dimensional space. In the first binary matrix, named AL (alive), the elements take the value 'true', if the corresponding cell is still connected to the solid, or 'false' if it has been previously eroded away. The second integer matrix, named FB (free boundaries) contains the number of unbounded boundaries for each cell.

The simulation of the marble erosion process as a function of time was conducted under the following hypotheses. A) The original top surface of the rock is perfectly horizontal. B) Only the two aforementioned erosion mechanisms are taken into account, which implies that any other source of rock deformation, such as mechanical abrasions due to ice motion, is excluded. C) The intensity of the erosion mechanisms is uniform along the entire upper surface of the sample block.

The simulation procedure is divided into 2000 time steps: the duration of one step is, by convention, set to one year. Furthermore, within a single time step the two erosion processes, which in reality actually occur simultaneously, are applied in sequence: the first one due to the freeze-thaw cycles, while the second one from chemical dissolution of calcium carbonate. Applying the two processes sequentially instead of simultaneously seems a reasonable approximation and greatly simplifies the simulation procedure.

Virtual surface erosion was achieved using the following method. Two probability distributions $P_a(FB)$ and $P_b(FB)$ were defined as a function of the number of free cell boundaries for the two erosion processes, respectively. Both functions P_a and P_b , which increase with FB values, must be chosen with the following properties:

$$\begin{aligned}
 P(FB) &= 0 \text{ for } FB = 0 \\
 0 < P(FB) < 1 &\text{ for } 0 < FB < 26 \\
 P(FB) &= 1 \text{ for } FB = 26
 \end{aligned}
 \tag{1}$$

In a single time step, for each 'alive' cell close to the top surface, i.e. whose $AL(ix, iy, iz)$ value is 'true' and $FB(ix, iy, iz)$ value is greater than zero, the computer generates a random number S in the range 0-1 using a uniform stochastic distribution. Whether the cell is removal or not is determined by

the following conditions:

$$\begin{aligned} \text{If } S &\leq P[FB(ix, iy, iz)] \\ \text{then } AL(ix, iy, iz) &= 0 \\ \text{If } S &> P[FB(ix, iy, iz)] \\ \text{then } AL(ix, iy, iz) &= 1 \end{aligned} \quad (2)$$

After the erosion in the single time step is complete, both matrices AL and FB are updated.

The probability function Pa for the freeze-thaw erosion process was defined as follows:

$$\begin{aligned} Pa(FB) &= (FB/26)^\alpha \\ 0 &< \alpha < \text{infinite}, \end{aligned} \quad (3)$$

where the exponent α indicates the concavity of the curve: upward if $\alpha > 1$ and downward if $\alpha < 1$. Figure 2a shows the probability distribution Pa as a function of the number of free boundaries, FB , for several values of exponent α . Note that this parameter is directly related to the average erosion rate, Va' (measured in cells/step), so that the amount of material eroded in a single time step can be finely modulated by changing parameter α in the probability function. The dependence of erosion speed on the exponent is shown in Figure 2b, whose curve has been calculated from many simulation tests performed on a small-sized flat marble block using only the Pa function with several values of its exponent.

The shape of the second probability function Pb for the calcium-carbonate dissolution process was defined starting with the inverse of the well-known Fermi-Dirac function, governing the electron energy occupation probability in a semiconductor crystal. Pb can be calculated from the following function normalized to $FB=0$ and $FB=26$ in order to precisely satisfy the rules defined in equation (1):

$$\eta^3 + \frac{d}{b} \eta^2 - \frac{a+b \cdot h}{b \cdot K_1} \eta + \frac{w-c-d \cdot h}{b \cdot K_1} = 0 \quad (4)$$

The above function has two different parameters: b and FBo . The first (b) governs the shape of the curve, from a sharp vertical step (for b tending towards infinity), to a line (for b tending towards zero). The second parameter FBo can be defined as the

number of free boundaries for which $f(FBo)=0.5$. The resulting Pb function is shown in Figure 2c as a function of the free boundaries, FB , for $b = 12$ and for several values of parameter FBo . Also in this case, the average erosion speed due to the calcium-carbonate dissolution process Vb' (measured in cells/step) is directly related to the shape of the Pb curve and, in particular, to the FBo value, as can be seen from the plots in Figure 2d, where parameter FBo is plotted as a function of the erosion speed Vb' (cells/step) for several values of parameter b . As before, these curves were calculated by performing several simulation tests on a small flat marble block using only the Pb distribution.

The parameters characterizing the two probability distribution functions, α for Pa and FBo for Pb , were chosen considering the average erosion speed values, Va and Vb (measured in mm/year), found in the literature for environmental conditions similar to those of the rock under study. The corresponding velocities Va' and Vb' measured in cells/step can be simply obtained via the following relations:

$$\begin{aligned} Va' &= Va \times (nx Lx); \\ Vb' &= Vb \times (nx Lx), \end{aligned} \quad (5)$$

where the ratio Lx/nx is the lateral size of the cubic cells. The parameters of the two distribution functions were chosen using the plots in Figures 2b and 2d, respectively, as look-up tables.

3. Erosion Speed Correction Factors

According to hypothesis C (i.e., that both erosion mechanisms are uniform in intensity), the speeds, and hence the probability distributions of both erosion components, should always be constant for all exposed cells. This assumption may actually be quite unrealistic given the small dimensions in question, since local morphological factors may concur to modify the erosion process. Therefore, in each time step the Va and Vb speeds – and consequently the parameters of the probability functions – must be adjusted using suitable correction functions that depend on the depth of the cells within the engraved groove and the local morphology of the exposed upper surface.

In particular, the erosion process due to freeze-thaw cycles may be considered as slightly

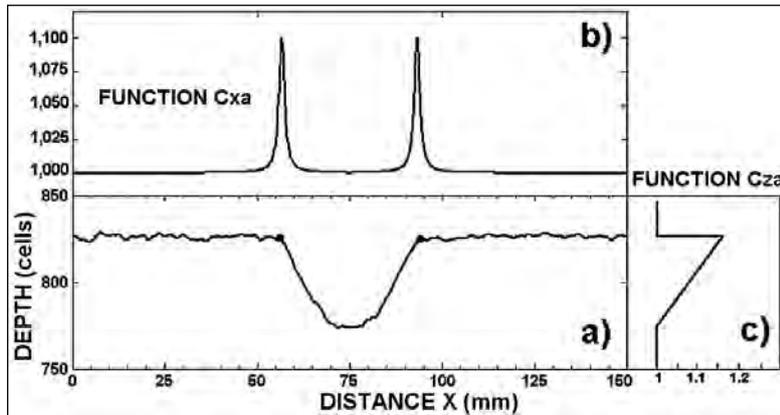


Figure 3. a) Horizontal and vertical correction functions, Cxa and Cza , for freeze-thaw erosion. b) Absolute and c) normalized chemical dissolution speed as functions of the annual rainfall for several values of the surface angle with respect to the horizontal plane.

more intense at the top surface of the engraved groove than at the bottom, and the percentage increase must be referred to the original depth of the small groove. This process may be described by a correction function $Cza(iz)$, which depends only on the z spatial variable of the cells. It must moreover be emphasized that the borders of the groove are subject to more intense freeze-thaw erosion than flat surfaces. Adjustment for this second source of divergence from hypothesis C can be made by introducing a further correction function $Cxa(ix)$, which depends solely on the x spatial variable. However, this implies that the program must be able to localize the groove borders after each simulation time step, which can be accomplished by using a suitable numerical algorithm. Therefore, for each cell, velocity Va , and consequently, velocity Va' and parameter a of the probability function Pa , must be corrected using the following relation:

$$Va(ix, iy, iz) = Vao \times Cxa(ix) \times Cza(iz), \quad (6)$$

where Vao is the standard value of the erosion speed. By way of example, Figure 3a shows the cross-sectional surface under simulation at a given time, together with horizontal and vertical plots of the $Cxa(ix)$ and $Cza(iz)$ functions, respectively. The maximum values for the two functions are open parameters of the simulation tests and in any case do not exceed 20% of the standard value.

Regarding the calcium-carbonate dissolution process, a large body of work can be found in the

literature concerning theoretical and experimental studies of erosion on limestone rock surfaces. In particular, a recently published study (Szunyogh 2005) describes an accurate mathematical model for carbonate-based rock dissolution, which takes into account many environmental parameters, such as annual rainfall and the slope of the rock surface with respect to the horizontal plane. According to the above model, and assuming vertical rainfall, the surface loss rate, W , can be written in compact form as follows:

$$W = (g \times Qa) / [u + Qa \times \cos(J)], \quad (7)$$

where Qa is the annual precipitation, and J is the local inclination angle of the surface with respect to the horizontal. Parameters g and u are defined by the following relations:

$$g = (k \times Ceq \times r_{rock}) \times td \times SH \times Na \times Mmm \quad (8)$$

$$u = k \times td \times SH \times Na, \quad (9)$$

where k is the velocity constant of chemical dissolution at 10 °C under open-air carbon dioxide content, Ceq is the equilibrium concentration of calcium carbonate, r_{rock} is the density of stone, td is hours/day of rainfall and SH (3600 sec/hour), Na (365 days/year) and Mmm (1000 mm/m) are suitable constants applied to convert the parameter dimensions to those used in everyday practice.

From equation (7) it can be clearly seen that velocity W is a function of the local slope of the rock surface. Figure 3b shows the plots of W versus annual precipitation (assuming $td = 10$ hours/day) for several values of the angle q , while Figure 3c shows the same plots normalized with respect to the curve with $J = 0^\circ$. Based on the above model, it can be concluded that the velocity of dissolution of a sample under simulation in each time step is a function of the local slope of the exposed surface. Therefore, although most of the original surface is nearly flat, the walls of the engraved groove undergo increased erosion with respect to the standard situation. Once a suitable value for annual precipitation has been

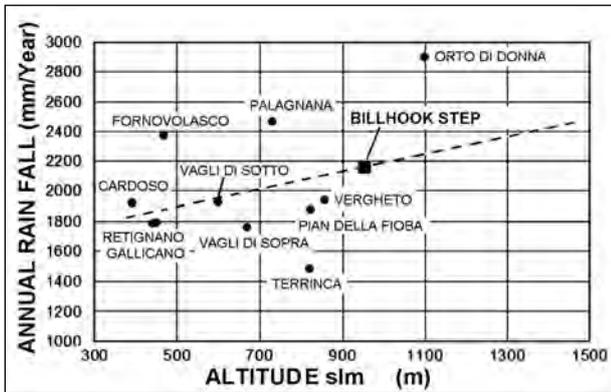


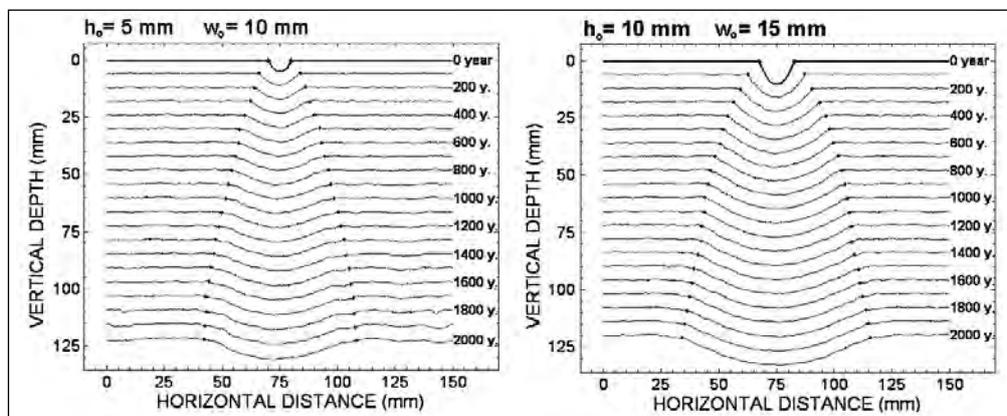
Figure 4. Annual rainfall data, from measurement stations of the Apuan Alps, as a function of altitude.

chosen, at the end of each simulation time step the local slope can be calculated from the absolute value of the first derivative of the surface profile along the x axis. Naturally, in order to eliminate the effects of surface roughness on the derivative function, the surface profile is previously smoothed using a mobile averaging filtering procedure. These data, interpolated within the plots of Figures 3b and 3c, allow us to calculate the horizontal correction function $Cxb(ix)$ for velocity Vb .

The corresponding vertical correction function $Czb(iz)$ for the dissolution process may be formulated in a similar way as that for the freeze-thaw cycles, though in this case, because of the accumulation of water, the standard value is located at the upper surface, while the erosion velocity is higher at the bottom of the engraving. Therefore, the local dissolution velocity Vb can be similarly expressed by the following relation:

$$Vb(ix, iy, iz) = Vbo \times Cxb(ix) \times Czb(iz), \quad (10)$$

Figure 5. Two simulations of the evolution of engraving profiles plotted in isometric scales for every 100 years.



where Vbo is the standard value calculated for $J = 0^\circ$.

4. Simulation Results

The values of the rock erosion rates for the two individual processes have been deduced from the data found in the literature. In particular, (Buhmann and Dreybrodt 1985) reported the average degree of erosion loss due to a single freeze-thaw cycle for several types of rock material. For carbonate limestone the value is about 5.5×10^{-5} mm. Using this value, and considering as a first approximation one cycle per day during four months per year, yields an average value of 0.010038 mm/year for the erosion speed, Va , corresponding to a Va' value equal to 0.0669 cells/year in the present simulation model ($Lx=150$ mm; $nx=1000$).

As far as the chemical dissolution rate is concerned, we used the data reported in (Szunyogh 2005) for the considered physical parameters to calculate equations (8) and (9). However, as shown in equation (7), the most important input parameter for the simulation procedure is the average annual rainfall, Qa . Lacking accurate scientific data from geological analyses on the local evolution of this parameter dating back 2000 years, the annual rainfall was provisionally deduced from historical data, spanning about 80 years, drawn from up to eleven meteorological measurement stations located at various different altitudes around the central massif of the Apuan Alps.

Figure 4 shows a plot of the Qa values for the various stations as a function of altitude; the dashed line represents the linear regression of the dataset.

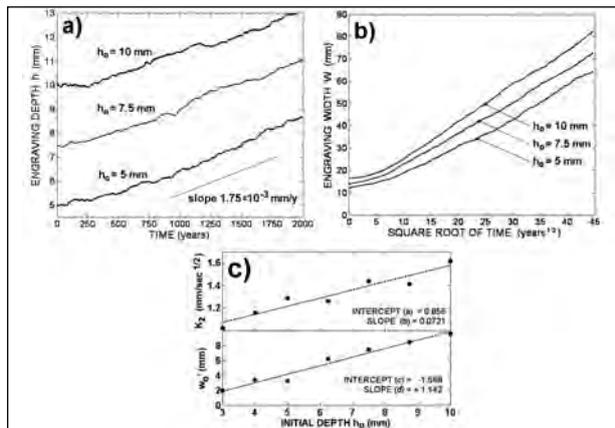


Figure 6. Plots of groove depth (a) and width (b) vs. time and square root of time for three values of initial depth h_0 . c) Fitting plots for calculation of parameters a , b , c and d in equations (11b).

The value used as the most probable one for the rock-art site under investigation was interpolated along the regression line and turned out to be 2150 mm/year. Consequently, the mean chemical dissolution velocity, Vb , for the horizontal condition ($q=0^\circ$) was calculated from equation (7) as 0.0459 mm/year, corresponding to $Vb'=0.3066$ cells/year.

Using the above input parameters, several simulation runs were performed on various models in which the initial depth (h_0) and width (w_0) of the engraved parabolic groove varied in the following ranges: $h_0=3 - 10$ mm and $w_0=5 - 22.5$ mm. Figure 5 shows an isometric plot of the rock surface cross sections along the x -direction of two sample models drawn every 100 years. The small dots indicate the borders of the groove as detected by the internal algorithm. As can be seen from this figure, the engraved markings are still observable after longstanding erosion, despite the progressive widening of the groove, smoothing of its borders and lowering of the top surface.

From the results of the simulations performed, a general trend in the evolution of the engravings may be detected, as shown by the plots in Figure 6a and 6b. In order to smooth the statistically induced roughness of the surface, each curve in these figures was obtained by averaging the plots of ten different simulations with the same parameters.

As can be clearly seen, the groove depth tends to increase slowly and linearly over time, and the

increasing speed ($K1$) was found to be independent of the original depth h_0 . Instead, the width evolves linearly with the square root of time (except in the initial stage), and the slope ($K2$) and intercept (w_0') are increasing functions of the original depth h_0 . Such behavior can be described as a function of time t by means of the following set of equations:

$$h = h_0 + K1 t ;$$

$$w = w_0' + K2 t^{1/2} , \quad (11a)$$

$$K2 = a + b \times h_0 ;$$

$$w_0' = c + d \times h_0 , \quad (11b)$$

where parameters a , b , c and d are constants that can be calculated from the plots of equations (11b) as a function of the initial depth h_0 , as shown in Figure 6c. By replacing the square root of time with the variable $x = t^{1/2}$ and the known current values of a engraving's width w and depth h , the time elapsed since an engraving was made can be obtained as the positive real root of the following third order equation:

$$f(FB) = 1 - \frac{1}{1 + \exp\left[\frac{\beta \cdot (FB - FB_0)}{26}\right]} \quad (12)$$

As a preliminary test of the algorithm described in the foregoing, a self-dating procedure was applied to the two simulation curves shown in Figure 5. For each time step (year), the height h and width w values obtained were used as the input data for equation (12), while the trend parameters a , b , c and d were calculated by averaging several simulation tests performed using the same h_0 and w_0 parameters and the same average annual rainfall Qa . The output value of the elapsed time was compared with the available initial value.

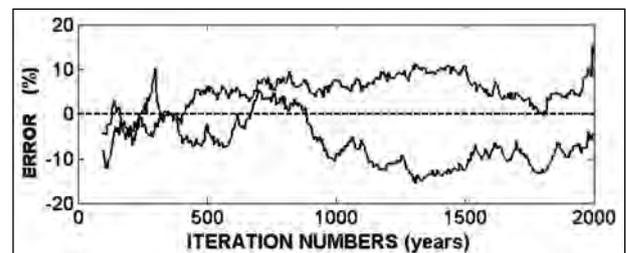


Figure 7. Percentage error between estimated and real time values for the simulation results in Figure 5.

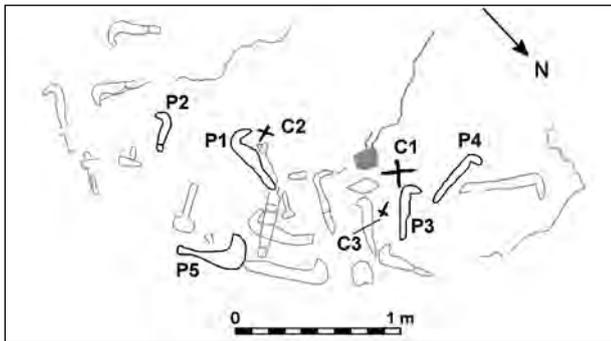


Figure 8. Map of the engravings in the Billhooks Step (mount Gabberi, Camaio). The billhook markings used for the dating experiment are drawn in black and labeled.

The results of this test are shown in Figure 7, where the relative percentage error between the actual elapsed time and the back calculated one using the dating algorithm is plotted as a function of time. As can be seen, the dating algorithm manages to estimate the engraving's age with a relative error in the range of 10-20%, at least under the specific test conditions and using a single data pair for h and w . This first result is quite encouraging for rupestral archaeology, which for the most part uses cultural and morphological considerations for chronological analyses.

However, it must be pointed out that the curves in Figure 7 are affected by random error only, having a mean value of zero. Such error is due to the granular nature of the phenomenon, which is responsible for the discrepancies between the real deformation curves and the corresponding average trends. In order to increase the accuracy of the evaluation, at least in terms of correcting for random error, the algorithm would have to be applied to several experimental measurements on the same sample (i.e. several pairs of h and w values measured on the same marking) and the output values averaged. The second source of error, which does not affect the plots in Figure 7, is the "systematic" one introduced by inaccuracies in the main parameters governing erosion speed (e.g., average annual rainfall), which consistently produce the same error in all the output data.

5. In Situ Experiment

The method described in the foregoing was applied to the rock-art site called the "Billhooks

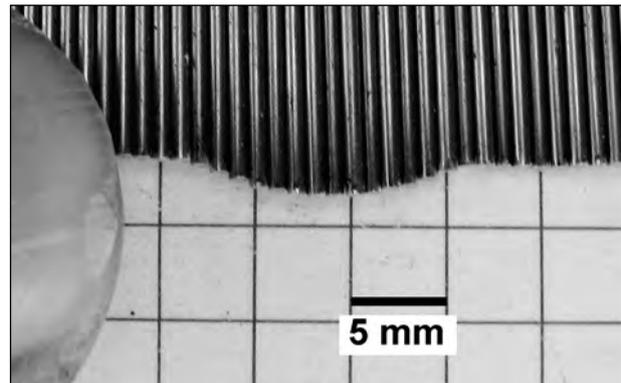


Figure 9. High resolution digital photograph used for experimental measurement of groove depth and width.

Step" (Italian "Ripiano dei Pennati") on the western side of Mount Gabberi (Camaio, Lucca). This flat rock, located on the top of a sharp cliff overlooking the seacoast at an altitude of 960 meter above sea level, is one of the sites described as having properties that, according to specific criteria, enable assigning it a high chronology (Bagnoli, Panicucci and Viegi 2005).

A complete map of the rock-art site is presented in Figure 8, which shows the location of several billhooks engraved into the stone surrounding a small rectangular tray and three crosses, two Greek (i.e. equal length arms) and one Latin. The *in situ* tests, which consisted of measuring several of the profile grooves in different markings, were carried out by the author with the invaluable aid of some volunteers from the Pietrasanta Archaeological and Speleological Group (GASP).

The dating method was applied to a set of five billhook markings (P1-P5), as well as the three crosses (C1-C3); Figure 8 indicates the selected markings drawn in black lines with the corresponding label.

The cross sections of the engraved markings were measured using a steel "needle" profilometer (1 mm diameter pins) and photographed using a high resolution digital camera with a paper grid in the background (Fig. 9). The digital image was then uploaded into a vector graphics program (e.g., AutoCAD) in order to easily measure both the height and width within an estimated resolution of less than 100 microns. Up to ten different profiles were measured for each tested marking to obtain an average value.

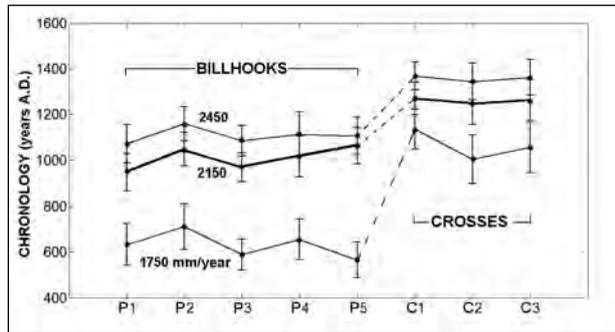


Figure 10. Plots of the chronologies of the eight markings for three different values of average annual rainfall (vertical bars indicate standard deviations).

Figure 10 Shows the results obtained in graphical form for easy reading. It plots the absolute date with respect to the year of the measurement (2006) for the five billhooks and the three crosses. The vertical bars on each point represent the standard deviations (random error). Moreover, in order to clearly demonstrate the sensitivity of the output results on the annual rainfall value utilized, Figure 10 reports the output data obtained for two other different Q_a values (1750 mm/year and 2450 mm/year) for comparison with those calculated using the most probable value (2150 mm/year) for Q_a .

Careful analysis of the results obtained reveals some important, particularly noteworthy details. Firstly, as anticipated, the dates are progressively more recent with increasing average annual rainfall, simply due to the higher speed of the chemical dissolution, which, at least at the altitude considered, predominates strongly over the effects of the freeze-thaw cycles. The second finding that should be stressed is the substantially uniform dating between the two separate groups of markings: the chronologies cluster around 1000 BC for the billhooks and about three centuries later for the crosses (at least for annual rainfall of 2150 mm/year), hence clearly indicating the medieval origin of the artifacts.

However, the fact that the dates for the two groups cluster around the same value does not necessarily imply that the markings were made at the same time - the uncertainty in the individual datings is about a century. Such an approximation is consistent with the random error from the combined effects of the stochastic granular erosion phenomena

and the different shapes of the small grooves in the various engravings' cross sections.

Finally, some words are in order regarding the large differences in dating between the billhooks and the crosses (even ignoring the results for annual rainfall of 1750 mm/year, which is very high). As such differences are much higher than the uncertainty due to random error, they seem to be due to historical or cultural causes rather than numerical ones.

6. Conclusions

The present work has presented a new method for absolute dating of rock art based on mathematical simulations and careful *in situ* experimental measurements. Despite the uncertainty affecting the input data required for the mathematical analyses, we believe that the proposed method is very promising for obtaining reliable absolute dates for rock art, at least under the specified study conditions, namely, markings etched in contour on the horizontal flat surfaces of limestone rock under conditions of open-air exposure.

Furthermore, the chronologies obtained by experimental application of the dating method, in particular the fact that the dating of the crosses indicates that they are more recent than the billhooks, appear consistent with the occurrence of so-called "Christianization" of rock art once used, or deemed to have been used, by pagan peoples for their rituals - a process that typically occurred precisely in the early centuries of the Middle Ages.

It must also be borne in mind that the results obtained for the Billhook Step may still be affected by systematic error due to the uncertainties in some of the input parameters, especially the average annual rainfall. Nevertheless, we believe that the results obtained from the chronological analysis of the engravings is substantially consistent and tentatively demonstrates the validity of the proposed method.

Further research on the proposed dating method should aim at appropriate validation via application in other rock-art sites whose chronology may be determined with certainty by alternative methods. Other future improvements to

the method should also include refinement of the annual rainfall data by investigating the results of geological research in the Apuan Alps, enhancement of the simulation procedure to include long-term fluctuations in annual rainfall, and the possible use of more sophisticated techniques for measuring the engraving profiles, such as a portable laser interferometer.

References

- Bagnoli, P.E., N. Panicucci, and M. Viegi. 2005. "Manifestazioni di arte rupestre figurativa sulle Alpi Apuan Centrali." *Acta apuana* IV-V: 105-116.
- Bednarick, R. G. 1979. "The potential of rock patination analysis in Australian archaeology – part 1." *The Artefact* 4: 14–38.
- Buhmann, D., and W. Dreybrodt. 1985. "The kinetics of calcite dissolution and precipitation in geologically relevant situations of karst areas. 1. Open system." *Chemical Geology* 48: 189–211.
- Citton, G., and I. Pastorelli. 1995. *Incisioni rupestri sulle Alpi Apuan in alta Versilia*. Massarosa, Lucca: Pacini.
- Citton, G., and I. Pastorelli. 2001. *I monti "scritti"*. Viareggio, Lucca: Baroni.
- Dewdney, S. 1970. "Dating rock art in the Canadian Shield Region." Accessed January 2013. <http://www.archive.org/stream/datingrockartincoodewd#page/n5/mode/2up>.
- Guidi O. 1992. *Incisioni rupestri della Garfagnana*. Lucca: Pacini Fazzi.
- Sani G. 2006. "Le incisioni rupestri dei Pennati sulle Alpi Apuan e il culto del dio Silvano." Accessed January 2013. <http://www.symbolisullarocchia.it/archivio/lelamepennate/Lelamepennate.pdf>.
- Svensson, U., W. Dreybrodt. 1992, "Dissolution kinetics of natural calcite minerals in CO₂-water systems approaching calcite equilibrium." *Chemical Geology* 100: 129–145.
- Szunyogh, G. 2005. "Theoretical investigation of the duration of karstic denudation on bare, sloping limestone surface." *Acta Carsologica* 34/1 (1): 9-23.

Cultural Heritage Application Schema: a SDI Framework within the Protected Sites INSPIRE Spatial Data Theme

Antonio Uriarte González, César Parcero Oubiña, Alfonso Fraguas Bravo, Pastor Fábrega Álvarez, Juan Manuel Vicent García, Esther Pérez Asensio, Carlos Fernández Freire, Isabel del Bosque González

Consejo Superior de Investigaciones Científicas, Spain

Abstract:

We present a Cultural Heritage data model built under the European INSPIRE Directive. This model extends the Data Specification on Protected Sites –one of the reference spatial data themes of the INSPIRE Annex I– through the development of cultural issues by means of new classes and attributes. The aim is to achieve an interoperable schema that allows organising and sharing georeferenced cultural heritage information via Spatial Data Infrastructures. This involves the use of standards and norms about several topics, such as geographic information, cultural heritage and documentary resources. The data model includes three parts: 1) a legal part about the administrative definition of protected sites, 2) a cultural part devoted to the description of the cultural entities that are subject of valuation and protection, and 3) a documentary part for the inclusion of information resources (texts, images and so on) about these cultural entities.

Keywords:

Cultural Heritage, Spatial Data Infrastructure, INSPIRE, Conceptual Data Model, Interoperability.

1. Introduction: INSPIRE, Protected Sites and Cultural Heritage

This paper presents a proposal of application schema about cultural heritage under the INSPIRE Directive (Directive 2007/2/EC), which establishes an Infrastructure for Spatial Information in the European Community. We use the term *application schema* in the way INSPIRE does, as the formal description of a conceptual data model about a certain part of the real world.

This model develops the INSPIRE Data Specification on Protected Sites (INSPIRE Thematic Working Group Protected Sites 2010) for the topic of cultural features. According to INSPIRE principles, the background objective is to provide a basic interoperable framework for describing, organising and sharing georeferenced information to institutions devoted to the management and research of cultural heritage.

The INSPIRE Directive addresses several thematic areas of spatial information, fixed in 34

spatial data themes distributed in three annexes. One of these themes is *Protected Sites*, defined as “area designated or managed within a framework of international, Community and Member States’ legislation to achieve specific conservation objectives” (Directive 2007/2/EC, 11). It is included in Annex I, which means that it is considered as reference data and so it is priority in its development.

INSPIRE Data Specification on Protected Sites has been developed by the Thematic Working Group Protected Sites. It consists of two application schemas, *Simple* and *Full*. Although complex, they only fix a minimum part as compulsory: one class (*ProtectedSite*) and two of its attributes (*geometry* and *inspireID*), belonging to the simple schema.

In spite of their development, we have found that both schemas show limitations for the description of cultural issues, for they are mainly centred on legal aspects (McKeague, Corns and Shaw 2012) and, beyond those, on the description of natural protected sites (e.g. Natura 2000). Cultural heritage, as a thematic layer, could have been addressed to Annex III, but none of its spatial data themes is conceived to fulfill this issue –except

Corresponding author: antonio.uriarte@cchs.csic.es

partially *Buildings*, that might be extended to incorporate historical architecture, although it is not aimed at describing the historical or cultural dimension of constructions.

Therefore we thought it was desirable to make a specific development for cultural heritage information, so that anything else than mere legal questions could be incorporated. In the Spanish context this has been tackled within the Working Group of the Spanish Spatial Data Infrastructure², whose mission is to apply, develop and extend INSPIRE in the Spanish regulations. It is articulated by several workgroups devoted to specific issues. One of them is the Thematic Working Group on Cultural Heritage³, an interdisciplinary team formed by experts on Information and Communication Technologies, mainly about geographic information, and Cultural Heritage (for further details, see the acknowledgments at the end of the paper). Through 2010 and 2011, work was made on the development of a proposal of an interoperability framework for cultural heritage information (the data model is available on-line, see GTT-PAH 2011, and will be thoroughly published, see Fernández-Freire, Parcero-Oubiña and Uriarte-González *in press*). This paper represents a further and personal elaboration made by the authors.

2. Cultural Heritage Application Schema: Guidelines

The Cultural Heritage Application Schema, according to INSPIRE philosophy, is based on the following principles: *generic*, in order to embrace any kind of cultural heritage data, direct or indirectly georeferenced; *extendable*, to allow any kind of data producer to adapt the model to the nature of their own information; and *interoperable*, to combine spatial datasets from different sources through network services, via Internet.

For achieving these objectives, it has been fundamental the use of the following standards and norms, where applicable:

² Grupo de Trabajo de la Infraestructura de Datos Espaciales de España (GT-IDEE): <http://www.idee.es/grupo-de-trabajo-idee/> (Accessed June 20, 2012).

³ Grupo de Trabajo Temático de Patrimonio Histórico (GTT-PAH).

- Regarding geographic information, the INSPIRE Directive itself and also the ISO 19100 series, on which the former is based.
- The CIDOC Conceptual Reference Model (CIDOC CRM)⁴ (Doerr 2003; Crofts et al. 2010), also an ISO standard (ISO 21127:2006) (ISO/TC46/SC4 2006), designed for the description of concepts and relationships for the documentation of cultural heritage.
- The Dublin Core Metadata Initiative (DCMI) and its Dublin Core Metadata Element Set, established as an ISO standard (ISO 15836:2009) (ISO/TC46/SC4 2009), that defines a basic list of attributes for describing and cataloguing information resources.

As we have said before, the conceptual data model presented in this paper is expressed as an INSPIRE application schema, that contains the complete and precise definition of the content and structure of a dataset (in our case, georeferenced cultural heritage data). The application schema, as a formal description, has been fully developed through the Conceptual Schema Language (CSL), which is used for the data specifications of the INSPIRE Directive. This is a profile of the Unified Modelling Language (UML) version 2.0, established as an ISO standard (ISO/TS 19103:2005) (ISO/TC211 2005).

3. Cultural Heritage Application Schema: Description

The Cultural Heritage Application Schema extends the Protected Sites Full Application Schema by means of additional classes and attributes (Fig. 1). The extension, that is, the cultural heritage schema in strict sense, is formed by three main parts (Fig.2): 1) a legal or administrative part, 2) what we call a cultural part, and 3) a part devoted to documentary information. It is important to emphasize this distinction between cultural and legal entities, since it is one of the conceptual foundations of the model. It is based on the idea that, within the realm of cultural heritage, a difference must be told between real-world cultural entities, that are the result of human or natural-human past actions, and legal artefacts, that are the result of a process of heritage

⁴ CIDOC Conceptual Reference Model (CIDOC CRM): <http://www.cidoc-crm.org/> (Accessed June 20, 2012).

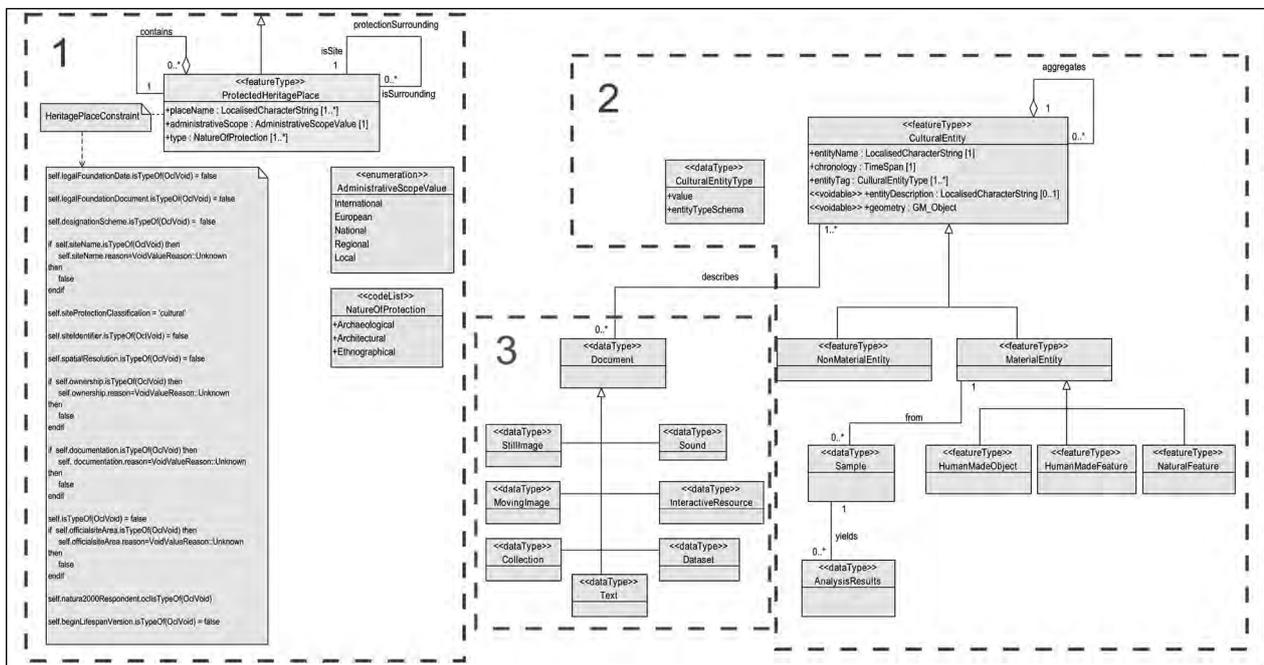


Figure 2. Cultural Heritage Application Schema: 1) Legal part. 2) Cultural part. 3) Document part.

entity can be modified without necessarily modifying the description of the legal entity related to it, and vice versa. When it comes to geographic information, the coherence and also the benefits of such approach are clear: for instance, the geometry (e.g. the extent) of a protected place may change (like being extended to accommodate changes in heritage management policies), while the geometry of the related cultural entity remains unchanged. Or inversely, the geometry of a cultural entity may change (such as a building losing some parts after a partial collapse) while the protected area around it remains unchanged.

In fact, the extended application schema for natural protected sites is designed in a similar way, separating protected sites as legal entities from the things protected, such as habitats, biotopes or distribution of species.

3.1 Legal part

The central class of this part is the Protected Heritage Place (*ProtectedHeritagePlace*), defined as an area dedicated to the protection of cultural resources and managed through legal and administrative means (Fig. 3). *ProtectedHeritagePlace* is a subclass of Protected

Site (*ProtectedSite*), the central class of the Protected Sites application schemas. Its specificity relies on the cultural character of the items to which the protection is applied.

ProtectedHeritagePlace inherits the attributes from the *ProtectedSite* class for generic protection topics and adds other specifically devoted to cultural heritage issues. In a effort to offer a minimum common information, some of these attributes –thirteen– have to be fulfilled: two are originally compulsory in the Protected Sites application schemas (*geometry*, *inspireID*); eight are voidable in the same schema but have been changed to mandatory through constraints (*legalFoundationDate*, *legalFoundationDocument*, *siteDesignation*, *siteProtectionClassification*, *siteIdentifier*, *spatialResolution*, *dataSource*, *beginLifespanVersion*); and finally three new attributes belong to the *ProtectedHeritagePlace* class (*placeName*, *administrativeScope*, *type*).

Next we describe each of these thirteen attributes, grouping them thematically in order to help their comprehension (for a full description of the attributes, voidable or not, of the *ProtectedSite* class, see INSPIRE Thematic Working Group Protected Sites 2010):

Attributes about identification:

inspireID: External object identifier of the protected heritage place, according to INSPIRE rules.

siteIdentifier: Identifier for the cultural heritage site using some national or international identification scheme. Contrarily to *inspireID*, which is unique, there can be several site identifiers for a same protected site, for each one can be assigned by a different administration.

Attributes about spatiality:

geometry: Spatial object defining the boundary of the protected heritage place. This boundary is that defined by the administration responsible for its protection and management.

placeName: Place name of the protected heritage place, that used in the legal foundation document (see below). This attribute is different to *siteName* in the *ProtectedSite* class, a voidable attribute intended to be filled with a preexisting geographical name. Place name attribute, on the contrary, is not taken from a predefined gazetteer, but is entered by the info provider.

spatialResolution: The spatial resolution of the cultural heritage place geometry, according to the way the data have been gathered (equivalent scale or distance).

Attributes about legal and administrative issues:

legalFoundationDocument: Reference of the legal act that created the protected heritage place. This consists of a web URL or a text citation.

legalFoundationDate: Date of legal creation of the protected heritage place. It should be that of the foundation document.

siteDesignation: Type or category of protected heritage place, according to a certain designation schema. A designation schema is a set of protection categories established by a certain institution in a specific field and for a specific purpose. Designation schemas are accommodated

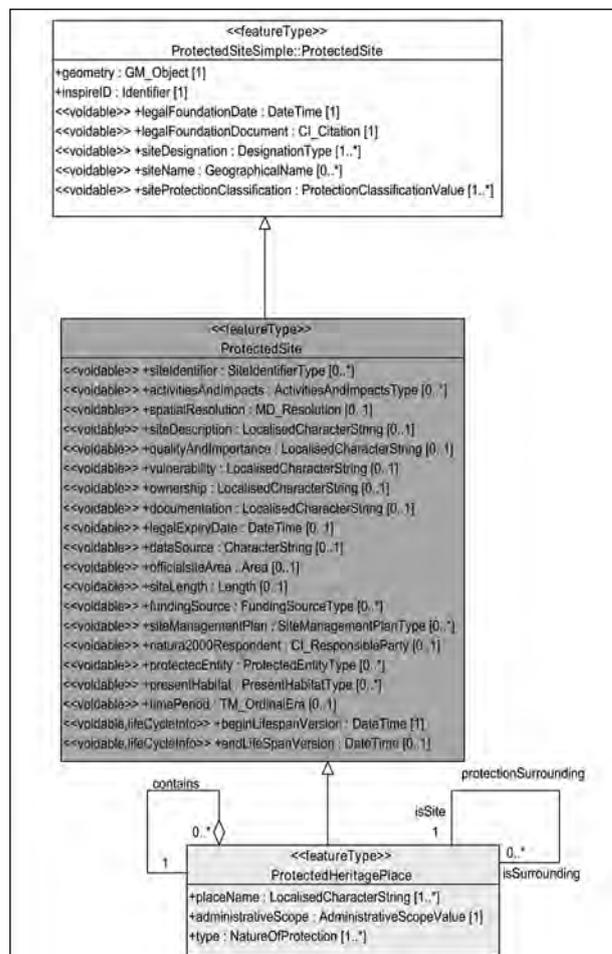


Figure 3. Cultural Heritage Application Schema: Legal part.

in the *DesignationSchemeValue* code list; the list of designations of each designation schema is stored in a code list. The Cultural Heritage Application Schema uses three designation schemas with their respective code lists, a preexisting one taken from the Protected Sites Simple Application Schema (*UNESCOWorldHeritage*) and two added by us in order to enrich cultural heritage description (*ProtectionTargetValue*, *BienesInteresCultural*). Besides, additional designation schemas may be incorporated to the data model by users in order to satisfy their specific needs. A same protected site can have several designations:

- *UNESCOWorldHeritage*: It includes the basic categories of World Heritage inscriptions classification established by UNESCO. Their values are in the code list

UNESCOWorldHeritageDesignationValue: natural, cultural and mixed.

- *ProtectionTargetValue*: Type of protected heritage place according to the UNESCO “Operational Guidelines for the Implementation of the World Heritage Convention” (UNESCO WHC 2011). Their values are in the code list *ProtectionTargetValue*: *Monument, Group of Buildings, Sites, Mixed* and *CulturalLandscape*.
- *BienInteresCultural*: This schema has been taken from the Spanish heritage management framework and added to this data model due to its central role in this context. *Bien de Interés Cultural (BIC)* is the highest protection status that Spanish heritage institutions can assign to a cultural entity. All nowadays existing categories of BIC are in the designation schema code list (see a full list of BIC types in Querol 2010, 71-75), being possible to add new ones as they are defined by Spanish heritage administrations.

siteProtectionClassification: The classification of the protected heritage place based on the purpose for protection. The enumeration *ProtectionClassificationValue* provided by the data specification includes the following values: *natureConservation, archaeological, cultural, ecological, landscape, environment, geological*. Only *cultural, archaeological* and partially *landscape* are related to cultural heritage. As long as archaeological sites and also landscapes are best regarded as a subtype of cultural sites, the constraint requires the filling of the attribute as *cultural* in all cases, leaving further considerations for a new attribute called *type* (see below).

administrativeScope: Administrative scope of the legal definition of the cultural heritage place. The enumeration *AdministrativeScopeValue* provides the values for this attribute: *International, European, National, Regional* and *Local*.

dataSource: The agency or organisation that is responsible for maintaining and providing the data about the cultural heritage place. In most Spanish cases it will be the corresponding Spanish regional administration, named *Comunidad Autónoma*, due to the decentralized character of heritage management in this country.

type: Reason advocated for the site’s protection, depending on the scientific or academic discipline that provides the arguments for the creation of the protected site: *Archaeological, Architectural* and *Ethnographical* are proposed as possible values. These values are in the code list *NatureOfProtection*, that may be extended with new categories if needed. It is important to note that this is not an attribute referring to the inherent nature of the culture features themselves, but to the heritage valuation one makes upon them. In fact many cultural entities can be “observed” and valorised from the perspective of several disciplines simultaneously. Here, the benefits of setting apart primary and derived entities can be well observed. For instance, a ruined church could be considered as an architectural and an archaeological element, and so could be part of both an architectural and an archaeological inventory. Sticking to a “traditional” disciplinary view on heritage classification would cause redundancies, since a single real-world entity (the ruined church) would be documented twice. Although the description and properties of the protected places created after it –including geometry– would be different depending on the disciplinary standpoint, and two different legal objects exist, the properties and description of the church –including geometry– are unique and do not need to be captured twice (and quite possibly in different ways).

Attributes about life cycle:

beginLifespanVersion: Date and time at which this version of the spatial object was inserted in the spatial dataset. There is also another attribute, *endLifespanVersion*, for indicating the date and time at which the version of the spatial object was declared obsolete. In this case, this attribute is avoidable, for there may be spatial objects that will not be removed at any point.

In order to consider potential links between different protected sites, two relations have been defined for the *ProtectedHeritagePlace* class to itself:

contents: This relationship is aimed at considering the inclusion of a protected heritage place by another (for instance, a monastery protected as a whole can include a church with a particular,

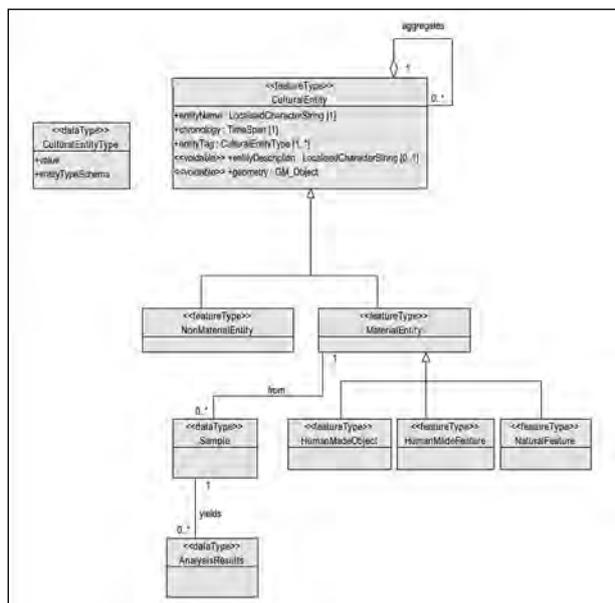


Figure 4. Cultural Heritage Application Schema: Cultural part.

specific protection).

protectionSurrounding: It is conceived to take into account protected sites complementary to others under the form of surrounding areas or buffers. As long as the protected surrounding will always have different attributes (e.g. geometry, identifier, etc.) it should be instantiated as a new object of the same class, and referred to through this self-association.

3.2 Cultural part

The central class of this part is *CulturalEntity* (Fig. 4). A cultural entity is any real-world feature result of human or human-natural action and, consequently, susceptible to be protected as cultural heritage.

CulturalEntity class is related to *ProtectedHeritagePlace* class to capture the relationship between legal objects (protected heritage places) and real-world entities to which they refer. *CulturalEntity* points to *ProtectedEntityType*, a preexisting class in the Data Specification on Protected Sites referring to any kind of a real-world feature susceptible of legal protection (e.g. a protected entity could also be a forest or an animal species); and *ProtectedEntityType* is related to

ProtectedSite by means of a voidable attribute that is *protectedEntity*. In this way, a protected site may contain one or more protected entities (in our case, cultural entities) and, reciprocally, a cultural entity may be “incorporated into” one or more protected sites.

It is worth noting that there is no need to relate every single protected heritage place to one or more cultural entities. In other words, the model allows entering information about a protected heritage place without being forced to separately describe and map the properties of the cultural entities protected by it. Although conceptually it is rather obvious that no protected place can exist if nothing is to be protected, this is a pragmatic decision to allow data providers (e.g. heritage managers) to publish even the very basic datasets they are able to produce. What is maybe more interesting, this also allows entering cultural entities without necessarily linking them to a certain protected heritage place. This means that the data model can also be used by data producers other than the administrative bodies (such as research institutions, private companies, free-lances, etc.). Furthermore, the model can be also conceived as a tool to foster cooperation and synergic processes of information maintenance and publication.

CulturalEntity has five attributes, three mandatory (*entityName*, *chronology*, *entityTag*) and two voidable (*entityDescription*, *geometry*):

entityName: Name of the cultural entity. It may differ from the name of the protected heritage place related to it.

chronology (Fig. 5): This complex issue has been tackled by means of ISO 19108:2002 (ISO/TC211 2002), which offers a temporal schema to describe the chronology of a cultural item, from its creation to its end, covering all the relevant events in its history. This temporal schema consists of a one-dimensional topology made up of two kind of primitives: *nodes* (instants) and *edges* (periods). This approach allows to data providers to describe the chronology of a certain cultural entity with the desired degree of accuracy, from just assigning a rough chronological indication to a rich and complex chronology that gathers all events with a precise date. The way of filling this attribute is through a

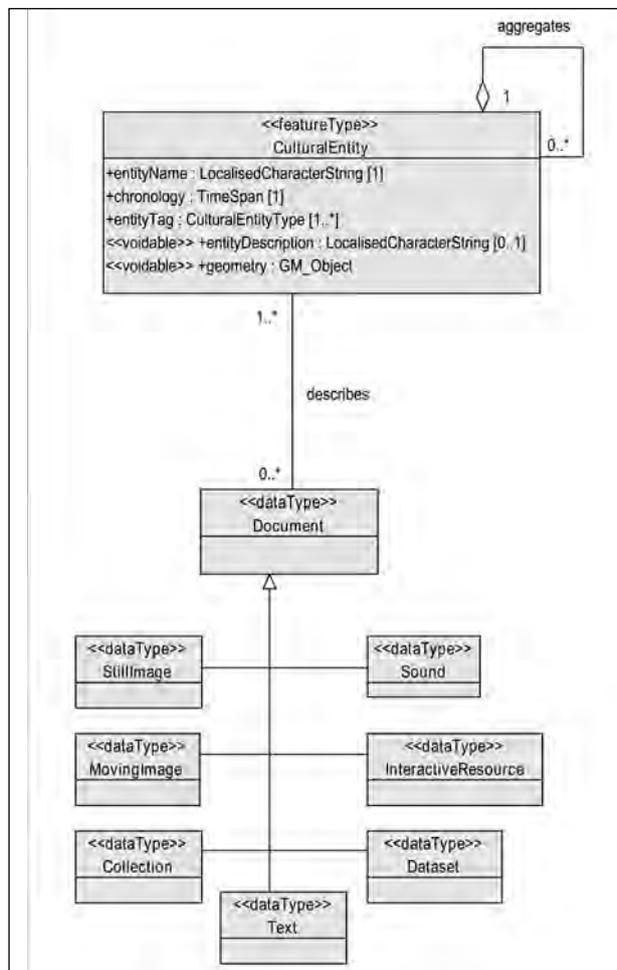


Figure 6. Cultural Heritage Application Schema: Documentary part.

new class (*TimeSpan*) that links the cultural entity with the chronological topology assigned to it.

In relation to this, we must point that the *ProtectedSite* class has a voidable attribute, *timePeriod*, intended to collect chronological information and related to the class *TM_OrdinalEra*, which belongs to the above mentioned topological temporal schema. Nevertheless, we have considered that it may be confusing for cultural heritage description for two reasons: firstly, it is ascribed to the legal entity, while we think that chronology is an attribute intrinsic to cultural entities themselves; and secondly, it only allows entering one chronological item for each entity, limiting variability in its chronological description.

entityTag: This attribute assigns cultural

values or categories (chronological, functional, etc.) to the cultural entity, as many as the data provider considers necessary. There are many ways to classify cultural entities according the subjects, chronological and geographical contexts, academic traditions and so on. Conceiving a universal classification of heritage places, no matter how generic, is obviously pure wishful thinking. Instead of that, data providers are encouraged to use a classification or schema (e.g. a thesaurus) of their choice (as most of them actually do), as long as they specify in this attribute the schema in use and the value adopted for the cultural entity within that schema. This information is collected by means of the *CulturalEntityType* class.

entityDescription: Textual description of the cultural entity.

geometry: Spatial definition of the cultural entity. The data provider does not necessarily have to assign a geometry to a specific cultural entity. Nevertheless, a cultural entity can be indirectly georeferenced if it is part of another cultural entity that really is (as we will see just now).

CulturalEntity class is related to itself in order to consider potential relationships of aggregation between specific cases. A cultural entity might be a whole building or an archaeological site. But it might also be a smaller feature such as a wall or a brick, being the building an aggregation of those smaller features and a cultural entity as well. Thus the disaggregation of the elements included in the model depends on the nature of the data gathered by the data provider. The self-aggregation relation named *aggregates* enables that behaviour.

In order to capture the diversity of cultural entities, several generic subclasses are proposed. A first distinction is made between material and immaterial entities (*MaterialEntity* and *NonMaterialEntity*), based on the very physical nature of things, depending on whether their existence is based on a tangible thing or on the performance of a certain event. Although widely accepted and used, the very notion of intangible or immaterial heritage has been actively challenged (Kirshenblatt-Gimblett 2004; Smith and Akagawa 2008). Among other reasons, it has been argued that any form of heritage involves both material and

immaterial dimensions (Munjeri 2004). However, one cannot deny that reality is composed of things of different nature –tangible and intangible– that are perceived and, consequently, described in different ways and using different terms. What are distinctly tangible or intangible are specific cultural entities, such as a building or a song; this is inherent in the nature of these things, and implies different mechanisms for describing and documenting them. The fact that tangible and intangible entities exist as separate concepts in the model does not mean that tangible and intangible cultural heritages could or should exist as separate things. Besides, the model allows linking a protected place to both tangible and intangible entities. This is, again, a benefit of setting apart primary (real-world) and derived (legal) entities.

Cultural material entities are classified according to the typology established by the CIDOC Conceptual Reference Model, that distinguishes three categories:

HumanMadeObject: Material things created by human activity and that are physically separate from other objects (for instance, a building, a pot or a chair).

HumanMadeFeature: Features that are physically attached to particular objects in an integral way (for instance, a petroglyph or a ditch).

NaturalFeature: Landscape features singularly identifiable, such as a tree or a rock. Although cultural heritage refers primarily to human-made things, it is rather frequent that natural features are related to cultural manifestations (such as a tree being related to traditional legends or a forest being the place for a festival).

The application schema also allows including analytical results taken from cultural material entities, and that provide with additional information about them (such as dates, physical composition, etc.). This is done by means of two classes:

Sample: Fraction taken from a cultural material entity for analytical purposes.

AnalyticalResults: Results of an analysis carried out on a specific sample.

3.3 Documentary part

Within this model, a document is not just a textual thing, but any resource that contains and provides information of a certain kind about anything. The application schema allows the inclusion of documents describing the cultural entities (e.g. implemented through a document repository), so the user can have access to them (Fig. 6). No cultural entity has to have any document necessarily linked to it; this is a piece of information that data producers and distributors may include or not, depending on the volume of information they have and/or they want to publish.

The Dublin Core Metadata Element Set and specifically its attribute *type* establish twelve specific kinds of documents⁵ according to the nature of the information they contain, seven of which have been selected for the aims of our data model:

Text: Primarily formed by words for reading (e.g. a journal article).

StillImage: Visual representation of static nature (e.g. a photograph).

MovingImage: Series of visual representations offering an impression of motion.

Sound: Acoustic representation.

Dataset: Set of data stored in a structured way (e.g. a table).

InteractiveResource: Resource requiring interaction from the user.

Collection: Aggregation of resources (e.g. a photographic archive).

4. Conclusions

To sum up, we want to remark what we consider some of the advantages of using a schema as the one proposed here for the organisation and distribution of cultural heritage information:

- Following the INSPIRE philosophy, it

⁵ DCMI Type Vocabulary: <http://dublincore.org/documents/dcmi-type-vocabulary/> (Accessed June 20, 2012).

would guarantee *interoperability* for the standardisation and distribution of georeferenced cultural heritage data.

- It considers and captures the *specificity* of cultural features as compared with natural areas.
- The *distinction* between cultural entities (as real-world things) and protected places (as administrative realities) captures a conceptually relevant difference, and has practical benefits for avoiding redundancies and simplifying the description and mapping of heritage-related entities.
- Pragmatically, it is aimed at *simplicity*, by limiting mandatory items just to one entity (*ProtectedHeritagePlace*) and thirteen attributes.
- Despite the former, it also incorporates *modularity*, introducing the possibility of including additional information through optional entities and attributes and also of extending the data model by other developers, by means of adding new specific subclasses.

Nowadays, the Cultural Heritage Application Schema is being used for developing some archaeological SDI. One of them is the IDEArq, a SDI planned within the Spanish National Research Council (CSIC) and aimed at the distribution of archaeological data archives of projects developed by this institution or in collaboration with it. Another example is IDEPatri⁶, a SDI for the publication of descriptive and analytical information available about Iron Age sites in the NW corner of the Iberian Peninsula after more than 100 years of archaeological research.

Acknowledgements

This paper is a further development of the conceptual data model on cultural heritage by the above mentioned Thematic Working Group on Cultural Heritage, within the Working Group of the Spanish Spatial Data Infrastructure. Many individuals and Spanish organisations have taken part in this group as experts: Juan M. Vicent García

⁶ IDE Arqueolóxica da Idade de Ferro en Galicia: <http://idepatri.cesga.es/> (Accessed June 20, 2012).

and Isabel del Bosque González, Centro de Ciencias Humanas y Sociales (CCHS), Consejo Superior de Investigaciones Científicas (CSIC) as coordinators; Alfonso Fraguas Bravo, Antonio Uriarte González, Juan Luis Pecharromán Fuente and María Ruiz del Árbol Moro, Instituto de Historia (IH-CCHS-CSIC); Emilio Abad Vidal, Centro de Supercomputación de Galicia (CESGA); Francisco García Cepeda, Dirección General del Catastro; Marta Criado Valdés, DMS-Group; Pilar Chías Navarro and Tomás Abad Balboa, Universidad de Alcalá de Henares; Miguel Lage Reis-Correia, Fundación Las Médulas – Junta de Castilla y León; Javier Márquez Piqueras, IDR–Universidad de Albacete; Victorino Mayoral Herrera, Instituto de Arqueología de Mérida (IAM-CSIC); Pastor Fábrega-Álvarez and César Parcero-Oubiña, Instituto de Ciencias del Patrimonio (Incipit – CSIC); Carlos Fernández Freire and Esther Pérez Asensio, Centro de Ciencias Humanas y Sociales (CCHS-CSIC); José Julio Zancajo Jiménez, Universidad de Salamanca; Antonio Vázquez Hoenhe, Arantza Respaldiza Hidalgo and Mercedes Farjas Abadía, Universidad Politécnica de Madrid.

This work has been partially supported by Spanish Government through the projects *Investigación en Tecnologías para la valoración y conservación del Patrimonio Cultural* (CONSOLIDER INGENIO 2010 CSD2007-00058) and *Definición de un modelo para la representación de aspectos espaciales del Patrimonio* (National R&D Plan).

References

- Crofts, N., M. Doerr, T. Gill, S. Stead, and M. Stiff. 2010. *Definition of the CIDOC Conceptual Reference Model*. Heraklion: ICOM/CIDOC CRM Special Interest Group.
- Directive 2007/2/EC. 2007. “Directive of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE).” *Official Journal of the European Union* 50(L108):1-14. Accessed June 20, 2012. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:108:0001:0014:EN:PDF>.
- Doerr, M. 2003. “The CIDOC Conceptual Reference Module. An Ontological Approach to Semantic Interoperability of Metadata.” *AI Magazine* 24(3): 75–92.

- Fernández-Freire, C., C. Parceró-Oubiña, and A. Uriarte-González, eds. *In press. A data model for Cultural Heritage within INSPIRE*. Santiago de Compostela: CSIC.
- González-Pérez, C., and C. Parceró-Oubiña. 2011. "A Conceptual Model for Cultural Heritage Definition and Motivation." In *Revive the Past. Proceedings of the 39th Conference on Computer Applications and Quantitative Methods in Archaeology*, edited by M. Zhou, I. Romanowska, Z. Wu, P. Xu, and P. Verhagen, 130-40. Amsterdam: Pallas Publications.
- GTT-PAH. 2011. "Borrador del modelo de datos de Patrimonio Histórico como Lugares Protegidos de INSPIRE. Consejo Superior Geográfico." Accessed June 20, 2012. <http://hdl.handle.net/10261/37300>.
- INSPIRE Thematic Working Group Protected Sites. 2010. "D2.8.I.9 INSPIRE Data Specification on Protected sites – Guidelines." Accessed June 20, 2012. http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_PS_v3.1.pdf.
- ISO/TC46/SC4. 2006. *ISO 21127:2006. Information and documentation – A reference ontology for the interchange of cultural heritage information*. International Organization for Standardization.
- ISO/TC46/SC4. 2009. *ISO 15836-2003. Information and documentation – The Dublin Core metadata element set*. International Organization for Standardization.
- ISO/TC211. 2002. *ISO 19108:2002. Geographic information – Temporal schema*. International Organization for Standardization.
- ISO/TC211. 2005. *Geographic information – Conceptual schema language*. International Organization for Standardization.
- Kirshenblatt-Gimblett, B. 2004. "Intangible Heritage as Metacultural Production." *Museum International* 56 (1-2): 52–65.
- McKeague, P., A.J. Corns, and R. Shaw. 2012. "Developing a Spatial Data Infrastructure for Archaeological and Built Heritage." *International Journal of Spatial Data Infrastructures Research* 7: 38–65.
- Munjeri, D. 2004. "Tangible and Intangible Heritage: from difference to convergence." *Museum International* 56(1-2): 12–20.
- Querol, M.A. 2010. *Manual de Gestión del Patrimonio Cultural*. Madrid: Akal.
- Smith, L., and N. Akagawa. 2008. *Intangible Heritage. Key Issues in Cultural Heritage*. New York: Routledge.
- UNESCO WHC. 2011. "Operational Guidelines for the Implementation of the World Heritage Convention. Paris: UNESCO." Accessed June 20, 2012. <http://whc.unesco.org/archive/opguide11-en.pdf>.

Old Places, New Ideas: New Routes into Canmore, the National Inventory of Scotland

Susan Hamilton, Rebecca Jones and Peter McKeague

Royal Commission on the Ancient and Historical Monuments of Scotland

Abstract:

The Royal Commission on the Ancient and Historical Monuments of Scotland (RCAHMS) maintains the national inventory and hosts the national collections of the archaeological and built heritage of Scotland, including its maritime waters. Both professional users and the wider public have been able to view information in the inventory online, through Canmore, since 1998. Yet there was no mechanism to actively contribute to the knowledge base. Since our centenary in 2008, new strategies have evolved to harness the knowledge of both our professional and public users, to contribute directly to the national inventory. This paper describes the opening-up - letting go - of Canmore, through a number of routes, to enable new partners to actively contribute their knowledge directly and efficiently to Canmore through a collaborative and flexible approach. The paper then discusses how information is disseminated using a similarly agile approach that aims to allow users to find and use the data that they need, while serving the needs of imperatives such as the European Union INSPIRE Directive. It is noted that pragmatism is required; government bodies, audiences and partners do not all move at the same pace as the available technologies.

Keywords:

National Inventory, Information Management, Big Data, Social Media, Web Services

1. Introduction

The Royal Commission on the Ancient and Historical Monuments of Scotland (RCAHMS) maintains the National Inventory and hosts the National Collections of the archaeological and built heritage of Scotland, including its maritime waters. In the last twenty five years, the Inventory's database has evolved from a paper-based card-index system, partly based on a reference system developed by the former Archaeology Branch of the Ordnance Survey supplemented by the paper catalogues to the RCAHMS Collections, to an Oracle database and ArcGis Geographic Information System with well-established online search facilities through Canmore (<http://canmore.rcahms.gov.uk>).

The transition from card-index to database has been progressive, reflecting changing priorities within the office as well as an enthusiastic engagement with the Internet, especially as a means of promoting RCAHMS' holdings. The RCAHMS Canmore online database was launched in April 1998 and has been evolving since. As part of the evolution, an Events – Monuments – Archive model for managing information created from specific

Corresponding author: susan.hamilton@rcahms.gov.uk

archaeological and architectural activities was introduced (McKeague and Jones 2008).

Following the development of this Events – Monuments – Archive model, there was a clear need to develop strategies to maintain and increase public engagement and to manage the increasing volume of information entering the organisation. A mechanism for submission of user-generated content to Canmore was created first. Two additional methods of data ingest were then developed, one offering 'direct' access into the database and another providing a streamlined, validated and more user-friendly experience (Fig. 1). Organisations wishing to submit information can use the method most appropriate for their needs, or a combination of all three.

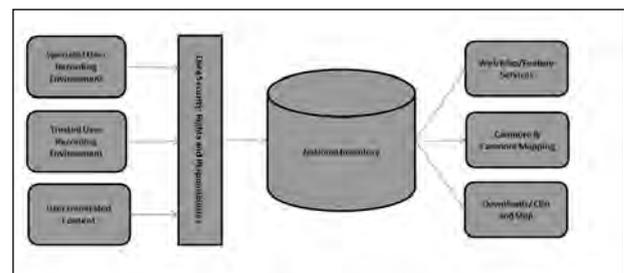


Figure 1. Routes into the National Inventory of Scotland.

The creation of this collaborative digital environment has proved a watershed in the management of heritage data in Scotland, demonstrating that the technologies available to us can be harnessed to create a 'toolbox' of techniques; no user or willing contributor should have their participation prevented by technology.

This paper describes the contents of that toolbox, which developed new routes into the database whilst preserving existing work procedures and retaining the structure and integrity of the data model. Balance is given to this discussion of ingest by describing the ways in which RCAHMS disseminates information. The - often contentious - subject of licensing is explored in the final part of the paper.

2. Information Ingest

From the time of the bulk data migration from card index to the Oracle database from the early 1990s until 2008, all information that became visible on Canmore had been added to the internal database by RCAHMS staff or volunteers. This included summaries of work undertaken by others, as featured in *Discovery and Excavation in Scotland* for example, as well as reports of RCAHMS fieldwork and information that supplements the holdings of the National Collection. It was clear that the volume of information being submitted to the organisation was growing and that new approaches were required which would ensure the wide range of work being carried out across Scotland could continue to be reported. Three primary methods of ingest are outlined here.

2.1 *Route One; Specialist User Recording Environment (SURE)*

SURE uses one database, the same as that used by staff at RCAHMS to hold the National Inventory, to input and share historic environment data, working to eliminate duplication of effort across the sector. That information is presented online to the user through Canmore. Using Canmore, now a well-established and heavily-used resource, public and professionals can find information in one place.

SURE has its origins in an existing partnership working between the National Trust for Scotland

(NTS) and RCAHMS. The NTS is an independent conservation charity which has a remit to care for cultural, built and natural heritage. It holds properties across Scotland that range from cottages to mountainous estates. Following a number of successful field survey partnerships, including work on the islands of Canna, Mingulay and St Kilda, RCAHMS agreed to provide the NTS archaeology team direct access to the internal database behind Canmore. This change went beyond technological because it required a culture change within RCAHMS. Staff had to relinquish full control of the national database and accept the proposal to allow other expert users to contribute to the record.

Through SURE, the NTS can curate the archaeological data generated through their research, field observations and conservation projects. The NTS archaeologists may add new records or update information for existing sites on their estates scattered throughout Scotland. Information is held alongside the existing records and is accessible immediately on Canmore for the benefit of all users. The ethos behind this – holding data once and using it many times – works towards a key Scottish Government requirement for spatial data (Scottish Government 2005).

SURE operates by granting permissions upon logging in to the secure database. At present, logins are restricted to specific IP addresses partly through a cautious approach to opening up the database to outside access for the first time and partly dictated by the use of Oracle Forms as an interface. A proposed migration to Oracle Application Express will enable a more flexible approach in the future. SURE partners and RCAHMS staff can access the input and query screens that they require, with appropriate read-only or read-write access. This ensures data security against unintentional updates and means that training can be efficiently targeted at a user's precise needs. The dispersed nature of the NTS estate posed an early challenge to the project; resolved through the creation of an internal, or system, event in the database granting editing rights for those records lying within the geographical extents of the NTS estate.

SURE partners can add new monument records or refine location information. Where a record already exists, data content is added, as

Events, through the Notes and Activities table, which handles the investigative history of a monument in addition to holding text descriptions and notes. The Notes and Activities table replaced a cumbersome, inflexible and monolithic descriptive text table where information was simply added cumulatively with specific records for each activity reported.

This approach allows the management of differences in opinion that are often associated with a monument; there is room within SURE for varying interpretations. Events or notes added by one partner's organisation cannot be deleted or edited by another (with the exception of members of the RCAHMS Data and Recording section who can access all users' information to amend typographical errors or similar). Data appears on Canmore in chronological order of Events, but can be sorted by the user into reverse order. Plans are in progress to enable more sophisticated searching in the future.

Following successful use by the NTS archaeologists, the SURE model was extended to provide the Orkney Islands archaeologist with a similar solution. Canmore now serves as the online Sites and Monuments Record for Orkney Islands Council. By sharing records, the burden of archaeological recording – inputting excavation reports and new publications – can be shared and exchanged instantly. On the technical side, the Orkney archaeologist (and other SURE partners) no longer has to deal with the 'nuts and bolts' of maintaining a complex database. The responsibility for upgrades, maintenance, resolving technical issues and meeting national and international data standards rests with RCAHMS; all of which is work that RCAHMS is permanently committed to. This not only frees up time for the archaeology team, but also for the council IT service.

This partnership has already borne fruit with the recent awarding of a grant in early 2012 to the Orkney Islands Archaeology Service. The award was from NESTA, an independent charity that supports innovation through investments and grants. The funding will be used to create web-based visitor materials using data from, and links to, Canmore that interpret monuments in the landscape and gather user-generated imagery and visitor responses to Orkney's Norse heritage (NESTA 2012).

The partnerships with Orkney and the NTS mean that SURE is now a tool in the conservation of three of Scotland's five UNESCO World Heritage Sites – St Kilda (NTS) and the Heart of Neolithic Orkney (Orkney Islands Council), as well as parts of Edinburgh's Old and New Towns (NTS).

A third partnership has been formed with the Garden History Society in Scotland (GHSS). This third-sector organisation produces high-quality surveys and research on gardens and designed landscapes across Scotland, and is largely volunteer-based. The Society's Scottish Conservation Officer will act as co-ordinator of the reports and be responsible for making them available on Canmore through SURE. The partnership provides an active community of interest with national exposure and recognition of the importance and quality of their work, and avoids the creation of separate databases or websites, whilst integrating valuable information on gardens within the National Inventory.

With each partnership, new elements of the system are being piloted. This ensures that SURE can develop, but that no new partners will have more than one 'untested' aspect to deal with. The GHSS will be the first SURE partners to utilise the Collections tables within the database, cataloguing PDF survey reports, which will become part of the RCAHMS digital archive downloadable through Canmore.

The most recent partner to join SURE is the Treasure Trove Unit at the National Museums Scotland, who will add detail to the existing picture by recording find spots directly into Canmore – instantly being able to map distributions or patterns of recovery and reporting through the use of Canmore mapping (<http://canmoremapping.rcahms.gov.uk>), while reinstating the link between object and place. This builds on partnership working previously carried out between RCAHMS and National Museums Scotland that explored linking catalogue records for museum objects with the relevant monument record in Canmore (Cowie and McKeague 2011). Furthermore, links between artefact and site were the subject of a workshop co-organised by RCAHMS in 2008 (Jones and Ralston 2011) and have been recognised as a systemic challenge for the profession by the Scottish Archaeological Research Framework (ScARF)

project run by the Society of Antiquaries of Scotland (Simon Gilmour, *pers comm.*). It was discussions between ScARF, the Treasure Trove Unit and RCAHMS that led to their inclusion into SURE, and continues the process of integrating information on object and place, as well as providing Treasure Trove with a system for managing information about their artefact record. During the course of 2012-13, it is anticipated SURE will continue to grow, with at least four additional Scottish local authorities set to become partners or start the process of joining.

2.2 Route Two: Trusted User Environment

A moderated approach to adding information to Canmore was pioneered by the Scotland's Rural Past (SRP) project (<http://scotlandsruralpast.org.uk/>) which enabled community groups, working in partnership with RCAHMS field staff, to report their project results via a mediated online form. SRP ran from October 2006 until September 2011 and saw over 60 community projects report their work through the form.

The Trusted User Environment is built in Oracle Application Express and provides an alternative, user-friendly, view of the database used by SURE partners. As with SURE, users can enhance information held for existing records and add new monument records. Once data entry is complete, the form is held within the system until it is checked and signed off by a registered and appropriately trained validator. On validation the information flows through into the database and onto Canmore. The project team produced extensive guidance to assist the community groups in not only recognising and recording the field monuments (RCAHMS 2011) but also on how to use and contribute to the online resources (Scotlands Rural Past 2012).

Recognising that users of this form may not be as familiar with heritage data as SURE partners, there are checks and simplifications built into this system. New national grid references must be entered twice - in the same manner as one would create a new online password - to reduce the risk of spatial errors. A macro then calculates the number of the 1:10 000 map sheet (a key reference identifier for RCAHMS) from the grid reference - a development which may be extended to all RCAHMS systems in due course. The form also offers fewer options to

report archaeological events, currently limiting the user to reporting documentary research, field visits or condition surveys. This ensured consistency of reporting throughout the SRP project; more terms can now be added but the aim is to always keep this form simpler than the direct database interface.

The NTS have now begun to use the Trusted User form. Seasonal ranger staff have been using it to report their visits to monuments on NTS properties. Once the pilot phase finishes in the summer of 2012, the role of validator can be undertaken by the relevant NTS area archaeologist. This means there will be no requirement for intervention by RCAHMS staff on any information that is generated by the NTS.

2.3 Route Three: User Generated Content

The most open and inclusive developments were the earliest to be implemented. As RCAHMS approached its centenary in 2008, there was a clear need to develop new strategies to maintain and increase public engagement. This started in the centenary year as part of RCAHMS 'Treasured Places' celebrations, with opening up the database to user-generated content, encouraging the active participation of Canmore users who previously had little or no access to 'official' methods of submission. The development of 'MyCanmore' saw the creation of a mechanism through which members of the public can add information to existing records on Canmore and share digital images via a Flickr photo-stream.

Users must register in order to post comments or share digital images. It is a self-moderated system, with a 'report it' button on every post; there has been little requirement thus far to remove any spam or inflammatory content. Data entered through these means is kept visually separate online, making it clear what data is user generated. However the information contributed is stored in the underlying Events and Notes table alongside other, more traditional recording events as 'User generated text/image Comment'. In essence, it is no different than any other field observation in that it represents a unique contribution in time. Despite being a relatively 'soft launch', with no large scale publicity, the opportunity to add content to Canmore has been a success. By May 2012, 20,000 images had been shared and over 1,200 pieces of text added.

Feedback from participants is guiding the next steps for Canmore and user generated content. RCAHMS is exploring ways in which images can be held on in-house servers. The Canmore User Group advised that contributors want to feel that they are actively contributing to an authoritative resource hosted by RCAHMS, not a photo sharing website.

There is also potential for the patterns of use to direct RCAHMS in its placement of resources. Prior to launch, there was a general assumption, shared by at least one author, that comments and photos would be higher in number on the 'big tickets' of Scottish heritage: Edinburgh Castle, Skara Brae and the Forth Bridge being three examples. As part of the Centenary Treasured Places project, investment had been made in writing jargon-free, summary records of one hundred such places. The reality was quite different. One of the first examples of users interacting with each other through Canmore was a discussion about a shop and flats within a suburb of a Highland town. There are other examples similar to this and it is a valuable lesson in demonstrating that what organisations such as RCAHMS may think is interesting to the public is not necessarily the case and therefore resources should be targeted at projects that simply make as much information available as possible.

2.4 Other routes

The concept of user generated content is not limited to 'the public'. With the Archaeology Data Service, RCAHMS has taken the lead role in introducing their OASIS application (Hardman and Richards 2003) to Scotland to standardise and ease the information flow from development-funded fieldwork to curators at both local and national level. This project has been running successfully in Scotland since 2006 although to date the information flow has not been integrated into Canmore.

3. Information Dissemination

The same flexibility and range of approaches that have been developed to allow information ingest are also being applied to dissemination. In addition to legislative imperatives such as the European Union INSPIRE Directive, through initiatives such as www.data.gov.uk there is a growing expectation that users should be able to

not only view, and contribute to, data, but extract and use it for many reasons, preferably under open access terms. It is also in the interests of RCAHMS, its partner organisations in SURE and contributors, to promote its holdings for the public benefit and contribute to the National Performance Framework (Scottish Government 2011).

The Web Services that provide data from the underlying internal staff database to Canmore are also used to contribute information to ScotlandsPlaces (<http://scotlandsplaces.gov.uk/>), a partnership portal with the National Records of Scotland and the National Library of Scotland enabling users to search aspects of the contributing organisations collections through geographic locations (Beamer and Gillick 2010).

Spatial information is increasingly accessible through Web Map Services (WMS) and Web Feature Services (WFS). RCAHMS has developed a WMS for its site location data, and is in the final stages of releasing a WMS for its aerial survey mapping programme (McKeague and Middleton 2013). The WMS, as well as providing content for RCAHMS web-mapping applications, can be accessed, subject to licence, for use by third parties within their own systems and viewed against their own datasets. Metadata for publically accessible WMS have been published on the Scottish Spatial Data Infrastructure website (Edina 2011) as part of a commitment to provide data for Scotland's Spatial Data Infrastructure and INSPIRE beyond the formally mandated datasets. Publically released Web Feature Services will follow once formal standards are released by the Scottish Government. In the meantime the internal WFS has been made available to the web developers working on the Orkney Islands Council Nesta project (see above) to develop innovative products under their own identity.

Users of ScotlandsPlaces are able to download results from Canmore in KML, whilst an API has been developed (but not yet released) as part of proof-of-concept research into participation and learning relating to the collections held by RCAHMS (RCAHMS 2012a)

It is not realistic to expect the wider community to have (access to) the expertise or the

capacity to adopt new technologies at the same time. It is, therefore, vital that the supply of data must be available to those who need it. Provision of ad hoc downloads on demand is continuing, primarily as database extracts and GIS Shapefiles, supplied through an online licensing system, for discussion of which see below. It is anticipated that provision of downloads – a resource-hungry process – will decrease with the increase in availability of WMS and WFS and development of ‘Clip and Ship’ solutions on Canmore. Clip and Ship will enable users to upload or create a shape around the area in which they are interested (‘clip’), and use that to download information (‘ship’), automatically through the online licensing system, thereby using the technology to make efficiency savings of staff time. It is accepted that more complex requests for data will still require staff time, for which a charging model is in place (RCAHMS 2012b).

Research has begun within RCAHMS into the best way to make information available for use as Linked Data. There is a will within the organisation to make data available in this way, but also concerns too about letting go of primary information; this must be balanced with the resources that a small public-sector body can expend on investing in technologies that have not yet been widely adopted. It is expected that forthcoming redevelopments of the RCAHMS database will at least include provision for stable URIs.

Through web services, the information RCAHMS hosts may be accessed by many portals. Whilst Canmore is the primary website for researching the National Inventory, ScotlandsPlaces also consumes web services to publish RCAHMS data alongside selected information from the National Records of Scotland and the National Library of Scotland: Map library. RCAHMS information is also published in PASTMAP alongside some of Scotland’s local Historic Environment Records and statutory information from Historic Scotland. With the release of WMS RCAHMS data can also be viewed remotely by other users alongside their own datasets in a GIS.

3.1 Social Media and other alternatives

RCAHMS has recognised that, as with ingest, dissemination can encompass a range of approaches.

This is most evident in the use of social media. The adoption of social media by RCAHMS as a method of raising awareness, communication and information provision has been a successful endeavour that complements existing methods – again the analogy of one tool within a toolbox is appropriate.

Through the use of Facebook and Twitter, in particular, RCAHMS has been able to present an informal, approachable public face and attempt to attract a new audience. There is evidence that this is working – beyond the usual counts of ‘Likes’ or ‘Followers’ – as followers are enthusiastically sharing information posted by RCAHMS. Through social media RCAHMS is able to take an active part in online events such as the worldwide ‘Day of Archaeology’ (<http://www.dayofarchaeology.com/>).

There can be tough lessons to be learnt in the use of social media. In 2011, RCAHMS attracted some disparaging comments on both Facebook and Twitter after requesting the removal of Crown Copyright imagery from the former; other members of the public then responded in the organisation’s defence. (Clari and Graham 2012) Engaging in social media means accepting its open nature.

4. Licensing

Opening up the RCAHMS database to external contributions has been a process of letting go and relaxing controls. The same challenges now face data dissemination where a balance must be struck between the need to generate revenue from image sales (gaining additional funds from external, non-government, sources is a key strategic priority for RCAHMS) and the protection of Crown Copyright and a genuine desire to make data available to as many users as possible. The complexities of this issue have been highlighted in a recent case involving re-use of high-definition images from the collection of the National Portrait Gallery in London (Wikimedia 2010).

In addition to protection of the information and collections for which it is responsible, RCAHMS must also operate within the license conditions set by other bodies, for example data providers including the Ordnance Survey and software vendors such as Oracle (in addition to the terms and conditions of websites such as Flickr, Twitter and Facebook).

For example, at the time of writing, redevelopment of RCAHMS mapping applications are complicated by Ordnance Survey concerns over potential unauthorised access and reuse of both the WMS and WFS. This is by no means a unique situation, but it is worth highlighting in the face of user frustration at the speed at which data can be made openly available. As an example, despite considerable goodwill on both sides, the negotiations surrounding setting up one of the SURE partnerships took almost one year as the partners discussed issues of copyright and data ownership prior to the signing of the Memorandum of Agreement that formalised the partnership.

Partnerships such as ScotlandsPlaces have begun to explore methods of making data available through the use of a click license prior to download of a selected kml file. A click through license may also be used in the planned development of a 'download zone' on the RCAHMS website. It is hoped to make data such as aerial photographic mapping layers available in this way. Downloads of data are made available to users through an online licensing system. Terms and conditions of use must be accepted before the requested file is made available (ScotlandsPlaces 2012). This system is also primarily used for the purchase of digital imagery at RCAHMS to protect ownership rights and prevent the creation of orphaned imagery.

5. Conclusions

The solutions described above rely on a collaborative and flexible approach to recording the archaeological and built heritage of Scotland and providing a seamless public experience and appreciation of that heritage. Anyone, from professional users and community groups to interested members of the public can contribute their wide ranging observations directly and efficiently to Canmore. Our recent experience has demonstrated that public participation should not be limited, or directed towards those records we perceive as the most significant. SURE, together with the complimentary alternative methods of data ingest, ensures a permanent home and ongoing legacy for the work of all partners, contributors and colleagues while reducing costs and improving the user experience.

The benefits of such a participative approach to knowledge creation are obvious: double-handling of records and duplication of effort is eliminated; recording mechanisms are simplified and the financial burden of maintaining separate systems and IT infrastructure is reduced. This last point is of particular concern in the current economic climate, where resources across the sector are becoming increasingly stretched. However, creating such a system requires complex negotiations, and an understanding that finding a way through existing licenses, terms of use and copyright legislation will take longer than desired. This can be frustrating given the speed at which technology evolves.

The developments that have been undertaken by RCAHMS in recent years have demonstrated that it is not necessary for public organisations to limit themselves to one or two methods of data ingest and dissemination. Offering a choice and flexibility can create a system which is agile and responsive to user needs. Continuous piloting of new developments, if well managed, can benefit host and partners and lead to a constantly improving suite of tools.

At the heart of all the developments described here, there lies a recognition that both ingest and dissemination need to be flexible, as it is only through building an agile system that the demands of the 21st century can be met. 'One size' does not - and should not - fit all users.

Acknowledgements

The work reported here represents projects carried out across RCAHMS and the authors would like to acknowledge Rebecca Bailey, Ashley Beamer, Piers Dixon, Mark Gillick, Philip Graham, Leanne McCafferty, Jo McCoy, Mike Middleton, Graham Ritchie, Hannah Smith, Robin Turner and Alistair Wilkie.

The collaboration of external partners has been essential. Partners from other bodies include Derek Alexander and Daniel Rhodes (National Trust for Scotland), Alison Allighan (The Garden History Society in Scotland), Stuart Campbell (Treasure Trove Unit), Trevor Cowie (National Museums Scotland) and Julie Gibson (Orkney Islands Council).

References

- Beamer, A., and M. Gillick. 2010. "ScotlandsPlaces: Accessing Remote Digital Heritage Datasets using Web Services." *Lecture Notes in Computer Science* 6436: 225-239.
- Clari, M., and P. Graham. 2012. "Learning to let go: Changing Patterns of Participation and Learning through the Digital Collections of the Royal Commission on the Ancient and Historical Monuments of Scotland." Paper presented at the *Museums and the Web, 11-14 April 2012 San Diego*. Accessed 4 February 2012. http://www.museumsandtheweb.com/mw2012/papers/learning_to_let_go_changing_patterns_of_partic.
- Cowie, T., and P. McKeague. 2011. "Mapping material culture: exploring the interface between museum artefacts and their geographical context." *Scottish Archaeological Journal* 32 (1): 73-93.
- Edina. 2011. "Scottish SDI Discovery Metadata Catalogue." Accessed 12 June 2012. <http://scotgovsdi.edina.ac.uk/srv/en/main.home>.
- Hardman, C.S., and J.D. Richards. 2003. "OASIS: Dealing with the Digital Revolution." In *Computer Applications and Quantitative Methods in Archaeology 2002*, edited by M. Doerr and A. Sarris, 325-28. Athens: Archive of Monuments and Publications, Hellenic Ministry of Culture.
- Jones, R.H., and I.B.M. Ralston (eds). 2011. *Artefacts, Records, Monuments and Sites*. Special edition of *Scottish Archaeological Journal* 32 (1).
- McKeague, P., and R.H. Jones. 2008. "The national database for Scotland: evolution of the digital resource." in *Layers of Perception. Proceedings of the 35th International Conference on Computer Applications and Quantitative Methods in Archaeology (CAA) Berlin, April 2-6 2007*, edited by A. Posluschny, K. Lambers and I. Herzog, 186-193. Bonn: Koll. Vor- und Frühgeschichte.
- McKeague, P., and M. Middleton (this volume) "Beyond Inspire: towards delivering richer heritage data in Scotland."
- NESTA. 2012. "Scottish local authorities open up data to tackle community issues." Accessed 8 June 2012 http://www.nesta.org.uk/press_releases/assets/features/scottish_local_authorities_open_up_data_to_tackle_community_issues.
- RCAHMS. 2012a. "Beyond Text Project." Accessed 11 June 2012 <http://www.rcahms.gov.uk/news/beyond-text-project>.
- RCAHMS 2012b. "Buy images and data." Accessed 12 June 2012. <http://www.rcahms.gov.uk/buy-images-and-data.html>.
- ScotlandsPlaces. 2012. "Legals." Accessed 15 October 2012. <http://www.scotlandsplaces.gov.uk/footer/index.php?action=view&id=47>.
- ScotlandsRuralPast. 2012. "Using the field recording form." Accessed 12 June 2012 http://www.scotlandsruralpast.org.uk/pdfs/recording_form_guidance_notes.pdf.
- Scottish Government. 2005. "One Scotland - One Geography; a geographic information strategy for Scotland." Accessed 12 June 2012. <http://www.scotland.gov.uk/Publications/2005/08/31114408/44098>.
- Scottish Government. 2011. "National Performance Framework: December 2011." Accessed 12 June 2012 http://www.scotlandsruralpast.org.uk/pdfs/recording_form_guidance_notes.pdf.
- Wikimedia. 2010. "Dcoetzee/NPG_legal_threat." Accessed 10 June 2012. http://commons.wikimedia.org/wiki/User:Dcoetzee/NPG_legal_threat.

ADS easy: an Automated e-archiving System for Archaeology

Ray Moore, Catherine Hardman, Julian Richards and Lei Xia
Archaeology Data Service, University of York, United Kingdom

Abstract:

This paper describes the development of ADS easy, an online system for facilitating the deposit of digital research data with the Archaeology Data Service, a trusted digital repository for the UK historic environment sector. ADS easy provides an online costing tool and a mechanism for the upload of data and metadata files. The semi-automation of the deposit and ingest process will reduce costs but will also give users great control of their archiving. It is argued that this may, in turn, lead to greater engagement with selection and retention decisions.

Keywords:

ADS easy, Digital Repository, Digital Archive, Automated Ingest, Selection and Retention

1. Introduction

The growing dependence on digital data in archaeology has raised awareness of the need for long-term preservation of the datasets resulting from archaeological research. The Archaeology Data Service (ADS) is the mandated repository for many organisations within both commercial and academic sectors within the UK, providing archiving services for the digital outputs of archaeological fieldwork and research² (<http://archaeologydataservice.ac.uk/>). The ADS has been able to harness over fifteen years of experience in preserving and disseminating archaeological digital data, to develop an online system that will both streamline and enhance elements of the archiving process. It is hoped that *ADS easy*, an online tool to allow depositors to estimate archiving charges and to upload files and metadata to the ADS repository, will reduce archiving costs and make digital preservation an economic reality for those working within the archaeological community. At a time when the historic environment sector is suffering from reduced funding the need to develop a sustainable funding model is an imperative (ADS 2012).

ADS easy is being developed under the SWORD-ARM project, and has been funded by

Corresponding author: ray.moore@york.ac.uk

² For more information on the ADS see the ADS website. Accessed June 20, 2012. <http://archaeologydataservice.ac.uk/about>.

the Joint Information Systems Committee (JISC).³ Within SWORD-ARM the ADS is refining and enhancing the archive ingest and charging processes, using a SWORD client that will both streamline and semi-automate the deposition process. SWORD-ARM will deliver real benefits to depositors in terms of their ability to deposit data, create and validate metadata, and engage in the process of selection and retention. It will also allow them to manage single and multiple deposits and, perhaps most significantly, to manage cost estimates and accounts. SWORD-ARM presents the opportunity to enhance ADS data management systems and to build upon existing business infrastructure and role as a discipline-based repository.

SWORD-ARM builds upon previous ADS experience with automated or semi-automated ingest. The electronic submission of grey literature using OASIS (*Online AccesS to the Index of archaeological investigations*), has illustrated the real cost benefits of a semi-automated system for the submission of digital data, specifically unpublished fieldwork reports, and has been firmly entrenched within the ADS workflow for 10 years (Hardman and Richards 2003, Hardman 2006).⁴ More recently

³ For an overview of SWORD-ARM and for developments within the project please refer to SWORD-ARM blog. Accessed June 20, 2012. <http://archaeologydataservice.ac.uk/blog/sword-arm/>. Project partners include the University of Glasgow, the University of Manchester and the University of Southampton.

⁴ OASIS provides an index of fieldwork carried out by individuals, units and organisations across the UK within the academic, but particularly within the private, sector. OASIS records event level metadata. The data created through the OASIS

the re-use of metadata, submitted using the OASIS system, has become embedded within existing workflows. Southampton Arts and Heritage, aware of the growing quantities of digital data produced during archaeological fieldwork within the authority and the need for a long term preservation strategies for that data, are working with the ADS to provide a secure archive for the digital outcomes of archaeological fieldwork in the Southampton area. Previously this digital data had been deposited with the Southampton Arts and Heritage Service along with the 'physical' archive. However the Arts and Heritage Service had limited experience of dealing with digital data. The resultant *Southampton's Designated Archaeology Collections Programme* has incorporated this digital data within the ADS where it can be archived and migrated to suitable preservation formats, while at the same time making this data available to a wide audience through the ADS catalogue.⁵ During the archiving process project metadata submitted by fieldworkers in Southampton using the OASIS system has been utilised to create the necessary metadata that forms the backbone of the collections on line. This process extends the re-use of metadata captured by the OASIS system, signposting the broader spectrum of digital outcomes from commercial archaeology, including databases, images, CAD and GIS files.

The deposit of this digital material with a dedicated, specialist data centre such as the ADS has obvious benefits for the survival in perpetuity of these digital outcomes, making them available for further research, whilst the re-use of the associated metadata has real cost-saving benefits which can be shared between the archive and the depositor. The *Wessex Archaeology Image Archive*,⁶ has amalgamated this re-use of metadata with a semi-automated ingest procedure involving the electronic submission of the digital photographic

on-line form is accessible for both regional Historic Environment Record's (HER) and National Monument Record (NMR), but is also used to create the metadata which forms the index to the grey literature library. Through OASIS users can upload flat 'grey literature' reports with the OASIS record forming the metadata that accompanies that report.

⁵ Southampton's Designated Archaeology Collections Programme, <http://archaeologydataservice.ac.uk/archives/view/southampton/>, accessed June 18, 2012.

⁶ Wessex Archaeology Image Archive, http://archaeologydataservice.ac.uk/archives/view/wessex_images/, accessed June 18, 2012.

records and metadata from commercial fieldwork. This system utilises a Python-based client to upload files to the ADS, with the re-use of existing metadata submitted using OASIS, much like the Southampton Designated Archaeology Collections Programme. Both projects have highlighted a need for automated systems for file and metadata deposition, particularly if a scalable and sustainable digital archive is to be achieved. It is the experiences garnered under the auspices of these projects that have positioned the ADS to develop a bespoke and specific system for semi-automated file upload and metadata deposition.

2. Current Workflows

Within the current ADS workflow data and metadata are generally deposited using physical media, with CDs, DVDs, and occasionally hard-discs, sent through the post. Smaller archives may be delivered by email within the obvious limitations on file sizes that can be experienced using this method. Increasingly, however, depositors will ask to submit data using an external file hosting service, such *Dropbox* or *MediaFire*, illustrating a desire amongst depositors for a more convenient mechanism for file delivery. Despite these developments the most common method of deposition still remains 'the CD in the post', but data sent using these physical media require significant intervention from the archivist *prior* to ingest in the archive. Data must be virus checked, opened to verify that files have been correctly copied to the media, checksums generated to authenticate the transfer of data from media to archive server, etc. The process can be prolonged further if the collection is lacking the necessary metadata or documentation, requiring the archivist to spend time negotiating with the depositor. Documentation may arrive in a 'physical' form requiring it to be scanned in order that it can be stored alongside data, while project metadata often arrives in a 'flat' pro-forma style requiring the archivist to manually input it into the Collection Management System (CMS). Current workflows are already in place to allow for the collation of some file level metadata. The ADS implementation of Fedora has allowed files to be held within a Unix directory structure that reflects the project and the resultant original and preservation versions of data (Austin 2010), allowing metadata to be held alongside technical metadata created using automated tools

like DROID⁷ (Digital Record Object Identification) (Mitcham 2011). Only when this process is complete can proper archiving begin.

3. SWORD and SWORD-ARM

As the name implies the SWORD-ARM project uses the SWORD protocol to transfer data from a remote client (depositor) to the host (archive). SWORD⁸ is essentially a light-weight profile of the Atom Publishing Protocol (known as APP or ATOMPUB) that was developed by JISC with the intention of developing a standard mechanism for depositing digital data into repositories and other systems (Allinson et al. 2008, Lewis et al. 2009). The development of the protocol came as a response to a trend within the digital community for standardising the output of digital repositories through OAI-PMH, Dublin Core (DC) metadata etc. although there were few standards for getting data *into* these repositories. The implementation of the protocol amongst a broad range of repositories DSpace, EPrints, Fedora, Intralibrary and others (see Allinson et al. 2008, Currier 2009) has aided interoperability (Lewis et al. 2009). The recent development of SWORD v2 has extended the flexibility of the protocol beyond simple upload, to include support for the updating, replacement and deletion of resources, thereby providing support for deposit through the lifetime of a research project (Lewis et al. 2012). The significant advantage of SWORD v2 is that it allowed the development of broader, interoperable deposit management systems, extending the original 'file and forget' approach adopted for the original protocol.

The ADS implementation of the SWORD protocol within *ADS easy* proved problematic in that the SWORD protocol's initial development stemmed from a requirement for the deposit of simple scholarly communications outputs into academic digital repositories; such outputs were based upon small text-based items. Unfortunately, the problem with archaeological data is that is neither simple, nor entirely text-based. As a result

⁷ A tool developed by The National Archives of the United Kingdom to profile files using the PRONOM technical registry (see <http://www.nationalarchives.gov.uk/information-management/our-services/dc-file-profiling-tool.htm>, accessed June 18, 2012).

⁸ Sword standing for *Simple Web service Offering Repository Deposit Protocol* (Allinson et al. 2008).

while the SWORD protocol will be used for simple deposits, particularly those involving reports, for more complex data *ADS easy* will utilise a more robust protocol.

4. ADS easy

ADS easy builds upon existing ADS infrastructure, but affords a more rigorous control of the ingest process allowing the automation of certain processes, whilst simultaneously allowing greater flexibility for the depositor who can control what is and what is not included in the archive right up until final submission. This allows for better data management by the depositor, allowing the individual(s) most familiar with the dataset to select the significant elements of their archive according to their own research criteria. *ADS easy* allows for the electronic submission of both project and file level metadata, which is stored digitally and routed to the ADS Collections Management System (CMS). Actual data can also be uploaded directly to the ADS servers, rather than relying on an external file hosting service, or the postal service. In a significant development, reflecting the desire to become more transparent, a costing module has been added to *ADS easy*, allowing depositors greater control in making financial decisions about their archive; it allows them to receive estimates *before* a project commences allowing them to formulate a costed data archiving plan at the outset.

4.1 Project and file level metadata and e-licensing

As Kintigh and Altschul have identified:

The largest component of this [curation] process is the provision of structured metadata, which provide the detailed technical and semantic documentation that are required for the data to be meaningfully used and technically sustained into the future (2010, 268).

The importance of 'good' metadata cannot be underestimated as traditionally it has played a significant role in the selection and retention policies employed by digital archives. As Gutmann et al. observe

The documentation and supporting information that are included in the data collection

are also important in the decision-making process. Data should have comprehensive technical documentation that provides ample information on sampling procedures, weighting, recoding rules, and data collection procedures. Data collections that include such information are strongly preferred in the archiving decision, as they allow users to assess the quality and analytical reliability of the data (2004, 213).

The difficulty comes in making the judgement about metadata obligations in order to fulfil OAI-PMH/DC metadata requirements, and what amount of metadata people are, in reality, willing to manually input. This issue is one familiar to many digital archives, but the ADS experience has shown that engagement with depositors has led to a more considered awareness of the significance of such metadata.

Within the current ADS ingest and accessioning workflow a significant amount of duplication of effort is involved in manually entering metadata into the CMS. *ADS easy* will address this issue by allowing project metadata to be submitted electronically through an online form. The benefits of this process are obvious, removing repetition, but also reducing the potential for the inaccurate transfer of metadata; a significant issue when this information will form the basis of archive discovery amongst ADS collections.⁹ While *ADS easy* will streamline internal procedures it offers similar benefits for depositors too, enabling the re-use of existing project metadata, primarily from within the OASIS system. Many of those who will be using *ADS easy* are already frequent users of OASIS. This shortcut allows for more effective use of time by both depositor and archivist. The resultant metadata will mirror the standards of the current ADS workflow: the OAI-PMH requirements for DC metadata. With regard to file level metadata *ADS easy* utilises a spreadsheet *pro forma* system for collection and deposit. This spreadsheet can be downloaded and worked on locally, before being passed electronically to the ADS servers and reused in its workflows.

ADSeasy “has the potential of directly engaging those who created the data and know it best” (Kintigh and Altschul 2010, 268), encouraging the data

⁹ For example Archsearch (<http://archaeologydataservice.ac.uk/archsearch/>, accessed June 19, 2012).

creator to generate information about the data they have created. Of course the issue of quality remains, particularly when metadata is input by those who are arguably not so *au fait* with technical standards. Previous experience using controlled vocabularies¹⁰ in OASIS has shown that once embedded in workflows it becomes more straightforward, while the spectre of the ‘mandatory field’ will provide the gentle hand to guide the unwilling. In order to help those perhaps less familiar with metadata generation significant investment will be made in FAQ and online assistance that can effectively guide depositors in the creation of appropriate metadata.

While digital archivists have long understood that having good metadata is a prerequisite to producing good datasets, it is also important to address the use of data selection and retention policies. Such awareness has often proven difficult to pass onto data creators (Gutmann et al. 2004, see above). The active engagement of depositors in the selection and retention process, which *ADS easy* affords, certainly contributes to a wider and more thorough awareness and understanding of the need for documentation; a significant by-product of this engagement. *ADS easy* also offers an opportunity to increase awareness of archiving strategies amongst the wider archaeological community, highlighting the need for relevant, well documented archives that are suitable for preservation, and subsequent re-use.

Once the selected files have been identified and their associated metadata has been generated either through the online form, or using the downloadable spreadsheet, the archive can be uploaded electronically and passed to a ‘clearing house’ where it can be held until it has been checked and validated by an archivist. Once this process is complete it can be passed into the CMS, and both data and metadata held securely until archiving work begins.

4.2 Costing tool

One of the more significant developments of

¹⁰ As outlined elsewhere an awareness for the need for ontologies and controlled vocabularies, such as the *National Monuments Record Thesauri* (<http://thesaurus.english-heritage.org.uk/newuser.htm> and <http://thesaurus.english-heritage.org.uk/frequentuser.htm>, accessed June 19, 2012) is an important issue within archaeology (Richards and Hardman 2008).

ADS easy is the development of an online costing tool. Indeed one of the projects principal aims has been to make charging more transparent and to simplify the entire process (ADS 2012). The ADS already has a charging policy based upon a one-off fee levied at the point of deposit, which has proven to be robust and sustainable (Richards et al. 2010). Yet at Richards et al. observe such a strategy

is not easy in a field where future costs are unknown and no one knows how long the future will be. The ADS preservation policy is based on migration of files to the latest formats, and it is impossible to predict how many times this might have to be undertaken (2010, 258).

Under current ADS guidelines cost is judged on a project-by-project basis and is highly dependent upon archive content and the projected level of involvement of the digital archivist in preparing suitable preservation formats, not only today but in perpetuity. It is perhaps for this reason that depositors can regard the costing of archiving as something of 'dark art', making comparison between projects somewhat difficult (ADS 2012). Yet for the ADS the issue of cost of the digital archiving of archaeological fieldwork has been an important issue for a number of years. The *Digital Archiving Pilot Project for Excavation Records* (DAPPER) project funded by English Heritage in 1999, for example, examined the underlying costs of digital archiving, whilst investigating the benefits of the dissemination and re-use of this same data (Austin et al. 2001). The ADS has always taken a self-reflexive approach to charging and in an attempt to be transparent has published its own charging policy.¹¹ The ADS charging policy continues to be used as an exemplar in wider studies of digital preservation (Beagrie et al. 2008; Beagrie et al. 2010), with such reports feeding into further development of the existing policy. A significant observation from these reports has been an acknowledgement of "the high proportion of staff costs associated with data curation" (Beagrie et al. 2010, 34), something which automation and the generation of metadata in a digital form directly from data providers seeks to address. This is something that should certainly result in real cost benefits for

depositors. These activities should certainly reduce direct costs for archive setup, whilst the automation of certain processes should allow for lower costs on a file by file basis. With this in mind the costing tool implements a simpler charging policy based around the number and type of files uploaded, once an initial 'set up' fee has been levied. As is summarised elsewhere

we have tried to keep things simple with a start up fee covering elements of management and administration for each project and then a range of prices usually per file; the more complex the file type the more expensive to deposit. So we'd charge £1 for a .csv file and up to £6 for an .au audio file that takes a lot more time to check, document and preserve (ADS 2012).

Within this costing system those 'simple' files (e.g. images or csv files) that require minimal archival intervention have a much lower price per file than the more complex data files which demand a much greater amount of input from the archivist and are consequently more expensive (e.g. audio or video). At the same time measures have been taken to preserve the existing ADS cost profile, whereby economies of scale are utilised for certain file types (ADS 2012, Beagrie et al. 2010). Ultimately, it is hoped that *ADS easy* will allow the ADS to keep costs low and allow a closer degree of comparability with the size and content of the archive (ADS 2012).

On a practical level the costing module allows depositors to set up and manage 'accounts' on an institutional or company basis; but it also affords them an opportunity to create and manage costs for individual archives on a file-by-file basis. Significantly the costing tool also allows depositors to calculate the actual cost of archiving at the outset of a project, to generate estimates and consequently accommodate these costs within their budgets. This has particular benefits within commercial archaeology where archive costs can be passed on to developers. Prior to logging into *ADS easy* there is also an area where individuals and organisations can experiment to estimate the costs associated with the upload of various permutations of files. For formal quotations prospective depositors must first register, login and begin to create an archive project. At a higher level *ADS easy* affords depositors the opportunity to manage accounts at

¹¹ *Charging Policy. Version 4, November 2007.* (<http://archaeologydataservice.ac.uk/advice/chargingPolicy>, accessed June 20, 2012). For a more focused discussion of charging see Richards et al. 2010.

an organisational level, allowing them to consider project histories and compare transactions from previous deposits, making the archiving process thoroughly transparent.

An important consequence of the release of the costing tool is that it allows depositors to engage with larger issues of selection, appraisal, and retention of digital data. An awareness of these issues at the outset, and at a grassroots level from amongst those generating digital data, has obvious benefits for informing preservation policy at an organisational level, but perhaps more significantly for the preservation policy within the digital archive as well. Discussions have, for example, highlighted that “archivists might be more pragmatic than scientists when it comes to the selection, appraisal and disposal of digital data” (ERPANET 2004, 25). *ADS easy* therefore offers real potential to examine how depositors and archivists conceptualise the significance of data, their methods of appraisal and selection, and the relationship between archivists, depositors and data users. As Chowdhury contends:

It may be noted that digital preservation research has always focused on meeting one main objective: to make sure that the information can be used in future. But use by whom and in what context? In the digital age this is a major question. A given content (information) may have different types of potential users, each with a different characteristic, need and expectation, and the same content may be viewed and used by different types of users differently (2010, 212).

While predicting the future is impossible, a useful starting point for making such informed appraisals is a thorough awareness of current conceptions and what is regarded as significant amongst depositors, archivists and users. As is noted “[s]ome researchers propose the focus of digital preservation research needs to be shifted from systems to users” (Chowdhury 2010, 213); *ADS easy* certainly fosters a much wider engagement with the archiving process by depositors who are, in turn, potential users of the data.

5. Conclusions

The significant advantage of the use of *ADS easy* is that it negates the involvement of the archivist

within the deposition process, passing responsibility for the formation of the collection and the creation of associated project and file-level metadata back to the depositor. This of course streamlines workflows whilst reducing cost, both of which are critical if cost effective, long-term preservation is to become a reality and deposition is to become the norm. The use of the SWORD protocol allows the upload of files directly into the archive in a proscribed form, producing consistency in structure, content and metadata that will facilitate automation within the accessioning process. In addition, some of the processes currently initiated ‘manually’ by the digital archivist during the accession process are also automated using *ADS easy* which generates technical and administrative metadata at the point of deposition. Experience has shown that every interaction with the deposit during ingest increases cost in real terms and also time, indeed the “[a]utomation of processes and more tools for practical application are keys to cost-effective life-cycle management of digital objects” (Harvey and Thompson 2010, DPE 2006). Yet the *Research Roadmap*, produced by Digital Preservation Europe (DPE), recognised that the “level of automation in digital preservation is still very low” (DPE 2006, 28). Although the number of projects have ‘experimenting’ with aspects of automation during ingest and preservation processes is increasing (see Strodl and Rauber 2011 for a useful overview).

The *ADS* has already made significant strides towards automation. *Archaeotools*, through its use of natural language processing, has allowed the creation of tools that have automated the creation of resource discovery metadata (Jeffery et al. 2009), whilst the implementation of *DROID* (Mitcham 2011) and more recently *Apache Tika* have allowed the extraction of file level technical metadata. *ADS easy* allows the potential for further automation within the archiving process, this includes the actual movement of files to the relevant preservation directory, virus checking and checksum generation. There is also some potential for the automation of processes within the preservation process, particularly for ‘simple’ files such as images. Complete automation is impossible; the validation of metadata and the verification of the ‘success’ of file conversions, for example, require the involvement of the digital archivist; but we believe that the development, and future use, of *ADS easy*

will mark a sea change in people's understanding of and engagement with the archiving process.

References

ADS. 2012. "The dark art of costing for digital preservation." *Sword-Arm Blog* March 21. Accessed 20 June, 2012. <http://archaeologydataservice.ac.uk/blog/sword-arm/2012/03/the-dark-art-of-costing-for-digital-preservation/>.

Allinson, J., S. Francois, and S. Lewis. 2008. "SWORD: Simple Web-service Offering Repository Deposit." *Ariadne: Web Magazine for Information Professionals* 54. Accessed June 20, 2012. <http://www.ariadne.ac.uk/issue54/allinson-et-al>.

Austin, T. 2010. "ADS+ and Fedora Commons". *The CSA Newsletter* 23, 1. Accessed 18 June, 2012. <http://csanet.org/newsletter/spring10/nls1001.html>.

Austin, T., D.J. Robinson, and K.A. Westcott. 2001. "A Digital Future for Our Excavated Past." In *Computing Archaeology for Understanding the Past: CAA 2000*, edited by Z. Stančić and T. Veljanovski, 289-296. Oxford: ArchaeoPress.

Beagrie, N., J. Chruszcz, and B. Lavoie. 2008. *Keeping Research Data Safe: a cost model and guidance for UK Universities*. London: Joint Information Systems Committee. Accessed June 19, 2012. <http://www.jisc.ac.uk/publications/publications/keepingresearchdatasafe.aspx>.

Beagrie, N., B. Lavoie, and M. Woollard. 2010. *Keeping Research Data Safe 2*. London: Joint Information Systems Committee. Accessed June 19, 2012. <http://www.jisc.ac.uk/media/documents/publications/reports/2010/keepingresearchdatasafe2.pdf>.

Chowdhury, G. 2010. "From digital libraries to digital preservation research: the importance of users and context." *Journal of Documentation* 66 2: 207-223.

Currier, S. 2009. "SWORD: Cutting through the red tape to populate learning materials repositories." *JISC E-Learning Focus*. February 2009. Accessed June 20, 2012. http://elearning.ac.uk/features/SWORD_paper09.pdf/download.pdf.

Digital Preservation Europe. 2006. *DPE Research Roadmap, DPE-D7.2*. Glasgow: HATII. Accessed June

20, 2012. http://www.digitalpreservationeurope.eu/publications/dpe_research_roadmap_pdf.

ERPANET. 2004. *The Selection, Appraisal and Retention of Digital Scientific Data: ERPANET/CODATA Workshop, December 15-17, 2003*. Lisbon: Biblioteca Nacional. 19 June, 2012. <http://www.erpanet.org/events/2003/lisbon/LisbonReportFinal.pdf>.

Gutmann, M., K. Schürer, D. Donakowski, and H. Beedham. 2004. "Sword-Arm offers real potential to examine how depositors and archivists conceptualise the significance of data." *Data Science Journal* 3: 209-221.

Hardman, C., and J. Richards. 2003. "OASIS: Dealing with the Digital Revolution." In *The Digital Heritage of Archaeology: Computer Applications and Quantitative Methods in Archaeology Archive of Monuments and Publications*, edited by M. Doerr and A. Sarris, 325-329. Athens: Hellenic Ministry of Culture.

Hardman, C.. 2006. "OASIS: sharing information across the profession." *Conservation Bulletin* 51: 26.

Harvey, Ross, and Dave Thompson. 2010. "Automating the appraisal of digital materials." *Library Hi Tech* 28 2: 313-322.

Jeffrey, S., J. Richards, F. Ciravegna, S. Waller, S. Chapman, and Z. Zhang. 2009. "The Archaeotools project, faceted classification and natural language processing in an archaeological context. UK e-Science All Hands Meeting 2008." *Philosophical Transactions of the Royal Society A* 367: 2507-2519.

Kintigh, K.W., and J.H. Altschul. 2010. "Sustaining the Digital Archaeological Record." *Heritage Management* 3 (2): 264-274.

Richards, J.D., C. Hardman, and T. Austin. 2010. "Covering the Costs of Digital Curation." *Heritage Management* 3 (1): 255-263.

Richards, J., and C. Hardman. 2008. "Stepping back from the trench edge: an archaeological perspective on the development of standards for recording and publication." In *The Virtual Representation of the Past*, edited by M. Greengrass and L. Hughes, 101-112. Farnham: Ashgate Publishing Company.

Lewis, S., L. Hayes, V. Newton-Wade, A. Corfield, R.

- Davis, T. Donohue, and S. Wilson. 2009. "If SWORD is the answer, what is the question?: Use of the Simple Web-service Offering Repository Deposit protocol." *Program: electronic library and information systems* 43 (4): 407 – 418.
- Lewis, S., P. de Castro, and R. Jones. 2012. "SWORD: Facilitating Deposit Scenarios." *D-Lib Magazine*, 18, 1/2. Accessed June 20, 2012. <http://dlib.org/dlib/january12/lewis/o1lewis.print.html>.
- Mitcham, J. 2011. "Using DROID to create file level metadata – a case study from the Archaeology Data Service." Paper presented at the Future of File Format Identification: PRONOM and DROID user consultation, November 28, 2011. Accessed June 18, 2012. <http://media.nationalarchives.gov.uk/index.php/jenny-mitcham-using-droid-to-create-file-level-metadata-a-case-study-from-the-archaeology-data-service/>.
- Strodl, S., and A. Rauber. 2011. "A cost model for small scale automated digital preservation archives." In *8th International Conference on Preservation of Digital Objects (iPress 2011)*, 97-107. Singapore: National Library Board Singapore and Nanyang Technological University. Accessed June 20, 2012. http://www.ifs.tuwien.ac.at/~strodl/paper/strodl_ipres2011_costmodel.pdf.

Archaeology in Broad Strokes: Collating Data for England from 1500 BC to AD 1086

Chris Green

University of Oxford, United Kingdom

Abstract:

Landscape and identities: the case of the English landscape 1500 BC - AD 1086 (EngLaId) is a new five-year project, bringing together multiple large datasets in order to map continuity and change in English archaeology from the middle Bronze Age to the Domesday book. This paper discusses issues encountered in collating many very large datasets and presents one possible solution for synthesising such a varied mix of data into a single explorable dataset, based upon spatial and temporal proximity and monument type.

Keywords:

GIS, Large Datasets, Data Collation

1. Introduction

Beginning during the second half of 2011, *Landscape and Identities: the case of the English landscape 1500 BC to AD 1086* (or EngLaId for short) is a five-year project funded by the European Research Council, based at the University of Oxford. The central concept of EngLaId lies in bringing together as many large spatial datasets as possible in order to try to understand identity, continuity and change in the English landscape from the middle Bronze Age to the Domesday book.

Inspiration for this project came from several previous similar surveys of England's archaeology, all of which took a particular period and / or dataset as their central focus (Roberts & Wrathmell 2003; Taylor 2007; Yates 2007). EngLaId is attempting to take a much wider temporal view in order to discover and explain how the rural landscape of England came into being. The middle Bronze Age was taken as a starting point as that is the time from which farming started to settle down into a more fixed pattern, making a more permanent impact upon the landscape than earlier forms of agriculture. The Domesday book was chosen as an endpoint, partly due to its detailed national character and partly due to most of the dominant agricultural trends of the following centuries already being in a quite well-developed state by that time. Our methodology is based around a national survey, followed by a series of regional case studies.

2. Datasets

In order to achieve what we have set out to attempt, we are sourcing and collecting many different datasets. The first of these is the National Mapping Programme (NMP), an English Heritage directed project to map archaeological features from aerial photography. The aim of the NMP is ultimately to cover as much of the country as possible. In practice, this data is a mixture of raster images of line drawings (for earlier projects) and CAD layers (for later projects). Features seen in the NMP are then interpreted, with the results being imported into the National Record of the Historic Environment (NRHE), which has also subsumed the National Monuments Record (NMR). In practice, the NRHE consists of a series of shapefile layers (points, lines, and polygons) associated with an external database (providing attribute data).

To date, we have received data from English Heritage for their south west, south east, and eastern regions, which are based upon the relevant EU parliamentary constituencies. We, thus, have coverage for three out of nine regions: however, by area this is very nearly half of England (47.6%). For these three regions, the total number of records in the NRHE comes to 199,607, of which nearly 70,000 are relevant to our period of interest (including records of uncertain date). Once we have the full database from English Heritage, we can, thus, expect over 100,000 records to fall within EngLaId's timeframe.

Corresponding author: christopher.green@arch.ox.ac.uk

In addition, there are some areas where NMP

interpretation has taken place but for which the results have not been collated into the NRHE. These come from English Heritage's MORPH databases, which map each feature discovered as a separate point (unlike the NRHE, which collates individual objects into sites). So far, we have 38,096 records of this type, of which around 26,000 are relevant to our period of interest (again, including records of uncertain date). Once these are included within our dataset, we can expect the total number of relevant records obtained from English Heritage to be approaching 150,000 (or perhaps much more if many further areas of MORPH coverage exist).

We are also collecting data from England's Historic Environment Records (HERs). These are local bodies responsible for taking care of our country's heritage, particularly in regard to making decisions about planning. They all maintain their own databases and these, due to their more localised focus, inevitably contain many more sites than feature in the databases maintained by any national body. There are over eighty different HERs within England, which were largely established along individual lines before any relevant data recording standards came in. As a result, there is a large variety in the ways in which different HERs record and store their data. To overcome this, we are importing HER data into our own standardised database.

In addition, we have been provided with access to data from the Portable Antiquities Scheme (PAS), which (largely) collects details of archaeological finds made by members of the public. We are hoping to use this to compare distributions of artefacts against landscape features. We are also collecting or hoping to collect data from various other sources, including the Archaeological Investigations Project (AIP) at the University of Bournemouth, the various coin indexes maintained for England, and the data from previous projects (such as Taylor 2007). Once all of these various datasets are gathered together, we can expect our project's total number of data records to number well over 500,000.

3. Synthesis of Multiple Datasets

There is, inevitably, a great deal of overlap between the data contained in the HERs and the data contained in the NRHE. This applies equally to most of our other datasets in addition. However,



Figure 1. Randomly shaded 2 by 2 km grid square tessellation for England.

there remain considerable differences in the content of the datasets, meaning that any one dataset in isolation could only produce an incomplete picture. Therefore, we had to find a way to combine our datasets and to either identify (on a record to record basis) or, at the very least, remove this duplication in order to undertake our national survey. Due to the immense number of data records to cross-reference, this is a not a task that could be performed manually. If every data object had a name or perhaps an identifier code common to all input datasets, then automatically identifying duplication / relationships would be fairly trivial. However, this is unfortunately not the case.

Therefore, we concluded that the only way in which to assess whether two or more records are related is spatially and categorically. In other words, if two or more objects are "close" in space and of the same date and type, we can assume that they are, in fact, the same object (or a continuation / close neighbour of the same object). However, this is further complicated by the fact that objects within our datasets are recorded using different geometries

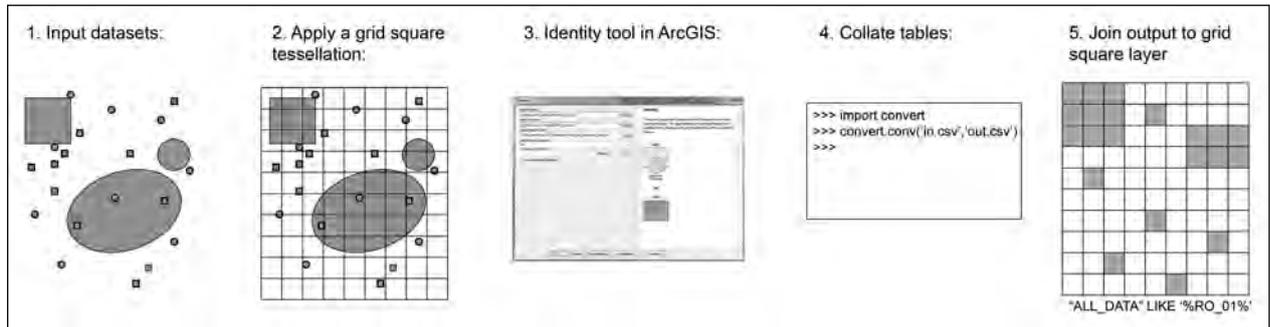


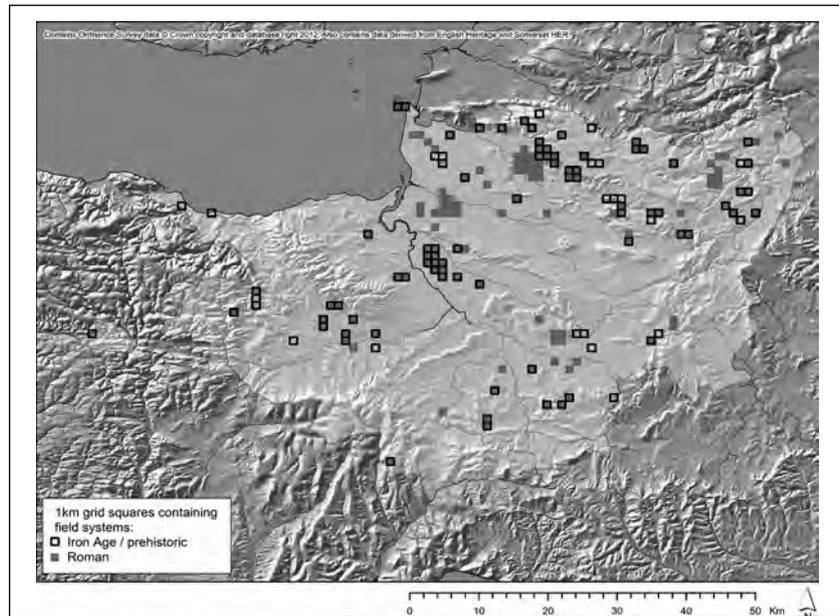
Figure 2. The synthesis process.

(points, lines, areas) and to differing levels of spatial precision. Therefore, how close we mean by “close” is the key question. When plotting objects on the scale of all of England, spatial variations of less than two kilometres would be very hard to differentiate (Fig. 1). As such, we decided to define closeness in this instance as being within a two kilometre grid square (using the British National Grid as coordinate system [OSGB 1936 / EPSG 27700]).

The following process was, therefore, devised by the project’s GIS specialist to test for data duplication using spatial proximity, using ESRI’s ArcGIS 10 (Fig. 2):

- Calculate the spatial precision of point objects within the data based upon the number of zeroes on each coordinate (i.e. no zeroes is the nearest metre, one zero is ten metres, two zeroes is 100 metres etc.). Remove records of worse spatial precision than two kilometres (i.e. those known only to the nearest 10 by 10 kilometre grid square or worse; so far, these records are only a small proportion of the total dataset). Move the remaining points to the centre of their precision squares, so that they do not potentially register as a presence in four squares if they fall on an intersection between squares when they are actually within a single square at some uncertain location (this is done via a simple Python script).
- Simplify the multiplicity of period and monument type terms used across our datasets down to a fixed set of terms. This is accomplished using a Python script that searches through the relevant field in each dataset in order to discover relevant terms related to the types of site which we are interested in. The list of input terms used to generate output terms will be reviewed upon receipt of each new dataset. The output terms are still to be finalised, but are expected to be broad types such as “field systems”, “small settlements”, “large settlements”, “routeways”, etc. However, exactly what output terms to use is still very much an open question. These types are represented by a period code (i.e. “BA” = Bronze Age; “IA” = “Iron Age”, etc.) combined with numerals for each type class (i.e. “field system” = 01, “small settlement” = 02, etc.; e.g. resulting in IA_01 for Iron Age field systems) in order to avoid difficulties caused by the inability of ArcGIS to cope with field lengths of greater than 256 characters (for shapefiles at least).
- Create a tessellation of grid squares (vector) of two by two kilometre resolution. This was done using the Geospatial Modelling Environment software developed by Spatial Ecology LLC (successor to the well known Hawth’s Tools). Each grid square is assigned a unique identifier by the software. This was then restricted down to the extent of England using a spatial intersection query based upon an Ordnance Survey base map.
- Use the Identity tool in ArcGIS to identify which grid square each object in each input dataset falls within (or, in the case of many polygons and lines, overlaps). The tables from the resulting identity layers are exported from ArcGIS and collated into one large table, sorted by grid square unique identifier.
- Using another Python script, work out which types of object (by period) fall within each square and remove duplicated entries. When performing this task, the unique identifiers from each original object are also tracked, so that we can return / build links to the original objects.

Figure 3. Example result of synthesis process for Somerset HER and English Heritage NRHE data. The filled squares show areas of Roman field systems. The squares with thick outlines show Iron Age / uncertain prehistoric field systems. Notice the spread in the distribution from higher ground in the earlier period down onto the marshy Somerset Levels during the Roman period: this implies drainage schemes. This is a preliminary result and should not be over-interpreted, however, especially as the dating of field systems is always difficult.



The result of this script is another table, recording the presence of types of object by period for each square, and also listing the source unique identifiers.

- Join the output table to the grid square layer in ArcGIS.

Once this has been done, the result is a layer consisting of two by two kilometre grid squares recording the presence of different types of archaeological site across England, which can be queried using the standard attribute query tool built into ArcGIS. The main difficulties with this approach are:

That it provides no indication of whether multiple original records are genuinely duplication or genuinely multiple within each square. However, as spatial patterns of this scale would not be apparent when plotting results on a national scale and as unique identifiers are tracked, this should not be seen as a large problem for this type of survey.

That small polygonal objects falling across the intersection between two or four grid squares will appear in each square within the final result, perhaps overstating their importance. We could test for this by offsetting the grid by half the width of each cell east and north, rerunning the analysis, and comparing the results.

That the result is inevitably a simplified one, in terms of spatial location, type and period. However, simplification is the nature of all acts of synthesis.

4. Somerset Test Case

This methodology has been tested initially using data for the county of Somerset, as that was the first relatively large area for which we had both HER and NRHE data. Due to the more detailed spatial scale of this test dataset, 1 by 1 kilometre grid squares were used. The simplification thesaurus used was defined solely by the project's GIS specialist, so the results will undoubtedly be subject to change as this is refined and as more data is gathered. Nevertheless, our test case did prove that the methodology works and produces results which can then easily be interrogated to try to extract patterns of change and continuity in the past (Fig. 3).

The main challenge encountered in implementing this system in practice was in the removal of sites which had been included in the source datasets as being Bronze Age but which were, in fact, early Bronze Age (bearing in mind our middle Bronze Age starting point). This is particularly important for funerary sites as most round barrows (and similar) tend to be early Bronze Age. However, it is still important to include some representation of these sites, as they remained

upstanding features during later periods which would, thus, have affected the decisions made about the location of new landscape features upon their creation. For this scale of survey, however, it seemed acceptable to simply remove objects that were described solely as Bronze Age barrows from the input datasets: removal of other early Bronze Age sites would be more difficult to achieve (where not explicitly recorded), without reading detailed descriptions for each entry.

Once this synthesis process has been undertaken, it is then possible to test distributions against various other spatial data. For example, we might test the distributions of different types of site against coarse-grained slope or elevation data. It is also possible to compare distributions between two or more periods (as in Fig. 3) or between two or more different types of site (e.g. larger against smaller settlements or settlement against field systems). The main difficulty with these latter forms of analysis is that it is not easy to display much more than three or four different distributions at once, but this could be taken as a virtue in itself, as it encourages a focused approach to map creation: mapping too many variables at once can all too often be very confusing to those you expect to make sense of your maps.

5. Conclusions

As with all models of the world, it is inevitable that detail is lost as part of this process and also that some errors will be introduced. For example, if two different representations of the same site in two different databases occur within two different grid squares, these will be registered as the presence of that type / period of site in both locations. It is generally very difficult in these cases to discern whether one record holds incorrect spatial information or whether this is a representation of site that crosses the boundaries of two (or more) grid squares. Furthermore, where multiple sites of the same period and type fall within a single grid square, these will simply be registered as the presence of that type / period of site within the grid square. It is impractical to test for these types of situation across the dataset as a whole, but small sample areas (coinciding with our case study locations) will be tested to attempt to quantify on average how often these issues occur. In order to do so, the system will

be modified to also record counts of each occurrence of each type / period of site in each square. On the whole, however, the reason we create models is to simplify complex processes so that we can attempt to understand them (Yarrow 2006, 77), so relatively minor issues such as these should not overly bias the distributions plotted on such broad spatial and temporal scales.

Naturally, the distribution maps produced as a result of this synthesis process will be biased by different spatial patterns of archaeological work, by different archaeological working practices, by different preservation conditions, and by different masking factors within the natural / built environment. For example, developer funded archaeological excavation is inevitably spatially biased towards areas in which much building or quarrying activity takes place (e.g. AIP data, much HER data, etc.), whereas survey work tends to focus on more open areas (e.g. NMP data, some HER data). Further, the prospects for aerial photographic survey are greatly affected by many factors including soil type, tree cover, the presence of buildings, etc. As a final example, survey work is also greatly affected by factors such as the visual prominence of sites and by the different scales of work undertaken. Some of these biases can be tested for by comparison against measures of ground surface obscuration or density of different types of fieldwork, but some are harder to quantify. By attempting to collate as much data as possible, however, it is to be hoped that these biases can be minimised as much as is feasible.

There is now an immense and ever-increasing amount of archaeological data in existence, especially for western Europe. This provides a brilliant opportunity for producing new studies over long periods of time and broad swathes of space, with modern computerised tools providing the means by which we can bring this vast multiplicity of data together within an analytical environment. However, when combining a large variety of sizable datasets in this way, there are inevitable difficulties over duplication. It is not feasible to test for this manually and automated methods are, thus, essential. One method, as outlined in this paper, is to test for duplication based upon spatial proximity.

This can be done fairly simply by applying a tessellation of grid squares, and through the use

of standard GIS tools and some bespoke scripts. The methodology has been proven to work and can produce interesting results when used to synthesise and interpret archaeological data. The most important element of this is to define a grid square tessellation of an appropriate resolution to the spatial scale of analysis to be undertaken. Both the main advantage and main difficulty of the approach lies in its simplification of complex input data. However, so long as this is not taken to the extreme of oversimplification, painting archaeology with a broad brush has the potential to produce a compelling picture of continuity and change in the past.

Acknowledgments

Many thanks go to the European Research Council, English Heritage (in particular, Simon Crutchley and Lindsay Jones), and Chris Webster at Somerset HER (and all of England's other HEROs).

References

Roberts, B.K., and S. Wrathmell. 2003. *An atlas of rural settlement in England*. London: English Heritage.

Taylor, J. 2007. *An atlas of Roman rural settlement in England*. York: Council for British Archaeology.

Yarrow, T. 2006. "Perspective matters: traversing scale through archaeological practice." In *Confronting scale in archaeology: issues of theory and practice*, edited by G. Lock, and B. Molyneaux, 77-86. New York: Springer.

Yates, D.T. 2007. *Land, power and prestige. Bronze Age field systems in southern England*. Oxford: Oxbow.

Relevant websites

Project blog: <http://englaid.wordpress.com>.

English Heritage's Pastscape (the portal to the NRHE): <http://www.pastscape.org.uk/>.

English Heritage's National Mapping Programme: <http://www.english-heritage.org.uk/professional/research/landscapes-and-areas/national-mapping-programme/>.

English Heritage's National Monuments Record Thesauri: <http://thesaurus.english-heritage.org.uk/>.

Heritage Gateway (links to local HERs and searching of many records online): <http://www.heritagegateway.org.uk/>.

Portable Antiquities Scheme: <http://finds.org.uk/>.

Archaeological Investigations Project: <http://cswb.bournemouth.ac.uk/aip/aipintro.htm>.

Beyond Inspire: Towards Delivering Richer Heritage Data in Scotland

Peter McKeague and Mike Middleton

Royal Commission on the Ancient and Historical Monuments of Scotland, United Kingdom

Abstract:

Through INSPIRE, public bodies are publishing metadata, view and download services for datasets, mandated by the European Union, to an agreed timetable. Mandated datasets help focus attention on priorities but INSPIRE should be seen very much as a catalyst rather than a checklist to unlocking richer archaeological data. Recognising the potential for publishing information as view or Web Map Services (WMS), the Royal Commission on the Ancient and Historical Monuments of Scotland (RCAHMS) has released a point dataset for the National Inventory and is finalising a WMS for its aerial survey mapping programme. In creating remote services, there is a need to consider the user who is often unfamiliar with the complexities of the data presented. Through Defining Scotland's Places, RCAHMS is undertaking the definition of the site extents to assist the end user, including archaeologists and land managers in better understanding the complex information we curate.

Keywords:

National Inventory, INSPIRE, Spatial Data Infrastructures, Web Map Services

1. Background

In 2007 the European Parliament published Directive 2007/2/EC establishing an Infrastructure for Spatial Information in the European Community (INSPIRE) in the official Journal of the European Union (European Communities 2007). The Directive covers 34 Spatial Data themes, grouped in three Annexes, including the Protected Sites theme in Annex I, which covers aspects of the historic environment. The INSPIRE Directive was adopted as a Statutory Instrument by both the United Kingdom and Scottish Parliaments in 2009 with a view to developing the metadata, Web Map and Web Feature Services, to an agreed timetable, over the next decade. The Scottish Government and Geographic Information community in Scotland both recognise that although the mandated datasets are helpful in focusing attention on priorities within the context of creating a Scottish Spatial Data Infrastructure and delivering efficiencies across all tiers of Scottish Government, the INSPIRE Directive should be seen very much as a catalyst rather than a checklist.

2. Adopting INSPIRE

There are two principal drivers for
Corresponding author: peter.mckeague@rcahms.gov.uk

implementing INSPIRE. The legislative approach compels organisations responsible for mandated datasets to publish datasets. However, there are also more informal thematic approaches, where there is a will and a wish to share information for use by the wider community.

2.1 INSPIRE implementation in Scotland

In 2004 the then Scottish Executive, in association with the Association for Geographic Information (Scotland), published One Scotland, One Geography; a Geographic Information Strategy for Scotland (Scottish Government 2005). If One Scotland, One Geography established a vision for sharing spatial information across the public sector in Scotland, adoption of the INSPIRE Directive by the Scottish Parliament (Scottish Government 2009) provides a legal mandate for delivering that vision. To date, a metadata service, the Scottish SDI Discovery Catalogue, using Geonetwork Opensource, has been established (Edina 2011) and guidance 'Cookbook 1: How to serve a Scottish SDI and INSPIRE compliant WMS' (Duffy 2011) published. Work on a further cookbook for publishing WFS is in preparation. Collaborative Public Sector agreements with vendors of core geographies (aerial photography suppliers and Britain's national mapping agency, the Ordnance Survey) provide

access to key spatial reference datasets have been established. A Spatial Information Board has been set up to provide governance and identify and encourage the publication of relevant datasets. However it has been up to the owners of individual datasets to deliver the data they are responsible for.

2.2 The marine environment

If implementation of INSPIRE in Scotland is driven through the legislative process, spatial information about the marine environment is coordinated through the Marine Environment Data Information Network (MEDIN) who host a Data Discovery Portal (Oceannet 2012) incorporating UKLocation, INSPIRE and Gemini 2.1 metadata standards. In contrast to the Scottish SDI, which, to date, has published formally curated datasets, such as protected places, mostly maintained by public bodies the MEDIN recognises that many of the primary data creators are from industry and that new information is created on a regular basis. The less formal approach includes signposting many more digital resources than those published as WMS or WFS.

3. Historic Environment Record in Scotland

In Scotland a number of public agencies hold information relevant to the management of the historic environment. Historic Scotland is responsible for the designation of statutory information, whilst the Royal Commission on the Ancient and Historical Monuments of Scotland (RCAHMS) maintains the national inventory, and associated archival collections, of Scotland's archaeological, built and maritime heritage. Sixteen local authority services also maintain Historic Environment Records (HERs) / Sites and Monument Records (SMRs) either independently or as part of the Specialist User Recording Environment (SURE) in partnership with RCAHMS (Hamilton, Jones and McKeague 2013). Two major landowners, The Forestry Commission (Scotland) and The National Trust for Scotland, also maintain records for management of their respective estates (the latter as SURE partners (Hamilton, Jones and McKeague 2013)). As well as informing the policies and working practices of each organisation, information, curated on behalf of the Scottish public, is essential for the stewardship and management of the historic

environment as well as stimulating research and more general interest in that resource. Historic Scotland, RCAHMS and a number of local authority services have each developed WMS and WFS for their internal business use, including presentation on web-GIS portals (McKeague 2011).

Although information is accessed by a range of users including professional and academic archaeologists, environmental researchers, land managers and the public, in terms of INSPIRE, only those datasets defining statutory designations, including scheduled monuments, Listed Buildings, Gardens and Designed Landscapes and the Register of Battlefields, all maintained by Historic Scotland, are unequivocally covered by the strict interpretation of the Annex I Protected Site theme. Metadata and the WMS addresses for these datasets are published on the Scottish SDI Discovery metadata portal. For other datasets, there is no immediate requirement to publish data and it is likely that most Historic Environment record datasets will be released only as part of the Annex III Area Management/Restriction Reporting Zones/Reporting Units theme.

The arguments for adopting a more expansive interpretation of the Protected Sites theme have been explored elsewhere (McKeague, Corns and Shaw 2012) but within the context of One Scotland, One Geography and contributing to the Scottish Government vision for a Scottish SDI there should be no debate. Retaining information in silos may form a short term coping strategy for spatial information but as the SDI evolves, organisations must publish if they wish their information to remain relevant.

3.1. RCAHMS contribution towards the Scottish Spatial Data Infrastructure

RCAHMS recognises the need to and value in sharing the information it curates on behalf of the Scottish public with partner organisations and the wider community for the benefit of the promotion and appreciation of Scotland's heritage. Through *Future RCAHMS* (RCAHMS 2010), the business plan for 2010-2015, RCAHMS recognises the need to widen digital access to information on Scotland's places, making it more interactive and an integral part of the burgeoning world-wide network of cultural heritage data. Publication in line with emerging Spatial Data Infrastructures

helps meet that objective. Although the majority of records in Canmore (<http://canmore.rcahms.gov.uk/>), the national inventory of the archaeological and built heritage of Scotland and its maritime waters, do not reference sites protected through statutory designation, RCAHMS has argued that the information it curates is relevant to and should be considered alongside INSPIRE Annex I Protected Places theme. To date RCAHMS has released a point-based WMS of the National Inventory with a further WMS for a sub-set of information: maritime losses. The maritime losses WMS comprises separate layers for known wreck sites, documented losses (or casualties) and obstructions. At the time of writing, release of a further WMS publishing the results of the ongoing aerial photographic mapping programme is imminent. The WMS comprises two layers. An interpretative layer maps the archaeological landscape revealed as cropmarks on oblique and vertical aerial photography. As part of the transcription process, using John Haigh's Aerial 6.0 software the source imagery is scanned and geo-referenced and presented as a screen-resolution mosaic complementing the interpretative data. Most of the photographs, informing the transcription process are from RCAHMS Collections but for a minority the copyright of the original material rests elsewhere. Resolution of copyright issues relating to third party material are partly delaying publication of the WMS. In turn, there are also concerns within RCAHMS about the commercial re-use and exploitation of information published through WMS, and particularly WFS, and issues of licencing need to be addressed.

It is a natural progression to assume that remote sensing data, or excavation data, could be similarly supplied as a WMS. However, in contrast to the aerial survey mapping programme undertaken by RCAHMS, remote sensing surveys and excavations are carried out by a range of commercial and research institutions, gathering data for specific purpose. Although some organisations such as Wessex Archaeology have the ability and vision to create an SDI for their data (Cripps this volume) it is beyond the realm and expertise of many smaller organisations and researchers to publish their own data in WMS and WFS formats. As a national inventory and archive, RCAHMS sees a need for coordinating data from third parties to create collaborative datasets.

3.2 Meeting a public need

Promoting the undesignated heritage of Scotland through INSPIRE raises a number of questions over the appropriateness of applying specifications for regulatory environmental data to the wider cultural heritage and how information, so published, can be understood and used remotely by non-specialists. Most archaeological data is ill-defined and incomplete. Would those accessing data remotely necessarily understand the incompleteness, bias and variability of the record in contrast to the fixed boundaries of most designated datasets? Does the information published under INSPIRE meet both non-specialist and specialist audiences – or are separate services required?

The initial release of the RCAHMS point data as a WMS reflects the interests of the heritage professional. Yet in publishing the WMS we cannot anticipate the knowledge and understanding to those accessing that information. As a profession we need to give more consideration to end-users unfamiliar with the nuances of archaeological terminologies and interpretation and how they engage with the information so published. To the non-specialist user, such as a land manager, the form of, or evidence for a monument may be more meaningful than a detailed classification of a monument. In order to provide stewardship for the archaeological resource, users need to know if a site is upstanding, a cropmark, or known only from documentary reference, but not necessarily that the site is a burnt mound, cursus or a fort.

One of the issues in releasing Canmore point data is that the points identify the presence of a site yet leaves the user uncertain of the site's extent. This need is now being addressed through the Defining Scotland's Places project.

3.3 Defining Scotland's Places: a case study in mapping area extents.

From the earliest days of recording at RCAHMS, defining the extent of monuments in the landscape has been a fundamental part of the recording process. Originally mapped by drawing directly on 1:10,000 Ordnance Survey quarter sheets the manual update of what are known as the Record Sheets continued until the late 1990s

Organisation	Historic Scotland	Local Authorities	RCAHMS
What are they mapping?	Designation	The Potential	The Known
Mapping class	Constraint Mapping	Trigger Mapping	Inventory Mapping

Table 1. Scotland's historic environment curatorial sector and what they map.

when integration of the RCAHMS architectural records into the database rendered this process impractical for urban areas. With the advent of digital systems and RCAHMS' policy of sharing data with Local Authority HER/SMR curators a number of local digital maps have been developed by and for specialist heritage users. However, heritage is embedded in planning through guidance including PAN 2/2011: Planning and Archaeology (Scottish Government 2011). Heritage is also a key concern in land management legislation, (Historic Scotland 2011) as well as being a significant draw for tourists (HEACS 2009), meaning that a large, non specialist, audience has developed with an interest in knowing what heritage sites exist and how far they extend.

Since 1998, the public have been able to access information curated by RCAHMS online, with the ability to view site locations, as points, on a map first added in 2003. But, point data is limiting. For example, grid references may be accurate but not precise, meaning that records may plot in what seems to be the wrong place. Another example is a monument that is very large or very long, where a single point may be easily missed or even seen to be misleading. The next phase is, therefore, to reconnect the mapping (Fig. 1).

Known as Defining Scotland's Places (DSP), the project began by working with the sector (RCAHMS, Historic Scotland, ALGAO Scotland and the SMR Forum) to define what mapping already existed, who the audience was, and what their needs are (Casey 2009, Middleton 2009 a and b). One of the most important outcomes of this work was to clarify and define what the different parts of the sector were mapping (Table 1).

These results were used in combination with the obligations imposed by INSPIRE and MIDAS Heritage data standards (English Heritage 2007) to develop a set of polygon standards for Scotland (Middleton 2010).

A significant observation was that there is no nationwide map, made to a common standard, of

known extents available for the non specialist user. This is important as it means the known historic environment cannot be fully protected. Legislation cannot reference a map based data set that will deliver the same level of information to all; grants giving organisations will not use a data set unless it is nationally consistent – a fair playing field for all - and, planners, farmers and land managers remain uninformed to what is on their land as existing mapping is seen as needing specialist interpretation and is therefore not meaningfully promoted.

The aim of the DSP project is therefore to fill this gap by producing a nationally consistent map to a common standard (Middleton Forthcoming) with sufficient attribution to meet data standards

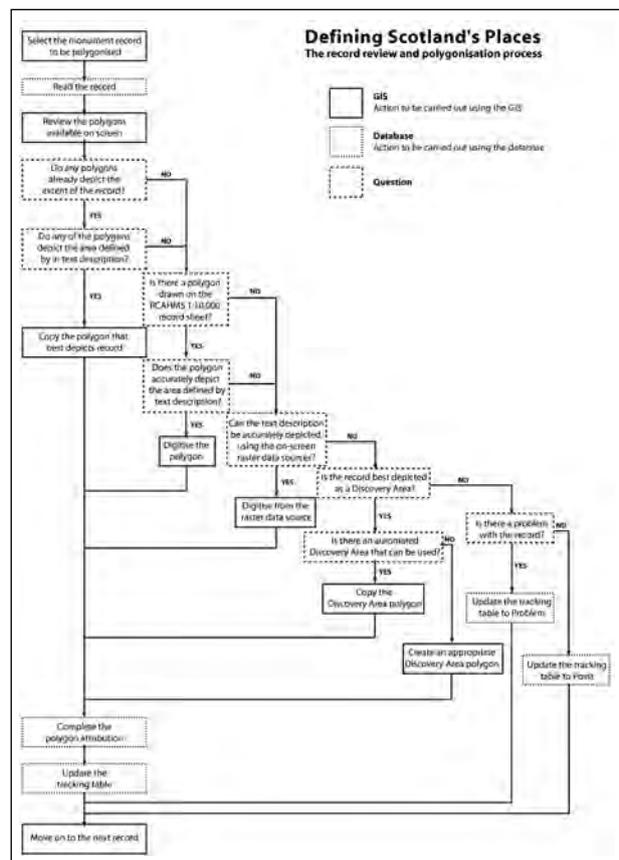


Figure 1. Mapping area extents makes the data immediately more understandable to all users.

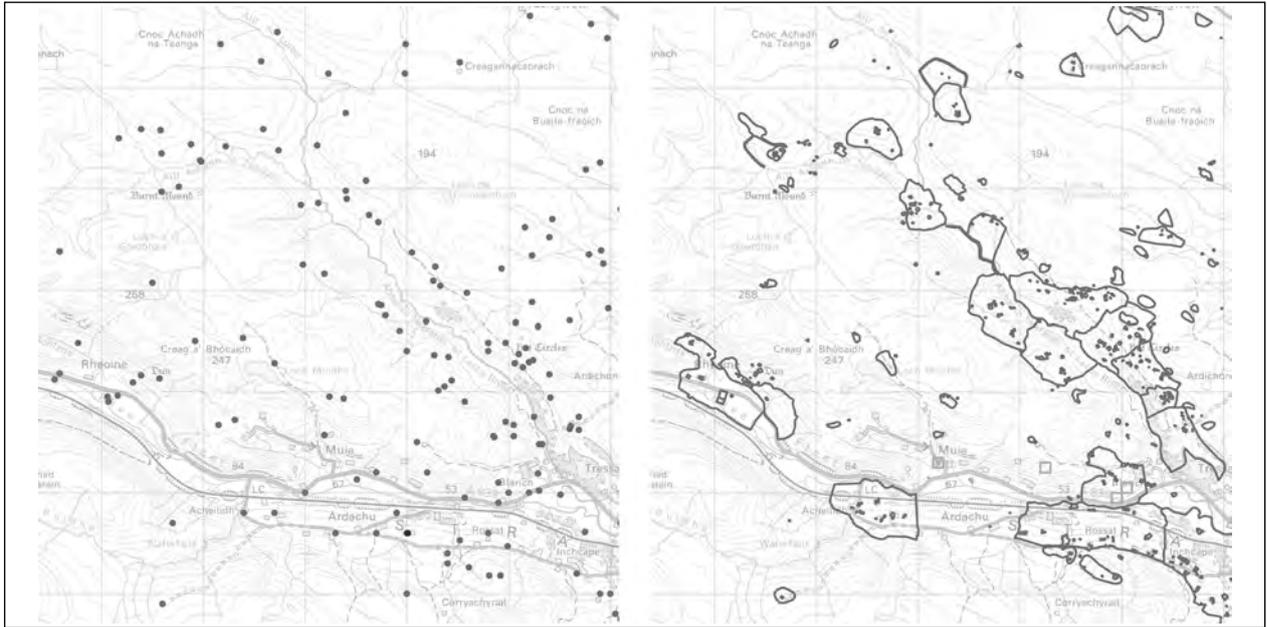


Figure 2. The record review and mapping process adopted by the Defining Scotland's Places project. Simple points are transformed to reflect the spatial extents of the monuments they represent. Base map (C) Crown Copyright and database right 2012. All rights reserved. Ordnance Survey Licence number 100020548.

and to be meaningful to the non-specialist user. We refer to it as a discovery layer with attribution giving the user information on who, what, where, why and when.

- Who: Who to contact if more information is needed and hyperlinks to online data
- What: What the monument is.
- Where: Where the shape came from – the source – and where it is on the ground.
- Why: Why it is of interest.
- When: When the mapped shape was created/updated.

RCAHMS already holds much of this information in the database or in the form of paper maps. But, instead of scanning these maps, RCAHMS has developed a process to review all on-screen data to produce a much more authoritative product and one that takes advantage of existing data (Fig. 2).

The process includes a review of every record and the use of on-screen ortho-rectified aerial

photography, historic maps, survey data and the reuse of existing digital mapping, where available, to produce the most appropriate area extents. To date the pilot project has mapped approximately 10% of the 300,000 records in Canmore and by April 2013, following a yearlong partnership project with the Forestry Commission Scotland, we anticipate having mapped a total of 20%. RCAHMS has also adopted the methodology for all ongoing project work so that new data collected by RCAHMS is mapped and included in the DSP map as a matter of course. A DSP Web Map Service is planned for release in 2013.

The value of this work is already obvious and examples are many: empty highland landscapes which were previously covered with a 'measles' of point data are now maps of highland clearance and upland prehistoric landscapes: the green spaces of modern Erskine show clearly as surviving fragments of a now lost designed landscape (Fig. 3); gaps in the Scottish Roman road system shout out as potential areas for new research; sites previously ignored because point density in survey areas gave the impression they were already recorded, now clearly stand out as new to the record and lost canals, shown on modern maps as drains, regain their connection with our industrial past.

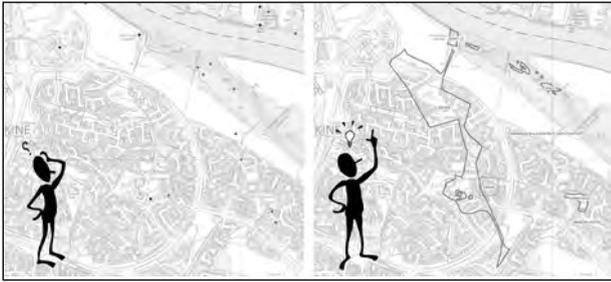


Figure 3. Point data v mapped extents. There is no doubt which is clearer to the user. (Base mapping: © Crown Copyright and database right 2012. All rights reserved. OS Licence No. 100020548)

Locally every area extent is an improvement on the point data. But the most important benefit is that the data is nationally consistent, produced to a standardised template with a core set of attributes that provide all users with basic discovery information on who, what, why, where and when; empowering them to make decisions based on their needs and giving them the detail on who they need to contact if they need more information.

4. Summary

Access to spatial information is changing. Through the development of Spatial Data Infrastructures, organisations are releasing their information for remote consumption. To remain relevant, information about the historic environment needs to be visible, compatible and understandable to those using it. If it is to be considered alongside ecological and bio-diversity data, information about the historic environment must be fit for purpose, conform to standards and be nationally consistent. Information released must consider the end-user and how they view and understand information they are unfamiliar with. Building on the release of a point dataset for the national inventory, RCAHMS is now developing data focused on the end user. Defining Scotland's Places is a dataset that aims to respond to user needs by mapping known extents complemented by user focused attribution. Yet these are just the first steps. A long view is required; for next we must look at data capture by those that deposit with RCAHMS and look to how deposition of data can feed into the next generation SDI.

References

- Casey, S. 2009. "Delivering Efficient Data Management." Accessed 7 June 2012. http://www.rcahms.gov.uk/rcahms_media/files/survey/rcahms_data_management.pdf.
- Cripps, P.J. "SDI: A perspective from a UK Archaeological Unit." Paper presented at the 40th Computer Applications and Quantitative Methods in Archaeology Conference, Southampton, United Kingdom, 26-30 March 2012.
- Duffy, T. 2011. "One Scotland: One Geography: One Information Network: Creating Scotland's Spatial Data Infrastructure: Cookbook 1 - How to serve a Scottish SDI and INSPIRE conformant Web Map Service." Accessed 10 June 2012. <http://www.scotland.gov.uk/Resource/Doc/311518/0098305.pdf>.
- Edina. 2011. "Scottish Spatial Data Infrastructure Discovery Metadata Catalogue. Accessed 10 June 2012. <http://scotgovsdi.edina.ac.uk/srv/en/main.home>.
- English Heritage 2007. "MIDAS Heritage." Accessed 22 October 2012. <http://www.english-heritage.org.uk/publications/midas-heritage/>.
- European Communities. 2007. "Directive 2007/2/EC of the European Parliament and of the Council, of 14 March 2007 establishing an Infrastructure for Spatial Information in the European community (INSPIRE)." *Official Journal of the European Union L 108* Volume 50 Brussels. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:108:0001:0014:EN:PDF>.
- Hamilton, S., R.H. Jones, and P. McKeague. (this volume) "Old places, New ideas: new routes into Canmore, the National Inventory of Scotland."
- HEACS. 2009. "Report and recommendations on the economic impact of the historic environment in Scotland." Accessed 10 June 2012 <http://www.heacs.org.uk/documents/2009/economicimpact.pdf>.
- Historic Scotland. 2011. "Scottish Historic Environment Policy. December 2011." Accessed 10 June 2012. <http://www.historic-scotland.gov.uk/shep-dec2011.pdf>.
- McKeague, P., A. Corns, and R. Shaw. 2012 "Developing a Spatial Data Infrastructure for Archaeological and Built Heritage." *International Journal of Spatial Data*

Infrastructure Research 7: 38-65.

Middleton, M. 2009a. "Polygonisation: The Shape of Things to Come: What are needs for Scottish polygonised Historic Environment data?" Accessed 7 June 2012 http://www.rcahms.gov.uk/rcahms_media/files/survey/rcahms_polygonisation_report.pdf.

Middleton, M. 2009b. "Inspired: The IT capabilities of the Scottish Sites and Monuments Records." Accessed 7 June 2012. http://www.rcahms.gov.uk/rcahms_media/files/survey/rcahms_inspired_report.pdf.

Middleton, M. 2010. "Historic Environment Polygonisation Standards (Scotland)." Accessed 7 June 2011. <http://www.rcahms.gov.uk/historic-environment-polygonisation-standards-scotland.html>.

Middleton, M. *In press*. "Mapping Scotland's Historic Environment: Guide to the Defining Scotland's Places (DSP) project methodologies."

Oceannet. 2012. "MEDIN Catalogue Search version 1.14." Accessed 10 June 2012. <http://portal.oceannet.org/search/full>.

RCAHMS. 2010. "Future RCAHMS" Accessed 7 June 2012. http://www.rcahms.gov.uk/rcahms_media/files/news/future_rcahms.pdf.

Scottish Government. 2005. "One Scotland, One Geography; a geographic information strategy for Scotland." Accessed 10 June 2012. <http://www.scotland.gov.uk/publications/2005/08/31114408/44098>.

Scottish Government. 2009. "INSPIRE (Scotland) Regulations 2009" Accessed 10 June 2012. <http://www.legislation.gov.uk/ssi/2009/440/contents/made>.

Scottish Government. 2011. "Planning Advice Note (PAN) 2/2011: Planning and Archaeology." Accessed 10 June 2012. <http://www.scotland.gov.uk/Publications/2011/08/04132003/0>.

SEAD - The Strategic Environmental Archaeology Database Inter-linking Multiproxy Environmental Data with Archaeological Investigations and Ecology

Philip Iain Buckland
Umeå University, Sweden

Abstract:

The volume of data on past environmental and climate changes, as well as human interactions with these, has long since passed the level where it is manageable outside of large scale database systems. The Strategic Environmental Archaeology Database project aims to not only store and disseminate such data, but also provide tools for querying and analysing them, whilst maintaining a close connection with the archaeological and ecological data that are essential for their comprehensive interpretation. Large scale, geographically and chronologically unrestricted databases provide us with essentially unlimited scope for putting individual sites into a broader context and applying locally collated data to the investigation of earth system level changes. By providing integrated access to data from a variety of proxies, including plant macrofossils, pollen, insects and geochemistry, along with dating evidence, more complex questions can be answered where any single proxy would not be able to provide comprehensive answers.

Keywords:

Environmental Archaeology, Database, Archive, Software, Palaeoecology

1. Introduction

SEAD, the Strategic Environmental Archaeology Database, is an open access archive for environmental archaeology and Quaternary science data. It aims to provide easy access to the raw data from a variety of investigation types where, primarily, biological or physical proxy data have been used to study the past. As an environmental archaeology database, the focus naturally leans towards the human and cultural aspects of prehistory, but the “natural”, Quaternary science, background data essential in the interpretation of these phenomena are considered fundamental to the long-term usefulness of the system. Modern calibration and environmental survey data (e.g. insect pitfall traps, vegetation surveys) are also within the scope of the system. These are limited in number at the moment, but are scheduled to play an increasingly more important role in quantitative landscape and archaeological reconstructions, as science demands more empirically supported interpretations. This paper discusses a number of data ingestion and archiving related aspects of the project, the initial development phase of which runs from 2008 to 2013. It is hoped that the information here will

Corresponding author: phil.buckland@arke.umu.se

prove useful to those considering undertaking similar ventures, or considering entering their own data into an existing archive such as SEAD. Readers interested in more applied aspects of the project are kindly directed towards Buckland et al. 2010.

2. Project Outline

The primary objective of SEAD is to provide a research infrastructure for environmental archaeology and palaeoenvironmental science. Specifically, systems and pathways for 1) data collation, storage and management; 2) integration, access and dissemination; 3) analysis and visualisation and 4) networking and support. These aims are to be achieved through the long-term financing of a stable resource based on expert domain science and technology. SEAD is a bottom up, research driven project, in that it originates from an expressed need from primary researchers and is designed and implemented by the same, in collaboration with technology experts. An appreciation of the need for large scale databases is essentially nothing new (e.g. Maurer et al. 2000; see earlier CAA volumes for archaeological examples), and the SEAD initiative stems from the Environmental Archaeology Lab's observation

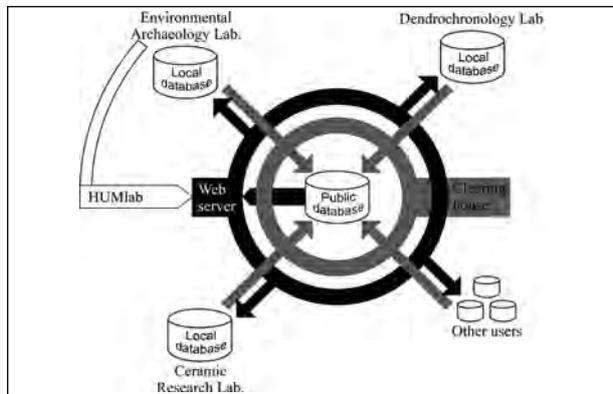


Figure 1. SEAD project structure and overview data flow model. Development is led by the Environmental Archaeology Lab (MAL), in collaboration with HUMlab, who in turn run the web server. Data submitted by partners and users are vetted by the clearing house before publication in the public database.

that the large amount of raw data accumulating its archives (paper and digital) were no longer manageable without a systematic data storage and management system. In common with much archaeological research and consultancy data, these data are also largely the results of publicly funded projects, and as such subject to open access requirements. An online database system is the ideal solution for public dissemination of digital data, and with the addition of data-entry and quality control systems can provide a trusted resource for archiving research data.

Whilst these are worthy goals, the SEAD team believe that the addition of advanced data interrogation and analysis tools are essential if we are to realise the vision of transdisciplinary science, especially with respect to the use of data from multiple institutions, regions and disciplines. These tools may not only save research users time when aggregating sample data from multiple sites, over multiple chronologies, or synthesising habitat data for species, but also inspire and empower them to interrogate or analyse the data in new ways. The team firmly believe that breakthrough science will be enabled by systems such as SEAD.

2.1 Project organisation

The database and associated software is a cooperative venture between the Environmental Archaeology Lab. (<http://www.idesam.umu.se/>

<http://www.humlab.umu.se/english/>) and HUMlab (<http://www.humlab.umu.se/english/>) at Umeå University, Sweden. It is funded by the Swedish Research Council, with co-funding from the Department of Historical, Philosophical & Religious Studies at Umeå University. Closely related research and development projects have been funded by Umeå University Faculty of Arts, and systems analysis and data entry of ceramics and dendrochronological data have been funded by the Nation Laboratories for Ceramic Research and Dendrochronology respectively at Lund University, Sweden (Figs 1 and 3). The project is advised by national and international reference groups, and close collaboration with the Neotoma Palaeoecology database ensures the optimal use of funds, by sharing ideas, software tools and data. As well as being an open access database, the project is run on open development principles, in that those interested in collaborating, both in terms of software development and data entry, are actively encouraged.

2.2 Data collation, storage and management

Data are entered into SEAD through a downloadable software application which includes an internal copy of the database (see <http://www.sead.se> for download). This copy can be synchronised with the online master at any time to provide local access to the latest datasets, and thus ensure that newly entered data can be analysed with respect to the latest data before publication. Users may of course wish to analyse their data with respect to a snapshot of the database, providing a stable base for quantification. Once ready for publication, the data can be submitted to the SEAD clearing house (one or more experts in the relevant field) for quality control (Fig. 1). This includes an assessment of the completeness of the data, with special attention being paid to the often neglected meta- and spatial location data, and a dialog with the submitter for any potentially inconsistent or extreme values. As archaeology is a science where it is most often the anomalies that are interesting, expert knowledge is essential for differentiating unusual data from errors.

2.3 Transparency

A fundamental concept in the SEAD project is that of data and analysis transparency; that research

results and interpretations should be traceable to the raw data underpinning them and that methods should be thoroughly described and reproducible. All datasets are uniquely identified (DOI is supported, but has yet to be implemented), permanent and version logged. The user can identify exactly which data were used to undertake an analysis, even when the datasets used have been subsequently revised. This also facilitates the retesting of earlier hypotheses with updated data, and the online interface supports saved viewstates which can be used to save, share and retrace analysis pathways. Datasets and selected metadata are linked to an internal bibliography, and the system is capable of providing citations at multiple levels, so that the nature of the publication's content can be assessed (e.g. publication of primary data; reanalysis of existing data; summary site interpretations). Virtual constituent databases (or master datasets) are used to help retain the identity of datasets ingested from other sources, or projects wishing to maintain their individual profiles, whilst taking advantage of access to the larger archive and analysis system, without the substantial costs involved in creating and managing it.

Transparency is also essential in connection with the storage of modern reference data. At the time of writing, this includes mainly insect ecology, distribution and climate data from the BugsCEP database (Buckland & Buckland 2006), but the aim is to expand the database, either internally or through linking, to cover other proxies and fields such as botany and ethnography. The methodology of environmental reconstruction, and the interpretation of archaeological deposits, ranges from subjective descriptions, based on the expert knowledge of an individual researcher, to complex mathematical models which are difficult to understand for those without expert knowledge. Giving users access to the raw source data allows them to see where interpretations come from and even assess the viability of reconstructions based on their understanding of the models and the description of the methods (Fig. 2).

2.4 System and database structure

SEAD is built around a PostgreSQL relational database, with online and offline components written in Java, Javascript and PHP. The highly normalised

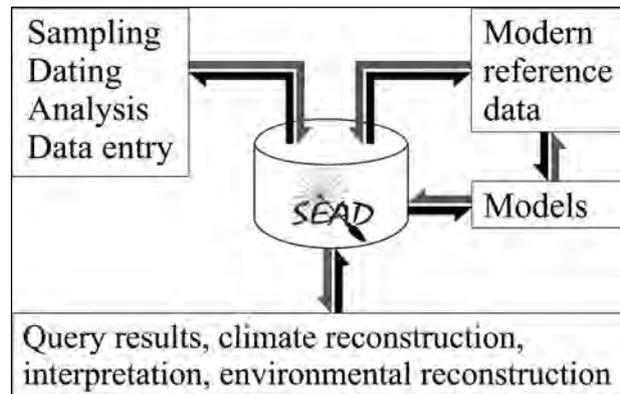


Figure 2. Schematic of transparent database/software based use of SEAD, showing the flow of data (light arrows) and transparency of process (dark arrows) from models to results, and vice versa.

relational model, though perhaps not the most flexible system (when compared to object orientated or more abstracting database models), provides a level of detail and analytical power difficult to realise through more abstract models without a more considerable programming overhead. The structure is extensible, in that data tables are able to accept any form of continuous, discrete or categorised values or measurements and that any type of sample or feature metadata can be included. The entry of metadata types is regulated by the clearing house to prevent duplication or ambiguity. As the database implements a standard minimum divisible unit, that of a single analysis of a single sample, further tables can easily be added in order to extend the scope into the territory of new data forms, such as spectroscopic data, or linked to external databases. The potential for linking to external databases is also facilitated by permitting the storage of multiple identifiers for objects common to archaeological research (e.g. sites, features, contexts, samples etc.). This means that the same sample can be referred to by lab number, museum number or field number as appropriate.

The database stores comprehensive details and references for all methods (e.g. proxy analysis methods, pre-treatments, measurements, coordinate systems, processing methods), and care is taken to assign method descriptions to any area of metadata which would require a qualified description for quality assurance. Methods may also be grouped, allowing for multiple methods for each analysis type, such as the use of different photospectrometers

in phosphate measurement. Whilst the details of many methods may be considered irrelevant for many users, they are essential for the assessment of different methods, and providing background data for improving existing and developing new methods. The system is also capable of recording uncertainty values where relevant and available. To paraphrase an anonymous attendee of another conference, “true transparency, is not achieved by providing the data, but by providing enough information to enable the assessment of the quality of the data”.

The latest database design, modelled using MicroOlap (<http://www.microolap.com/>), can be found on the project website at <http://www.sead.se/database/>.

2.5 Data scope

Whilst SEAD’s structure technically allows it to cover an essentially unlimited scope of methods in environmental archaeology, the current scope of the project (Table 1) is confined by its budget. This is in no way an indication of future limits, integration of the ceramics and dendrochronological components having begun after the project’s first three years. Other laboratories and research groups are actively encouraged to contact the SEAD team with respect to further expanding the scope of the database and analysis tools. In the long term, large scale, sustainable and queryable research archives can only be realised through community based data entry and development. The project is particularly interested in integrating archaeological and Quaternary science data, the latter dominated by studies of “natural” environmental change not directly related to archaeological investigations, and more often than not conducted from the point of view of the natural sciences. As well as being interesting in their own right, these studies provide data and information essential for the study of past human activities, not only in terms of the background climate and environment, but also for describing the immediate history and surroundings of archaeological sites.

SEAD has no chronological limits, meaning not only that any fossil data may be entered, but also any modern data. Currently, this extends only to insect survey data, which will be used as proof of concept for the integration of fossil and

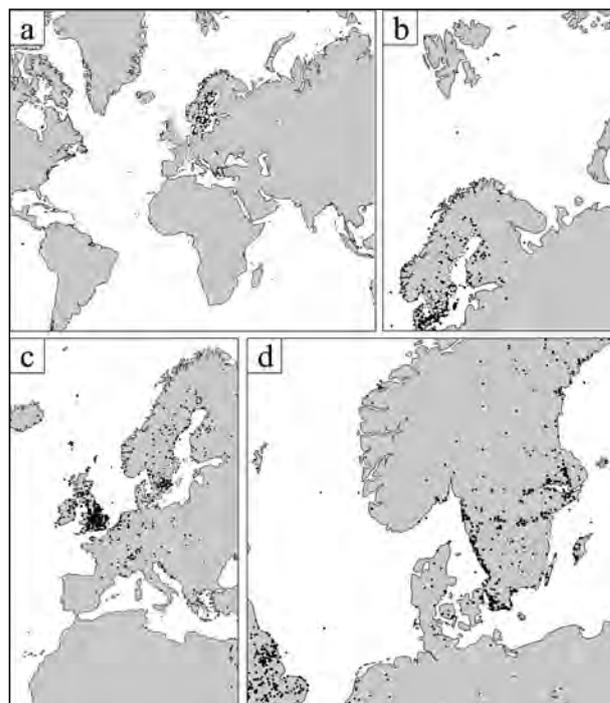


Figure 3. Four maps showing the geographical location of sites currently being ingested into SEAD. The white points in each map indicate the sites originating from one of the initial four master datasets: a) BugsCEP; b) Environmental Archaeology Lab, Umeå; c) National Laboratory for Ceramic Research, Lund; and d) National Laboratory for Dendrochronology, Lund (Småland test data). Note that coordinate errors may still be present as all the data have yet to be checked.

modern data, building on work undertaken in the Bugs database project (Buckland 2007). This will improve the efficiency with which researchers can interpret sites, but also opens up possibilities for refining palaeoenvironmental reconstructions. With time, similar data will be added for plants and geoarchaeology in order to build up a powerful reference database. Similar projects are well established for landscape reconstruction through pollen (Mitchell 2011), although they are not yet represented by integrated databases.

2.6 Geographical scope

The majority of data currently included in SEAD’s data entry schedule are European, reflecting the research scope and history of the labs involved in the initial development. Figure 3 shows the contribution of each of the data sources to the current

Proxy data sources	
Biological proxies	Raw counts of insects/arthropods, plant macrofossils, pollen, molluscs
Geoarchaeology	Soil chemistry (pH, phosphates) and physical properties (conductivity, organic content, colour)
Ceramics	Thin section quantification and properties (e.g. tempering material, inclusions, firing temperature, vessel characteristics)
Dendrochronology	Dates and support data, tree species, building history, dated object description and location (e.g. church, west tower, third beam from roof)
Dating evidence	
Scope	14C and other radiometric methods, dendrochronology, archaeological typology dates, period classifications, calendar dates and ranges, tephras
Chronological extent	Theoretically unlimited, but current range from 2.4 MyBP to present day.
Bibliographic data	
References	May be linked to site, sample group, sample and dataset levels as well as to methods, ecological codes and more.
pdf files	For references where available and not restricted by copyright
Modern reference data	
Abstracted text	Insect habitats and distributions, abstracted from trusted sources, with citations
Coded descriptors or classifications	Insect ecology, in-house system and Koch (1989-1992). Used for quantitative habitat reconstruction/visualisation (Buckland 2007)
Climate	Beetle Mutual Climatic Range (MCR) temperature reference data
Location data	
Coordinates	Three dimensional at site, sample group and sample levels (latitude, longitude, altitude and project survey grids). Capacity for national grid based storage. See section 2.6 and figure 3 for current geographical extent
Depth	Multiple types of depth recorded as positive or negative numbers, e.g. depth from lake/soil surface, depth from datum line or reference level (especially useful for stratigraphic sequences)
Archaeological, geological and sampling data	
Descriptive metadata	Site, feature and sample metadata to allow correlation between environmental and archaeology or lithology datasets. Sample names (e.g. field label, lab number, museum number). Descriptive information for all objects
Sample dimensions	Capacity for multiple measurements (volume, size, weight) at multiple stages of analysis (initial sample, analysed subsample, residue etc.). Position in sample group
Images	Scans, photographs, plans etc. for multiple levels in the site hierarchy and reference data (e.g. seeds, insects)

Table 1. Outline of the initial data scope of SEAD. A full list of metadata will be provided in the database documentation.

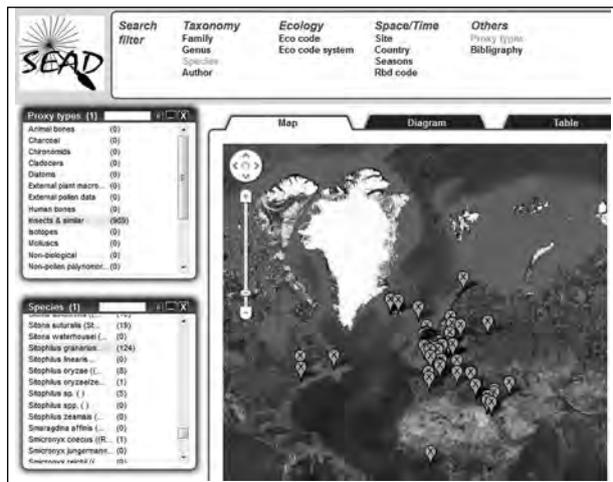


Figure 4. QSEAD, the online faceted browser for the SEAD database. Facets are classified by theme and activated by clicking on their names along the top of the screen. Selected filters appear on the left, collapsible to for an easier overview. Records may be multi-selected within each filter, the results cascading to subsequent filters and the results displaying after a short delay, dependent on the complexity of the query and the amount of data involved. The example uses Google Maps to show the location of the 127 sites which make up the databased fossil record of the grain weevil, *Sitophilus granarius* (L.). Map imagery ©2012 NASA, TerraMetrics. Insect data from the BugsCEP dataset (Buckland & Buckland 2006) in which citations for original sources can be found.

geographical scope of the database. At the time of writing, the dendrochronological data consists only of a small proof of concept digitalisation, roughly 1.5 % of the Dendrochronology lab's archived data (Meissner et al. 2012).

3. Interfaces and Analysis Tools

The SEAD online interface, referred to as QSEAD after its aesthetic and technical origins in the QVIZ inter-archive project (www.qviz.eu; Palm 2008, 2009), is based around a faceted browser system using selectable, rearrangeable and cascading filters (Fig. 4). Each filter passes its results to the next one on the list, and can be rearranged should the logic of the research question require it. Figure 5 shows an example of the type of (palaeo)biogeographical and archaeological research which is made considerably easier by the system. The inclusion of relevant archaeobotanical and geological data in the system will increase the scope for such studies, both through

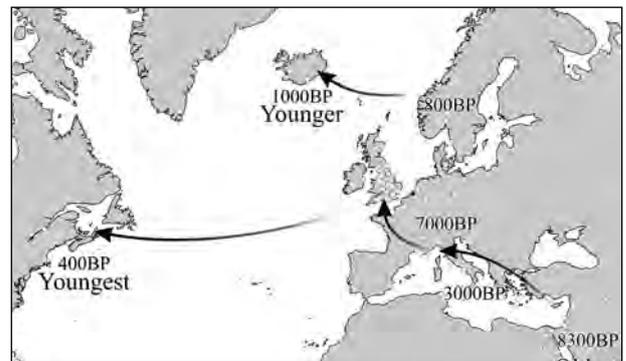


Figure 5. An example research application of the results exported from Figure 4. The spread of points and dates shows the state of (databased) knowledge on the palaeobiogeography of the species and highlights some obvious gaps, such as the relatively late dates from Greece and southern Scandinavia. Arrows show the probable spread of the grain weevil with agriculture and transport of store products. See Plarre (2010) and references for a more comprehensive study of the species.

the investigation of parallel lines of evidence and the capacity to undertake the synthetic analysis of multiple proxies.

The system is currently able to show results as a list of datasets in tabular form as well as on maps, and customisable reporting functions are in the pipeline. Data can be exported (csv, xml), as a composite dataset of aggregated samples where the query includes multiple sites. Later, the ability to transfer the same dataset to more advanced analysis functions within the online software will be provided. A climate reconstruction explorer is under development at the moment, which will allow the faunal implications of insect based climate reconstructions to be investigated, summary statistics generated, and the relative importance of species in an assemblage assessed. Further tools are planned, including more flexible versions of the abundance and habitat diagrams and statistics provided in BugsCEP (see Buckland 2007), which can be applied to any data of appropriate type.

The engine behind QSEAD is a flexible, semi-automated, abstracting query system based on a network graph of the database (Palm 2009). Whilst this works well for many queries and up to a certain level of database complexity, additional web services are under development to provide database subsets which can be chained together for more

complex queries. These services will also allow for easier integration with other databases and external software.

4. Data Entry and Ingestion

The four initial data sources are described briefly below with respect to their scope, quality and specific problems relating to their ingestion. In SEAD terminology, these are referred to as master datasets, denoting their nature as virtual constituent databases. There is no limit to either the number of datasets stored under a master dataset or the volume and scope of its data. Although each of the master datasets are continually being added to, the import of legacy data is the main focus below.

4.1. The BugsCEP master dataset

BugsCEP is a downloadable, Microsoft Access and Visual Basic for Applications based insect ecology database and software package which contains the large part of the Quaternary fossil insect record (<http://www.bugscep.com>, Buckland & Buckland 2006; Buckland 2007). It contains over 1050 sites (Fig. 3a), most of which include quantified abundance data in the form of minimum numbers of individuals of each taxon found in each sample. The database includes ecology, distribution and climate reference data, as well as a comprehensive bibliography. The software incorporates a number of advanced data interrogation and analysis tools. Bugs has its origins in the 1980s (Sadler et al. 1992) and was presented in a previous incarnation at the 2001 CAA meeting in Gotland, Sweden (Buckland & Buckland 2002).

Having existed in some form for over 30 years, there have been various versions and improvements. Each stage has refined the data, parsing text and memo fields, and added to the scope of the project. The database is partially normalised, and includes a number of compromises which enabled the more rapid development of interfaces. These compromises have occasionally resulted in a lack of structural referential integrity, which in combination with insufficient checks on data entry have allowed inconsistent data to be entered, and important data to be omitted. However, the ingestion of Bugs into SEAD has comprised of the relatively simple task of devising SQL schema for mapping the two

database structures, with a set of integrity catches which report to the data manager. This report can be used to correct any problems in BugsCEP or SEAD before publication, either using the interfaces or via backend hacking. Although simple in nature, this task requires a large amount of manual checking given the more than 10 500 taxa and 144 000 fossil records in the database. As always, the most vital resource in this undertaking is the knowledge of experts in the field. The assessment of data quality cannot be undertaken by inexperienced technicians or students, and training is therefore essential for long-term sustainability.

The BugsCEP software provides a number of analysis tools which could be usefully applied to other proxy data sources, and it is the author's intention, given time and funds, to reproduce all of these features online. Some of this will be achieved through the SEAD project, but until it is fully realised, and for the benefit of current users, the two systems must run in parallel. This has required a synchronisation system based on the SQL transactions which can cope with alterations as well as additions and deletions. Bugs unfortunately has a more limited and free text based approach to metadata, the automated linking of archaeological databases being outside of its scope. Sites are, by practice, uniquely named on data entry. The BugsCEP unique site identifier has no external relevance and is not seen by the user, and sites are unique on the basis of their name, location, date and references combined.

The primary linking attribute of the Bugs species reference data is that of the unique identifier for an insect taxon, the name alone being insufficient due to differences between international and the numerous national taxonomies. Where Bugs can only hold a single identifier for each taxon, SEAD can store multiple parallel taxonomies and thus record and export occurrences in whatever taxonomic system the user chooses. The database stores a comprehensive and referenced synonym list, but interrelating these taxonomies for data retrieval may be far from simple, and more problems will be encountered where pollen are concerned (e.g. Birks & Birks 2000). The approach of this project is, therefore, based on the concept of getting the data in and sorting out the details later, a principle which works best in the presence of

comprehensive metadata recording and consistency of data entry. Ontologies are being developed parallel to this process, rather than in advance, due to the complexity of the data and large number of unknowns.

4.2 The Environmental Archaeology Lab. (MAL), Umeå, master dataset

Research and consultancy in environmental archaeology has existed in some form since the 1970's at Umeå University. Consequently, MAL (Miljöarkeologiska laboratoriet) has amassed a considerable amount of data (Fig. 3b), much of it analogue and very little of it in database ready form. The lab's activities have encompassed a number of methods, but most frequently the analysis of plant macrofossils, pollen, geoarchaeology and insects (the latter covered by BugsCEP above). Although much of the recent data (from at least the year 2000) are stored digitally, far from all datasets are in easily ingestible spreadsheet or database form (Buckland et al. 2006). Even the spreadsheets, which should theoretically be based on a small number of templates, have had columns moved, renamed and new ones added due to the lack of a strict data handling regime. Occasionally, the wrong data have been entered under a heading, with an explanatory note in another column. These factors are a severe hindrance to the automated import of data, and have necessitated the development of data ingestion tools which allow spreadsheet columns to be checked and mapped to the appropriate analysis methods or metadata tables. Fully automated ingestion is out of the question due to the unpredictable nature of the variation.

A primary goal of ingesting the MAL dataset, and one of the pillars of the SEAD project, is that of making so called grey literature publicly available. These reports, although always available on request, are often only included as appendices in archaeological reports and would be more conveniently available online. SEAD will make the raw data which underpin or are included in the reports available online and queryable, with the reports available as pdf files. Metadata will be ascribed, national antiquities numbers checked, and English summaries provided for Swedish texts. The short reports presenting null results will also be included as a guide to future studies. This work is

extremely time consuming, and requires qualified staff capable of undertaking detective work in the reports and paper and digital archives (which span five generations of Macintosh). A considerable part of the project budget goes to such assignments, a fact that is sometimes forgotten when planning or reviewing large scale database projects.

It may be prudent to mention at this point that, from the contract side of archaeology, it is the scientific literature that may be considered as "grey" in that it is not accessible without membership of a university or through considerable fees. The publication of data from these publications, if not the papers themselves, in open archives is a step towards a more democratic dissemination of scientific data where those not fortunate enough to have a university post may make use of them.

If SEAD is to achieve the vision of its designers, then all sample and analysis data should be cross-queryable – that is to say that it should be possible to identify the commonality of samples based on their location, chronology and properties, between projects. This is no simple task, and the data detective must deal with obscure or badly recorded sample descriptions and identifiers, missing or incorrect dating information, partial data, missing metadata and even the occasional bit of cryptography (older ecologists have been known to code their identification data). Where multiple proxies have been used on an excavation their results may be published in separate reports, sometimes with little indication of having come from the same samples. Part of the solution has been the use of standardised spreadsheets for the intermediate storage of data whilst problems are sorted out. An alternative exists where pollen data are concerned, in that Erik Grimm and the Neotoma project's (Neotoma 2009) Tilia software is able to read older Tilia files and export them as XML or spreadsheets. These files can then be easily imported into SEAD, leaving the main issue as the finding of metadata associated with each sample. Unfortunately, such metadata are rarely stored together and in unfortunate cases common sample names were not used for all analyses. In such cases, it is only by the interrogation of staff involved in the projects that an acceptable level of metadata can be extracted (a process that is familiarly referred to as "chaining senior researchers to their desks" within the SEAD project).

4.3 The National Laboratory for Ceramic Research, Lund, master dataset

At the inception of SEAD it was never conceived that ceramics data would form part of its scope. After discussions with researchers at Lund, however, it was clear that certain commonalities existed between the environmental data and aspects of the ceramics data. In particular, it was decided that the ingestion of already partially digitised data from the analysis of ceramic thin sections would be highly desirable, and SEAD was adapted to accept them (Fig. 3c). A number of the recorded properties of thin sections are essentially geochemical or can be recorded as the presence/absence of specific components. The facility to cross-query the ceramics and macrofossil data of multiple sites holds great potential for archaeological research. Among other things, it could expedite contributions to the study of plants as foods and medicines, crop processing, agricultural development, diets and rituals by facilitating the integration of different streams of evidence for the same activities.

The ingestion of the ceramic data serves to illustrate an interesting problem in the assimilation of legacy data: that of recording absence. In this master dataset, representing data accumulated by several researchers over the past 30 years, but also including some data from the 1950's, there are five different absence recording methods: blank/null, "no", "no information", "missing" and "undefined". For the sake of simplicity, it has been decided to assume that all of these mean the absence of whatever was being recorded, although as the authors of earlier investigations are inactive we may never know the truth. There is always a danger of losing resolution when digitising or migrating between database systems. To ensure data transparency the intermediate spreadsheets, which contain this variety, will be preserved in case the variation has meaning. The general problem of recording absences is perhaps one of standardisation, and common to many databases. Problems occur more often when merging from different data sources, where internal standardisation may not have been consistently applied to each source. We can, however, at least hope to try to maintain that null is a state and not a value.

4.4 The National Laboratory for Dendrochronology, Lund, master dataset

With the exception of a few dates, dendrochronological data are the latest edition to SEAD (Fig. 3d), and a test dataset consisting of material from the Swedish county of Småland was entered in early 2012 (Meissner et al. 2012). Although data from a number of archaeological projects (35) were entered, using intermediate spreadsheets, the majority of the data relate to building heritage management (71 projects). The latter orientation was also an extension of the initial mandate of SEAD, and required the expansion of generic description tables to enable the attribution of multiple types (e.g. building type, building purpose, building structural form etc). Although dendrochronology is often considered primarily as dating evidence for the context in which wood is found, there is a large amount of measured and descriptive data behind every date. For SEAD to provide anything but a date retrieval system for these data, it was therefore essential that it be able to store these supporting data. This was achieved by a minor modification of spatial data tables, so that they could cater for a descriptive location within a structure; and the addition of a specific dendrochronological support data table. The existing structure provides for the storage of tree species, and allows the retrieval of all analyses performed on a particular species, be they macrofossil, pollen, ¹⁴C or dendrochronological.

Entering the remainder of the archived dendrochronological data is estimated to take the equivalent of four years (Meissner et al. 2012), a monumental task planned to start in 2013. In addition to the above, the Dendrochronology Lab. has also undertaken the analysis of wood in wrecks and living trees. With time, these datasets will also be entered into SEAD and provide an extremely useful continuity between ancient and modern measurements, land and sea.

5. External-linking

The BugsCEP project has shown the advantages of being able to rapidly summarise the ecology of fossil insects found in samples, with over 80 publications citing use of the latest version alone (see <http://bugscep.com/publications.html>). Bugs has been managed from the environmental

archaeology point of view, with an understanding of the importance to archaeology of Quaternary science data. Although the latter may be obvious to some, there are many archaeologists who believe that only samples taken on-site may reveal useful archaeological information. This is a misconception which must be overcome if archaeology is to progress and collaborate with other palaeoenvironmental sciences. By doing so, it may increase its contribution to the body of accepted knowledge on an integrated picture of past human activities and environmental and climate change. As ecologists become more interested in palaeobiogeography, perhaps with the explosion of molecular studies of potential origins and refugia (e.g. Schönwetter et al. 2005), there is an indication that they themselves may themselves become more interested in palaeoenvironmental databases (see Brewer et al. 2012), especially when the DNA, fossil record and morphology may disagree.

Within Sweden, the standard cultural heritage identifier is the RAÄ (Riksantikvarieämbetet) number, the equivalent of the British SMR number. The combination of parish name and RAÄ number are supposedly unique, although mistakes have occurred and this cannot be guaranteed. Discussions are underway on how to resolve this and similar issues and enable reliable linking between SEAD and the Swedish National Heritage Board's databases. The latter are undergoing a large scale redesign and there is good scope for improved low level accessibility (Lars Lunquist, RAÄ, pers comm.). Recently, the Swedish National Data Service (SND) has released excavation data in GIS form (<http://snd.gu.se/sv/node/564>), an important step in creating online archaeological databases with a level of detail sufficient for complex multi-site queries.

An important issue yet to be fully dealt with is the representation and visualisation of sites without spatial data or dates. These will not show up on maps or timelines respectively without special attention, and may contain useful data which should not be omitted from a query. A similar problem is posed by sites/objects with multiple locations, such as an Egyptian mummy stored at the British museum.

6. Concluding Remarks

The creation of large scale research data

infrastructures is essential for scientific advancement in an increasingly data-rich world. Open access to published data, as well data citation, encourages data reuse and large scale syntheses which would otherwise be extremely difficult. Providing easier access to previous results effectively bootstraps research within and across disciplines (Mooney & Newton 2012). Such initiatives may also help to satisfy public and political demands for greater research transparency, the importance of which was made particularly clear during the "climategate" affair, where lack of transparency was used as one of numerous arguments against the viability of evidence for global warming (Maibach et al. 2012). The examples provided in this paper highlight the fact that although there are considerable advantages to integrated data sharing, the construction and maintenance of such infrastructures is far from simple. It not only requires long-term funding and commitment from a core research community, preferably distributed over multiple research groups, but perhaps more importantly a persistent user base. SEAD is still at an early stage and it remains to be seen as to whether it will achieve the required momentum for long term success.

Expert systems have been discussed earlier in the history of the Bugs project (Buckland et al. 1997), the implications being of a software system that could provide explanations for new datasets based on the existing database. This is still a goal for many palaeoecologists, even if, in the absence of artificial intelligence, the preferred term may now be decision support systems. Ultimately the system should be able to provide a list of statistically similar sites or samples to those selected, along with multiple environmental reconstruction scenarios and other tools to aid interpretation. There is however, always a cost based limit to the amount of "intelligence" that can be built into a system, after which the designers must rely on the intelligence of the user. Even with advanced ecology based, quantitative environmental reconstruction, a detailed knowledge of environmental science and the ecology of the organisms/proxies used for reconstruction is required, if the reconstruction is to be reliably interpreted. It is never a case of push a button and get an answer, but always push the right button and get something that will help provide potential answers to the research questions.

Acknowledgements

The SEAD project is a team effort, in cooperation with the Lund laboratories, and the work of all those involved should be acknowledged. Thanks in particular to Fredrik Palm and Erik J. Erikson for help with the conference presentation. Roger Engelmark should also be mentioned for pressing the burgeoning idea of the database home with his colleagues. I am also grateful to the Swedish Research Council and its external reviewers who provided useful comments and permitted SEAD to be funded. As always, the database would be nothing without the data contributors (see the database for full citations).

References

- Birks, H.H., and H.J. B. Birks. 2000. "Future uses of pollen analysis must include plant macrofossils." *Journal of Biogeography* 27(1): 31-35.
- Brewer, S., S.T. Jackson, and J.W. Williams. 2012. "Paleoecoinformatics: applying geohistorical data to ecological questions." *Trends in Ecology & Evolution* 27(2): 104-112. doi: 10.1046/j.1365-2699.2000.00375.x.
- Buckland, P.I., Y. Zhuo, D. and P.C. Buckland. 1997. "Towards an Expert System in Palaeoentomology." In *Studies in Quaternary Entomology - An Inordinate Fondness for Insects*, edited by A. C. Ashworth, P. C. Buckland, and J. P. Sadler, 71-77. Chichester: John Wiley & Sons Ltd. http://bugscep.com/phil/publications/bucklandetal1997_expert.pdf.
- Buckland, P.I., and P.C. Buckland. 2002. "How can a database full of Bugs help reconstruct the climate?" In *Archaeological Informatics - Pushing the Envelope - CAA 2001 - Computer Applications and Quantitative Methods in Archaeology, Proceedings of the 29th Conference, Gotland, April 2001*, edited by G. Burenhult and J. Arvidsson, 453-461. Oxford: Archaeopress. http://bugscep.com/phil/publications/buckland&buckland2002_caa.pdf.
- Buckland, P.I., and P.C. Buckland. 2006. *BugsCEP Coleopteran Ecology Package*. Boulder: IGBP PAGES/World Data Center for Paleoclimatology Data Contribution Series # 2006-116. NOAA/NCDC Paleoclimatology Program. Accessed 24 June 2012. <http://www.ncdc.noaa.gov/paleo/insect.html> and <http://www.bugscep.com>.
- Buckland, P.I., J. Olofsson, and R. Engelmark. 2006. *SEAD - Strategic Environmental Archaeology Database, planning report*. MAL Reports 2006-31. http://bugscep.com/phil/publications/bucklandetal2006_SEADplanning.pdf.
- Buckland, P.I. 2007. "The Development and Implementation of Software for Palaeoenvironmental and Palaeoclimatological Research: The Bugs Coleopteran Ecology Package (BugsCEP)." PhD diss., University of Umeå, Sweden. <http://www.diva-portal.org/umu/abstract.xsql?dbid=1105>.
- Buckland, P.I., E.J. Eriksson, J. Linderholm, K. Viklund, R. Engelmark, F. Palm, P. Svensson, P.C. Buckland, E. Panagiotakopulu, and J. Olofsson. 2010. "Integrating Human Dimensions of Arctic Palaeoenvironmental Science: SEAD - The Strategic Environmental Archaeology Database." *Journal of Archaeological Science* 38(2): 345-351. doi:10.1016/j.jas.2010.09.011.
- Koch, K. 1989-92. *Die Käfer Mitteleuropas. Ökologie, 1-3*. Krefeld: Goecke & Evers.
- NEOTOMA. 2009. "The Neotoma multiproxy palaeoecology database." Accessed 30 May 2012. <http://www.neotomadb.org/>.
- Maibach, E., A. Leiserowitz, S. Cobb, M. Shank, K.M. Cobb and J. Gullede. 2012. "The legacy of climategate: undermining or revitalizing climate science and policy?" *WIREs Clim Change* 3: 289-295.
- Maurer, S.M., R.B. Firestone and C.R. Sriver. 2000. "Science's neglected legacy." *Nature* 405: 117-120. doi:10.1038/35012169.
- Meissner, K., P.I. Buckland, D. Hammarlund, and H. Linderson. 2012. "Pilotprojekt 'Dendro-databas' i SEAD." MAL rapporter nr. 2012-23. Lund: Umeå universitet & Lunds universitet. http://www.sead.se/files/pilotprojekt_dendro_2012.pdf.
- Mitchell, F.J.G. 2011. "Exploring vegetation in the fourth dimension." *Trends in Ecology & Evolution* 26(1): 45-52. doi:10.1016/j.tree.2010.10.007.
- Mooney, H., and M.P. Newton. 2012. "The Anatomy of a Data Citation: Discovery, Reuse, and Credit." *Journal of Librarianship and Scholarly Communication* 1(1): eP1035. <http://dx.doi.org/10.7710/2162-3309.1035>.

SEAD - The Strategic Environmental Archaeology
Philip Iain Buckland

- Palm, F. 2008. "Crossing the Archival borders." In *IST-Africa 2008 Conference Proceedings*, edited by P. Cunningham and M. Cunningham, 57-63. IIMC International Information Management Corporation. http://qviz.eu/downloads/ISTAfrica2008_Paper_ref_57.pdf.
- Palm, F. 2009. "Abstracting query building for multi-entity faceted browsing." *Lecture Notes in Computer Science* 5822:53-63. doi:10.1007/978-3-642-04957-6_5.
- Plarre, R. 2010. "An attempt to reconstruct the natural and cultural history of the granary weevil, *Sitophilus granarius* (Coleoptera: Curculionidae)." *European Journal of Entomology* 107: 1-11.
- Sadler, J.P., P. Buckland, and M. Rains. 1992. "BUGS: an entomological database." *Antenna* 16: 158-166.

Obscura itinera: a GIS-based Approach to Understand the pre-Roman and Roman Transhumance Pathways in Umbria and Sabina Regions (Central Italy)

Paolo Camerieri

Regione Umbria - Promozione della Qualità del Patrimonio Paesaggistico Regionale, Italy

Tommaso Mattioli

Independent Researcher

Abstract:

This paper presents two different methods (topographic and GIS based analysis) applied by the authors to recognize the layout of transhumance pathways in Umbria and Sabina regions (Central Italy) and to investigate the spatial correlation between pre-Roman 'castellieri' (Italic hillforts) and transhumance routes.

Keywords:

Roman Centuriation, Transhumance, Pre-Roman Hillforts, Cost Distance, Umbria and Sabina

1. Introduction

Horizontal transhumance is the seasonal migration of people with their domestic livestock between fixed summer and winter pastures, in Italy's south-central regions from the higher Apennines mountains to the southern lower valleys in Puglia region (Fig. 1).

Here the transhumance pathways, the so-called 'tratturi' in modern Italian language (and their branches, 'tratturelli' or 'bracci'), are a kind of very long 'green highways' with a considerable width, sometimes up to 100 mt, a strip of land not cultivated but left as enclosed tracts of pastures to allow the sustenance of the flocks during the long-distance journey (Fig. 2).

The most significant examples of these pathways are documented between Abruzzo and Puglia regions where this traditional practice has shaped the landscape by being practiced up to modern time since Late Antiquity and Middle Ages with similar underlying procedures for taking advantage of remote seasonal pastures (Fig. 3).

This network of pathways lasted until modern times almost intact in south Italy, as if it had been

crystallized in the landscape, thanks to Alfonso V of Aragona, king of Naples in 1442, who established in 1447 a new laws (the so-called *Prammatica Aragonese*) to regulate the transhumance of the Kingdom of Naples (Gabba & Clementi 1990).

Unfortunately the rest of the Italian peninsula, such as Papal States in Umbria and Sabina regions north of the border with the Kingdom of Naples, has undergone a different fate: the lack of a medieval legislation to regulate the 'tratturi' network caused their disappearance at the end of the Roman period and today they are barely recognizable in the landscape (e.g. in Umbria Spada 2002).



Figure 1. Summer and winter pastures in south-central Italy and the network of 'tratturi' in the Kingdom of Naples.

Corresponding author: pcamerieri@libero.it



Figure 2. The 'tratturo' between Trivento and Lucito (Campobasso, Molise, South Italy).



Figure 3. Aerial picture of the 'tratturo' between Trivento and Lucito (Campobasso, Molise, South Italy) (from GoogleEarth, 41° 44' 36" N, 14° 40' 33" E).

But we know from Latin sources that in Roman times the 'calles' (the transhumance pathways in Latin) came very far north, up in Umbria and Sabina². Probably these pathways are the 'obscura itinera' mentioned by the Latin author Frontinus³ as the best route chosen by the Roman consul M. Curio Dentato in the autumn of 290 BC to move fast from Samnium to the north for the conquest of Sabina: "M. ' Curius adversus Sabinos, qui ingenti exercitu conscripto, relictis finibus suis nostros occupaverant,

² Varro, *De Re Rustica*, 2.1.17; 2.2.9; 2.8.5; 3.17.9.

³ Frontinus, *Stratagemata*, 1.7.4.

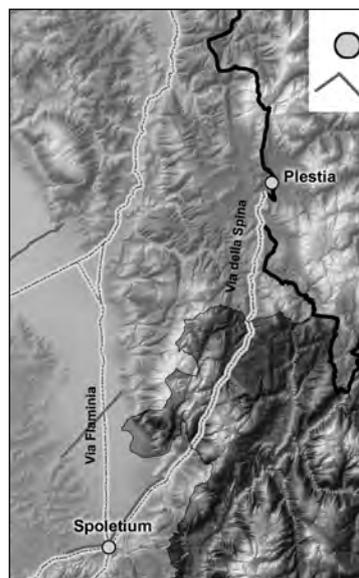


Figure 4. The study area with the 'callis' found in the Roman centuriation, the main Roman road and the main connecting point used in the GIS cost distance analysis.

occultis itineribus, manum misit, quae desolatos agros eorum, vicosque per diversa incendit".

Going back in time, according to several authors (Corbier 1991; Giardina 2005; Hermon 2001; Maggi et Al. 1991; Narcisio 1991; Whittaker 1988) we can therefore establish that the origin of horizontal transhumance in Italy dates to pre-Roman age but the great exploitation of this economy dates back to the Roman period: during the III century. BC the Roman conquest of inland south-central Italy, the confiscation of several lands once belonging to the Italic communities and then their transformation into 'ager publicus', both for grazing or agriculture, led Rome to a widespread exploitation of the migration of domestic livestock, even through the investment of considerable capital and the regulation of trades (Pasquinucci 1979).

This paper presents two different methods applied by the authors to recognize the layout of transhumance pathways in Umbria and Sabina and to investigate their relationship with pre-Roman settlements: a purely topographical approach and a GIS-based analysis. The study area (967 km²) is located in south-est Umbria and is included in the municipalities of Campello sul Clitunno, Cascia, Cerreto di Spoleto, Monteleone di Spoleto, Norcia, Poggiodomo, Preci, Sant'Anatolia di Narco, Scheggino, Sellano, Vallo di Nera (Fig. 4)⁴.

⁴ All coordinates are in the system EPSG:3004 Monte Mario / Italy zone 2 (see <http://spatialreference.org/ref/epsg/3004/>)

2. The Topographical Approach

The research carried out in 2009 for the celebration of the birth of the Emperor Vespasian has concerned the study of Roman centuriation of Umbria and Sabina regions (Camerieri 2009a, 2009b, 2009c; Camerieri et Al., 2009; Camerieri & Mattioli, 2011). The results have unexpectedly brought to light, in a singular and eloquent way, the existence of a single act of Roman land reorganization of this area that followed the brief military campaign of Manio Curio Dentato in 290 BC and the presence of transhumance paths within the geometric land division of the intramontane basin of Reate, Falacrinae, Nursia, Villa San Silvestro and Amiternum.

The Roman plot of land divisions of these basins, identified by historical cadastral maps and aerial photographs, showed us some modern roads (low-volume roads not paved or surfaced with gravel), that, although they are following and emphasizing the ancient location of roman Kardus and Decumanus, they are not tracing a straight direction: they seem to move like a zigzag in a constant range of amplitude (3 actus or 106 metres), while maintaining a direction strictly consistent with the Roman field division (Fig. 5).

This unusual road fluctuation within a fixed range showed evident analogy with the shape of the 'tratturi' in the Kingdom of Naples (Fig. 2), that closely follow the oldest roman routes and the main axis of centuriation. Here the modern road (if present) behaves in a similar way and surprisingly the amplitude of fluctuation is always maintained within a range of about 110 meters, a multiple of Roman actus and exactly the width of the 'tratturi reali' (royal 'tratturi') prescribed by the Aragonese law in the XV century.

In conclusion, the topographical study of Roman centuriation provided us with a new key to locate where the 'calles' or 'tratturi' (transhumance pathways) cross the intramontane plains during Roman period and, above all, informs us that Romans, during their massive public work projects, probably regulated an older transhumance pastoralism already performed by the Italic population.



Figure 5. An example of a modern unpaved road moving in a constant range of amplitude (3 actus = 106 metres) into the roman field division while maintaining a direction strictly consistent with the Kardus (Nursia, Umbria, Central Italy).

3. The GIS-based Analysis

Transhumant pastoralism involves the movements following already established and precise routes which are repeated every year. We know where these routes entered in (or left) the intramontane centuriated plains of Umbria and Sabina and, as consequence, we can try to identify what are the paths connecting them through the use of a GIS cost-distance analysis. A list of the connecting points is provided in Table 15.

As it was previously discussed, the transhumance pathways in Umbria and Sabina seem to date back to pre-Roman age. The Italic mobile pastoralists will probably require securing the resources they need as well as having flexible social and spatial patterns of resource use. During this activity, one of the principal problems they probably encounter is to deal with the right of passage through a landscape that may not be established as private property and to ensure control of the territory. During the pre-Roman period the archaeological record of Umbria and Sabina mountains is almost exclusively composed of particular structures, the so-called 'castellieri' (the name derives from 'castello', a fortified borough), developed from the late Bronze Age to the Roman conquest, a kind

⁵ The GIS analysis take into consideration also other archaeological sites well known in the archaeological literature (Bonomi 1985, 2001).



Figure 6. Aerial photograph of the 'castelliere' at Forca Canapine (Norcia) (courtesy of Umbria Region).

of hillforts usually located on the top of hills or mountains constituted by one or more concentric series of walls built with uncut stones, of rounded or elliptical shape within which was the inhabited or sacred area (Figs 6 and 7). As discussed by other authors (Bonomi 1985, 2001), it has been suggested that these structures functioned as markers in the landscape with the purpose of visual control of movements or boundary defense.

Following this argument, our secondary aim in the GIS analysis is to establish through a statistical

Name	Type	X	Y
Corvatello	'Callis' end point	2363270	4718152
Falacrinae Nord	'Callis' end point	2369249	4717984
Falacrinae Ovest	'Callis' end point	2368397	4717482
Fonte Marucigna	'Callis' end point	2364400	4722142
Interamna Nahars	pre-Roman settlement	2326523	4714079
Monteleone di Spoleto	pre-Roman settlement	2352209	4724287
Nursia	pre-Roman settlement	2364053	4739220
Piani di Castelluccio	main pastures	2372767	4736834
Pidiripa	'Callis' end point	2365315	4733862
Plestia	pre-Roman settlement	2349092	4766719
Reate	'Callis' end point	2343654	4696787
Spolegium	pre-Roman settlement	2334820	4734462
Via Salaria	main roman road	2363753	4709923
Villa San Silvestro Nord	'Callis' end point	2360620	4723292
Villa San Silvestro Sud	'Callis' end point	2359588	4718239

Table 1. The list of the connecting points analyzed by GIS cost distance tool.



Figure 7. Oblique aerial photograph of the 'castelliere' at Monte Pianciano (Spoleto).

computation the spatial correlation (if any) between pre-Roman 'castellieri' and transhumance routes identified by GIS cost distance and to verify if the 'castellieri' settlements could restrict or ensure the access to plentiful and predictable resources (the pastures). The list of 'castellieri' and their geographical features is provided in Table 2⁶.

Traditionally, it has been believed that transhumance routes were developed taking into account diverse natural variables: time of traveling, the effort consumed, terrain types and ground knowledge, the access to basic resources while traveling, territoriality, social networks and other relationships (Murrieta-Flores 2010).

In this research we performed two GIS experiments regarding two factors that we believe could affect mobility in pastoral societies: the time consumed during the traverse in all the landscape of the study area (path 1) and the optimal path on a friction surface obtained from a weighted overlay of vegetation cover, natural obstacles (rivers) and the altitude above sea level (path 2).

The first experiment concerns a simple concept of mobility (e.g. the best path to reach quickly one location from another), the second one takes into consideration the environment and its target is to

⁶ The archaeological data come from 'CAU (Carta Archeologica dell'Umbria) project and in particular from the analysis of aerial photographs made in 1989 by Prof. Francovich and Dr. Cosci at the University of Siena. The castellieri were classify as 'verified' and 'not verified', to according to data coming from direct surveys.

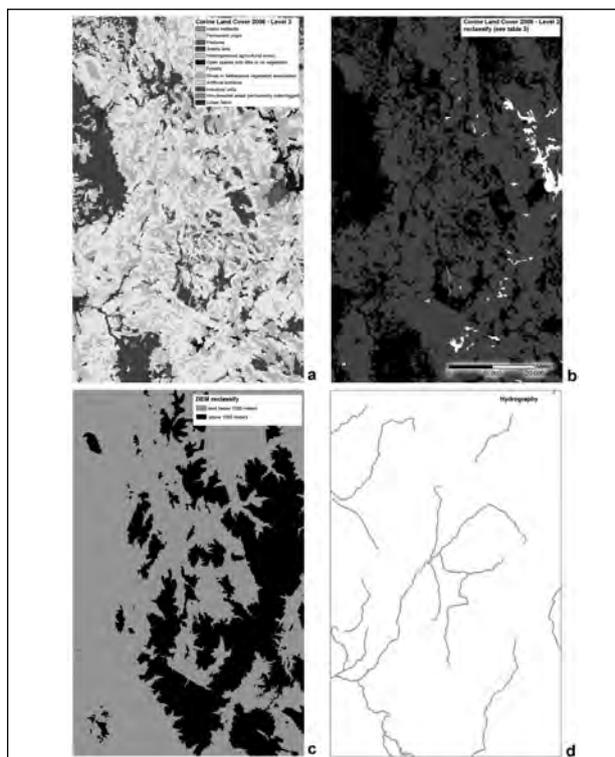


Figure 8. The three raster datasets reclassified and weighted as shown in table 3 to obtain the friction surface.

by deep, canyon-like river valleys with a high volume rate of water flow. We classify rivers by their discharge, the volume of water which passes through a given cross-section of the river channel per unit time.

4. Results

As shown in the figures 10 and 11, the GIS analysis has provided two different paths. Surprisingly the best network of paths to reach quickly one location from another (path 1 = 1861 km, 13,56° average slope) is longer and with an higher average steepness than the optimal network of paths to allow the shepherd and the herd to cross favorable surface with grass for grazing as longer as possible (path 2 = 1350 km, 9,57° average slope). This occurs due to the fact that usually the optimal paths (path 2) rapidly reach high altitudes and move in a straight line across sub horizontal plane, while best paths (path 1) avoid large gradients of slope.

In order to establish the visual correlation (if any) between pre-Roman 'castellieri' and the two pathways and to verify if the 'castellieri' settlements

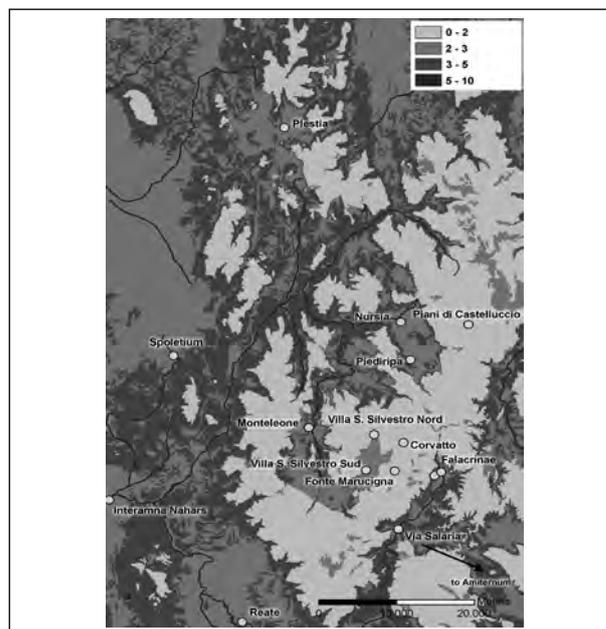


Figure 9. The friction surface obtained from a weighted overlay among the hydrography shapefile, the DEM and the Corine Land Cover 2006 (see Table 3).

could restrict or ensure by direct control the access to plentiful and predictable resources (the pastures), we explored the visual control of the 'castellieri' by the application of a viewshed analysis with Spatial Analyst extension in ArcGIS 9.3.1 (ESRI). Viewshed analysis was based on the same raster

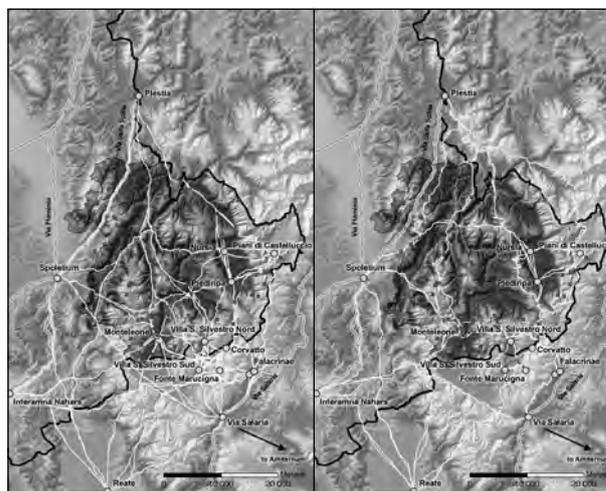


Figure 10. Left. Cost paths 1 based on Tobler's 'hiking function' with the verified 'castellieri' (red square) and unverified 'castellieri' (blue square). **Figure 11.** Right. Cost paths 2 based on friction surface with the verified 'castellieri' (red square) and unverified 'castellieri' (blue square).

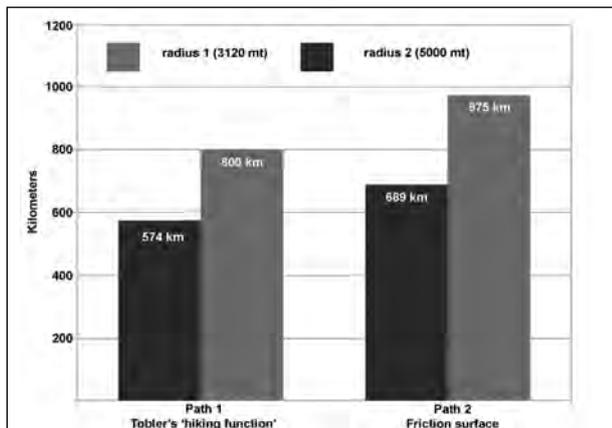


Figure 12. Kilometers of path 1 and path 2 under the visual control of 'castellieri' within a radius of 3120 and 5000 metres.

digital elevation model (DEM) (20m cell size); two offset (2 meters) were used to define the height of the observer point (castelliere) and the height of the feature considered for visibility (path1 and path2). According to a previous study on visual control based on the Higuchi human eye's degree of perception of different objects/targets (MATTIOLI 2008) the search distance was limited to 3120 metres (radius 1, territorial control of landscape elements) and 5000 metres (radius 2, horizon).

Figure 12 show how many kilometers of path 1 and path 2 are under the visual control of 'castellieri' within a radius of 3120 and 5000 metres. The Italic settlements seem to have a greater aptitude for visual dominance on the optimal paths for the transhumance (path 2) rather than best path to reach quickly one location from another (path 1).

Furthermore a confirmation of this hypothesis is also provided by Figure 13 that shows how many hectares of land with different vegetation cover are controlled by 'castellieri': their interest seems to be aimed primarily toward land with mixed vegetation (forest or meadows suitable for grazing,) carefully avoiding the visual control of open spaces with little or no vegetation (scree, cliffs, rocks and outcrops) or deep, canyon-like river valleys.

5. Conclusions

The origin of transhumance routes of Umbria and Sabina regions seem to date back to pre-Roman age. The Roman centuriation of the intramontane

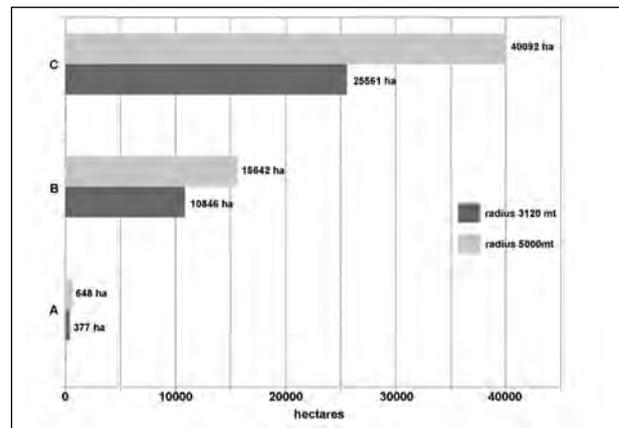


Figure 13. Hectares of land with different vegetation cover (Corine Land Cover 2006) controlled by 'castellieri' at 3120 meters and 5000 meters radius: A. open spaces with little or no vegetation (scree, cliffs, rocks and outcrops); B. permanent crop, pastures, grassland, field sown with seed, heterogeneous agricultural areas; C. forests, sclerophyllous vegetation, transitional woodland shrub.

basins in Apennines mountains seems to keep trace of this older network of pathways and, according to data deriving from GIS cost distance and viewshed analysis, the typical fortified settlements of Italic populations (so-called 'castellieri') seem to prefer, among the best pathways, the visual control of those paths crossing favorable surface with grass for grazing as long as possible.

Acknowledgments

We would like to thank the Regione Umbria and the Soprintendenza per I Beni Archeologici dell'Umbria for making available the records and cartography of the Carta Archeologica dell'Umbria (CAU).

References

- Bonomi Ponzi, L. 1985. "Topographic survey of the Colfiorito di Foligno plateau: a contribution towards the study of the population in the territory of the plestini." In *Papers in Italian archaeology 4*, edited by C. Malone, S. Stoddart, 201-238. Oxford: Archaeopress.
- Bonomi Ponzi, L. 2001. "Tra Appennini e Tevere: il ruolo dei Naharci nella formazione della cultura umbra." *Annali della Fondazione per il Museo Claudio Faina* 8: 319-341.
- Camerieri, P. 2009a. "La ricerca della forma del catasto

- antico di Nursia nell'odierno Piano di Chiavano." *Divus Vespasianus. Il Bimillenario dei Flavi. I templi ed il forum di Villa San Silvestro*, edited by Filippo Coarelli, 41 – 47. Roma: Quasar.
- Camerieri, P. 2009b. "La ricerca della forma del catasto antico di Reate nella pianura di Rosea." In *Divus Vespasianus. Il Bimillenario dei Flavi. Reate e l'Ager Reatinus, Catalogo della Mostra*, edited by Filippo Coarelli, 39-48. Roma: Quasar.
- Camerieri, P. 2009c. "Le valli interne dell'alta Sabina e le antiche vie di transumanza." In *Divus Vespasianus. Il Bimillenario dei Flavi. Le origini di Vespasiano*, edited by Filippo Coarelli, 40-44. Roma: Quasar.
- Camerieri, P., A. De Santis, and T. Mattioli. 2009. "La limitatio dell'Ager Reatinus. Paradigma del rapporto tra agrimensura e pastorizia, viabilità e assetto idrogeologico del territorio." In *Agri Centuriati, VI*, edited by P.L. Dall'Aglio and G. Rosada, 325 - 345. Pisa and Roma: Fabrizio Serra editore.
- Camerieri, P., and T. Mattioli. 2011. *Transumanza e agro centuriato in alta Sabina, interferenze e soluzioni gromatiche, Lazio & Sabina, VII*. Roma: Quasar.
- Corbier, M. 1991. "La transhumance entre le Samnium et l'Apulie: continuités entre l'époque républicaine et l'époque impériale." In *La romanisation du Samnium aux I^ee et I^{er} siècles av. J.C. (Atti del convegno, Napoli, 4-5 novembre 1988)*, 149 – 176. Napoli: Publications du Centre Jean Bérard.
- Gabba, E., and A. Clementi. 1990. *Atti delle Giornate internazionali di studio sulla transumanza (L'aquila, Sulmona, Campobasso, Foggia 1984)*. Padova: Bottega d'Erasmus.
- Giardina, A. 2005. "Uomini e spazi aperti: transumanza e transumanze" In *Storia di Roma, IV, Caratteri e morfologie*, edited by E. Gabba and A. Schiavone, 91 – 99. Torino: Einaudi.
- Hermon, E. 2001. *Habiter et partager les terres avant les Grecques*. Roma: Coll. École Française de Rome.
- Kantner, J. 2004. "Geographical approaches for reconstructing past human behavior from prehistoric roadways." In *Spatially Integrated Social Science*, edited by M. F. Goodchild and D.G. Janelle, 232-244. Oxford: Oxford University Press.
- Maggi, R., R. Nisbet, and G. Barker. 1991. *Archeologia della pastorizia nell'Europa meridionale (Atti della tavola rotonda internazionale, Chiavari, 22-24 settembre 1989, Museo Archeologico per la Preistoria e la Protostoria del Tigullio)*. Bordighera: Rivista di Studi Liguri.
- Mattioli, T. 2008. "Landscape Analysis of a Sample of Rock-Art Sites in Central Italy." In *Proceedings of the 35 International Conference on Computer Applications and Quantitative Methods in Archaeology*, edited by A. Posluschny, K. Lambers and I. Herzog, 342-343. Bonn: Kolloquien zur Vor- und Frühgeschichte.
- Murrieta-Flores, P. 2010. "Travelling in a prehistoric landscape: Exploring the influences that shaped human movement." In *Making History Interactive. Computer Applications and Quantitative Methods in Archaeology (CAA). Proceedings of the 37th International Conference, Williamsburg, Virginia, United States of America, March 22-26, 2009*, edited by B. Frischer, J. Webb Crawford and D. Koller, 258-276. Oxford: Archeopress.
- Narciso, E. 1991. *La cultura della transumanza (Atti dell'incontro di studio, Santa Croce del Sannio, 12-13 novembre 1988)*. Napoli: Guida editori.
- Pasquinucci, M. 1979. "La transumanza nell'Italia romana." In *Strutture agrarie e allevamento transumante nell'Italia romana (III – I sec. a.C.)*, edited by E. Gabba and M. Pasquinucci, 79 – 182. Pisa: Giardini editore.
- Spada, E. 2002. *Transumanza e allevamento stanziale nell'Umbria sud orientale*. Perugia: Quaderni del Cedrav.
- Tobler, W. 1993. "Three Presentations on Geographical Analysis and Modeling. National Center for Geographic Information and Analysis." Santa Barbara: University of California.
- Wheatley, D., and M. Gillings. 2001. *Spatial Technology and Archaeology: the Archaeological Applications of GIS*. London: Taylor & Francis.
- Whittaker, C.R. 1988. *Pastoral Economies in Classical Antiquity*. Cambridge: The Cambridge Philological Society.

Transparency, Testing and Standards for Archaeological Predictive Modelling

William Wilcox

University of East Anglia, United Kingdom

Abstract:

This paper starts with considering the extent of archaeological predictive modelling in Europe and the various different techniques used. One of the main criticisms against archaeological predictive modelling is that it is often viewed as a 'black box' technique and this paper suggests one possible way to make the procedure more transparent to the non-technical user and allow that user to test and interrogate a model. The paper stresses that, if possible, archaeological predictive models should be tested against new archaeological data, as opposed to how well a model predicts known archaeological data. The paper also suggests that the best way of justifying the use of the archaeological predictive modelling for cultural heritage management is by directly comparing the costs and results from the technique against the existing system of cultural heritage management used. This paper argues that whilst it would be impractical to write standards that cover every technique to produce a predictive model, it would be advantageous to start thinking now about standards for the output of these models. Thus, one model could be directly compared to another model and standardised attribute data could be exchanged between models and other applications.

Keywords:

Archaeological Predictive Modelling, Transparency, Testing, Standards

1. The Extent of Archaeological Predictive Modelling In Europe¹

In late 2011, the words 'archaeological predictive modelling' followed by the name of each of the countries of Europe were entered into the Google search engine. Thirty six countries (72%) had reference to research into archaeological predictive modelling within that country, fourteen countries (28%) had no reference to archaeological predictive modelling (Fig. 1) and twelve countries (24%) had reference to the technique being used (in part) for cultural heritage management. However, just because there was no reference to research into, or the use of, archaeological predictive modelling on the internet, does not mean that it does not exist in that country. Hence, the above figures are probably conservative. Internet references for research into archaeological predictive modelling were also found in Australia, the USA, Canada, parts of Africa, etc. The conclusion from this provisional survey is that there is a lot of interest in the technique world-wide and that some countries are starting to incorporate the technique into their systems of cultural heritage management.

Corresponding author: Archaeology53@gmail.com

2. Types of Modelling

From internet research and background reading, there appears to be as many archaeological predictive modelling techniques as there are archaeological predictive models. Examples include; models based on ethnographic and ethno-historic literature (Dalla Bona & Larcombe, 1996), adjusted deductive modelling (Altamira Consulting Ltd, 2009), least-cost path modelling (Verhagen, 2010), optimal foraging theory and diet-breadth modelling (Whitley, 2010), geographically weighted regression (Löwenborg, 2010), fuzzy logic (Hatzinikolaou, 2006), intelligent network structures, based on neural networks (Ducke 2003), K-means cluster analysis and Dempster-Shafer theory (Veljanovski & Stančič, 2006), binary logistic regression analysis (Cuming, 2002), and so on. The reason for this variety appears to be because each study area is unique. Some study areas have a large amount of known archaeological data and hence inductive (statistical) techniques can be applied and other areas have little archaeological data and hence deductive techniques have to be applied. Inductive techniques rely upon representative data, which is not always available and deductive techniques rely upon expert judgement, which can be biased.

Archaeologists are interested in where (and how) people lived or worked in the past and it is assumed that concentrations of artefacts and/or features are an indication of such areas, which are often called settlements or sites. Terrain can influence the power of a predictive model as some landscapes *funnel* human settlement or restrict the human exploitation of the landscape. For example, settlement patterns of hunter-gatherers are very different to settlement patterns of farmers as hunter-gatherers used temporary camps (based on access to foraging and hunting), whilst farmers lived in permanent houses (based on access to good arable or pastoral land). Further, the modelling of specific function sites would relatively straight forward as they tend to be located near to sources of raw materials. For example, pottery production would be based near a good source of clay, water and fuel; sheep farming would be based near to grass land that is not easily cultivated (possibly on steep slopes), etc. Hence, the type of site being modelled should be fully understood as the grouping of all site types together reduces the power of any archaeological predictive model. Due to these variants, it does not appear possible to directly compare one archaeological predictive model or technique with another model or technique, unless they are based in the exact same study area, the same archaeological and environmental data, for the same historical era and for the same type of site.

3. Criticisms Against the Use of Archaeological Predictive Modelling for Cultural Heritage Management

There are various criticisms against using the technique for cultural heritage management (for example see; Kamermans 2008, Wheatley 2004, etc.) and these criticisms must be addressed before the technique can fully accepted by the archaeological community. However, such criticisms should be considered in relation to existing systems of cultural heritage management as no system of cultural heritage management is, or can ever be, 100% perfect. For example, in England all planning applications for development (new housing, factories, roads, railways, etc.) are reviewed by the local archaeological department who, if considered necessary, will issue a written brief on what archaeological investigation is required before that development takes place.

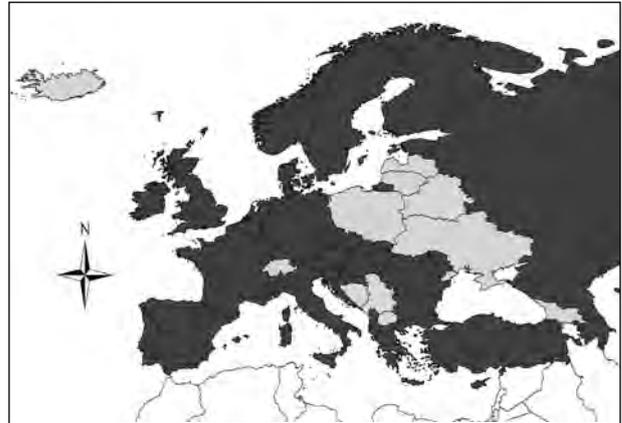


Figure 1. A provisional survey of countries in Europe that carryout research into archaeological predictive modelling or employ the technique (in part) for cultural heritage management. Dark grey = countries with research into, or use of, archaeological predictive modelling noted on the internet. Light grey = no reference to archaeological predictive modelling found on the internet.

Within four English counties in 2008, only 2% of all development applications involved some form of archaeological investigation and over a third of that investigation found no archaeological artefacts or features whatsoever. Therefore, it appears somewhat pedantic to argue that an archaeological predictive model in England should predict all archaeological data. Another criticism is that archaeological predictive models are self-fulfilling and that if areas of high prediction are only investigated, it is not possible to find archaeological data in low prediction areas, thus apparently strengthening the predictive power of the model. Whilst this is true, in several English counties the existing system of cultural heritage management is based on the proximity of known archaeological data. For example, within some English counties an archaeological investigation is required if a development is within a certain distance from a medieval church. Hence, this system of heritage management is also self-fulfilling. Therefore, to further the cause of using archaeological predictive modelling for cultural heritage management, we must directly compare the technique with existing systems of cultural heritage management. That is to say, would archaeological predictive modelling preserve more or less archaeological heritage than an existing system of cultural heritage management, for the same cost? In this respect, archaeological

predictive modellers appear to have a case to prove and further research into this issue is required.

Completed (and proposed) large building projects such as new road and rail schemes, housing estates, pipelines and areas of mineral extraction offer good opportunities to prove the worth of archaeological predictive modelling. First, determine the total cost of the archaeological work carried out, the time taken and, if some archaeological investigation was unproductive, how much? Second, carry out an archaeological predictive model, based on the known data *before* the archaeological investigation, and then determine (or not) how well the model agrees with the archaeological data found and how much unproductive archaeological investigation could have been saved should the archaeological predictive model have been used to determine where to carry out the archaeological investigation. Third, determine the potential cost saving to the project should an archaeological predictive model had been used.

Figure 2 shows a fictitious new housing estate where three evaluation trenches were demanded by the governing archaeological body, which were duly excavated, recorded and reported. Trench 02 uncovered no archaeological data! Figure 3 shows the same housing estate where an archaeological predictive model predicted that evaluation trench 02 would be archaeologically sterile. Should the cost of the archaeological predictive model be less than the excavation, recording and reporting of evaluation trench 02, the use of an archaeological

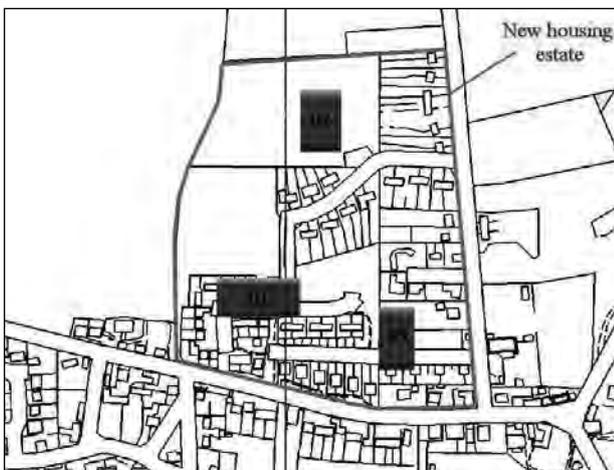


Figure 2. A fictitious new housing estate with no archaeological predictive model.

predictive model in this case would have saved the developer both time and money.

Have such comparisons ever been carried out? An archaeological predictive model was produced to determine where archaeological investigations should be carried out on a 7.0Km long proposed pipeline between Andijk-West and Wervershoof in the Netherlands. The model suggested that only 2.5Km of the pipeline was in an area of high potential. However, due to the unexpectedly large amount of archaeological data found in the high category, it was decided to extend the archaeological investigations to include the whole length of the pipeline, which resulted in further significant archaeological finds (Verhagen, 2007). Later analysis concluded that some of the original modelling assumptions were wrong and that ground disturbance played a greater part in archaeological survival than previously thought. Whatever the reason, the archaeological community could see this as a failure of archaeological predictive modelling and supporting evidence of why the technique should not be used for cultural heritage management.

Introduced in 1999 at a cost of \$4.5 million, the Minnesota archaeological predictive model is said to be saving the US state \$3.0 million per annum, compared to the original system of cultural heritage management (Madry et al. 2006). Unfortunately, when pressed for justification of these figures, it was revealed that these were ball-park estimates only, which had been reached by a consensus of opinion of the staff working on the project. Whilst there is

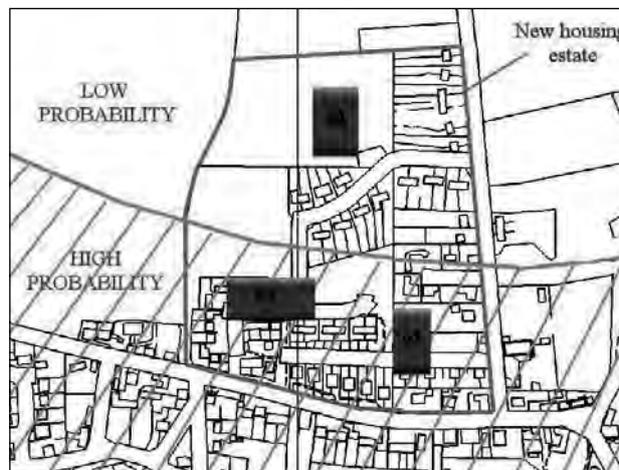


Figure 3. The same fictitious new housing estate with an archaeological predictive model.

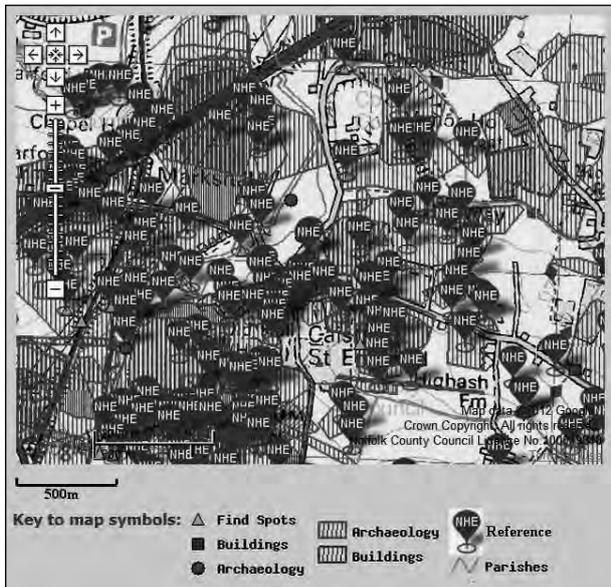


Figure 4. An area of dense archaeological data in Norfolk, England (Source: Norfolk Heritage website).

no doubt that the introduction of this archaeological predictive model saves the state money, it would be far more persuasive to the archaeological community if actual figures could be produced. Within the Weyerhaeuser Grande Prairie forest management area in Alberta, Canada, archaeological predictive models are used in advance of tree harvesting (see Altamira Consulting Ltd website). These models are said to have a success rate around 85% and the archaeological community (and contractors) see this as a good way of concentrating archaeological investigation, thus saving time and money. However, with the large Canadian wilderness containing little archaeological data, it is probably the only realistic technique available to cultural heritage managers.

So what about when there is a huge amount of archaeological data available within a study area (Fig. 4), is it necessary to produce an archaeological predictive model for cultural heritage management? Probably not, as there would be sufficient archaeological information with which to base sound archaeological decisions upon. In other words, an archaeological predictive model would probably not reveal much more than what is already known.

Whilst England contains large archaeological databases, there are some areas where the archaeological record is sparse and archaeological predictive modelling could be viable. For example,

between urban areas or areas of off-shore mineral extraction. Whilst there is no shortage of building material (sand, gravel and rock) in the UK, open quarries are often resisted by the local population. Hence, aggregate companies now obtain about 20% of aggregates by off-shore dredging in the North Sea and English Channel. However, when past sea levels were lower than today, the North Sea and English Channel were inhabited by man and modern aggregate dredging destroys any archaeological remains or features. The Regional Environmental Characterisation (REC) project (funded by the Marine Aggregate Levy Sustainability Fund), is an attempt (in part) to produce maps showing areas of potential underwater archaeological data. With large areas involved and little known archaeological data, archaeological predictive modelling is a possible candidate to help with this study - based on bathymetric data (old rivers and coastlines), geological data, etc.

4. Transparency

As previously discussed (see Wilcox, 2011), I believe that a significant portion of the archaeological community is suspicious of archaeological predictive modelling and view it as a *black box* technique. They do not fully understand the intricacies and limitations of the technique and as if by magic the technique produces beguiling coloured maps, which the modeller says are based on *sound mathematical principles* or *expert opinion* and simply asks the user to accept as being correct. It is like buying a second-hand car from a second-hand car salesman. Hence, every archaeological predictive model should be open to scrutiny by non-experts and hence should be accompanied by adequate data and a simplified explanation of the mathematical principles involved. To illustrate this point, Figure 5 is an inductive predictive model (based on binary logistic regression analysis) for Late Anglo-Saxon settlement in the county of Norfolk (UK), which is divided into above 50% probability (dark grey) and below 50% probability (light grey) of containing archaeological data.

Position A (above 50% probability) and position B (below 50% probability) can be chosen either randomly or specifically targeted. Hence, at position A the various values assigned to each environmental factor are: soil description (deep

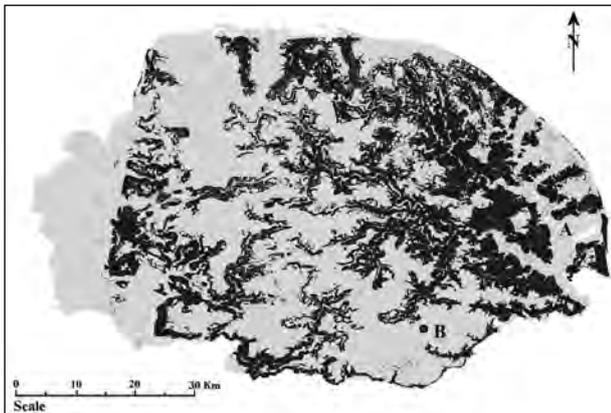


Figure 5. Archaeological predictive model of Late Anglo-Saxon settlement in Norfolk, England. Dark grey is above 50% probability and light grey is below 50% of finding archaeological data.

loamy soil) = 1; soil drainage (freely draining) = 1; contour level (below 32m above datum) = 1; ground slope (above 1°) = 1; distance to river (900m) = 0.9. Using the coefficients from the binary logistic regression analysis:

$$\text{Linear Predictor} = 0.235 + (-0.450 \times 1 [\text{ground slope}]) + (0.524 \times 1 [\text{contour level}]) + (-0.377 \times 0.9 [\text{river distance}]) + (-0.307 \times 1 [\text{soil description}]) + (0.394 \times 1 [\text{soil drainage}]) = 0.0567$$

$$\text{Predicted Probability} = 1/1+\exp(0.0567) = \underline{0.514}$$

Therefore, the probability of finding Late Anglo-Saxon archaeological data at point A is just above 50% (0.514 > 0.5). Thus, the model can easily be tested and interrogated by a non-expert. For example, the regression coefficient for distance from a river is -0.377, showing a negative preference for distance from a river and the coefficient for soil drainage is 0.394, which shows a positive preference for land with good drainage.

5. Testing

The main criticism against archaeological predictive modelling is that models are rarely tested. Defining an inductive model by how well it predicts its own input data is a circular argument and defining a deductive model by how well it predicts the known archaeological sites is also a circular argument. In both cases, the archaeological



Figure 6. The East Kent Access Road – 6.5Km long and 40 hectares of archaeological investigations (source: EastKent.owarch website).

data may not be representative as such data is often collected in different ways. For example, the spatial difference between known archaeological data in the counties of Norfolk and Cambridgeshire (UK) is large (Wilcox, 2011). Hence, the success of testing archaeological predictive models using data from these counties would vary considerably and would only high-light these spatial differences. The whole point of archaeological predictive modelling is to predict and models should be tested by how well they predict new data. Unfortunately, archaeological investigation to uncover new data is expensive and the whole point of archaeological predictive modelling is to avoid costly excavation. One way to resolve this *chicken and egg* situation is to use the technique in advance of large scale projects. New roads, pipelines and mineral extraction offer excellent opportunities to test predictive models, as archaeological investigations will be carried out regardless of any modelling. Further, it is these large projects that may have the finances to pay for the modelling process. Examples of such projects in England would be the proposed extension to Stansted Airport (32 hectares of archaeological investigation) and the new East Kent Access Road - A256/A299 (40 hectares of archaeological investigation) (Fig. 6).

Archaeological predictive models should also be interrogated upon completion. It is accepted that an archaeological predictive model will not predict every settlement but when a model does not predict certain types of settlement or settlements

in a particular region, questions should be asked as to why. Archaeological predictive modellers are beginning to split their study areas down into 'archaeo-regions' in order to account for the differences in terrain and so produce higher predictive gains. Such is the case for the Dutch national archaeological predictive model called the IKAW (Fig. 7). Whilst this is a perfectly acceptable technique, it does raise the question of how to merge these individual models to make one large model. The problem is that one model (say bounded by a river) and based on one set of variables may predict a high probability next to that river. Another model (on the other side of the river) and using different variables may predict a low probability next to that river!

In England, calculations by one Engineer are often checked by another Engineer for compliancy with the relevant standards and regulations. This system is regulated and controlled by the governing body for Engineers in the UK. Could such a system work with archaeological predictive models? One university or governing archaeological department produces an archaeological predictive model (to be used for cultural heritage management) that is then checked by another. I'm not suggesting a full mathematical check or that an independent model be produced to check the first model, but a *reasonability check*. For example, if the values for a predictive model state that the probability of discovering an archaeological settlement decreases the further away from a river, this could be viewed as a reasonable assumption, even if the checker may not fully agree with the actual values used. Such a system would require the establishment of a new governing body to control it. For example, it could be established within English Heritage or it could be incorporated with the Institute for Archaeologists (IFA). Could this be a simple way to induce confidence in the technique or would this just result in two sets of experts disagreeing?

6. Standards

With interest in archaeological predictive modelling growing and researchers using ever more inventive methods, it is proposed that some agreed simple standards at this stage could allow research to be integrated in the future and duplication avoided. However, given the ever increasing number

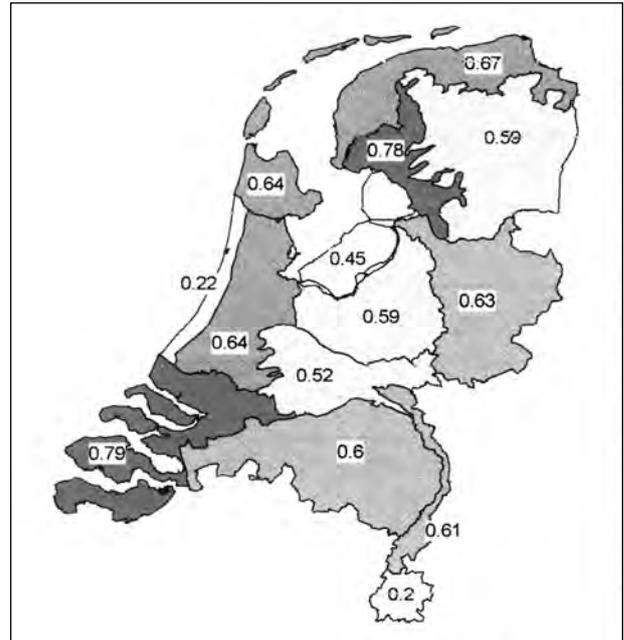


Figure 7. The various Kvamme's gains for different 'archaeo-regions' of the IKAW (source: Verhagen, 2007).

of different ways to produce an archaeological predictive model, writing standards for producing archaeological predictive models would not be an easy task. Hence, standards should concentrate on how models are presented and tested as opposed to how they are produced. The modeller would be free to model how he or she likes but if the model is to be checked by others, used for cultural heritage management or published, the model should comply with some standards. The only standard for archaeological predictive modelling that the writer is currently aware of is in British Columbia, Canada.

The following suggestions are offered:

- An agreed modelling resolution and projection should be agreed for each country.
- An agreed file format should be used for predictive models. Possibly ESRI files, as they are very common and compatible with most GIS programs, both commercial and Open Source.
- The Kvamme's gain of any archaeological predictive model should be quoted, so that models can be compared. Kvamme's Gain = $1 - (\text{percentage of study area} / \text{percentage of archaeological data})$.

- The archaeological site density within each probability category should be quoted, for risk assessment by cultural heritage managers. For example, High probability category = 1.23 archaeological sites per square kilometre, Medium probability category = 0.89 archaeological sites per square kilometre, etc.
- Specific data columns in the attribute tables should be agreed so that data can be used in other models or applications.

Metadata for each model should be given, that includes:

- The author, contact details, date and modelling background should be given.
- A detailed description of the study area, including geology, topology, etc.
- A detailed assessment of the archaeological and environmental data used for the model should be given, including any and all biases, restrictions, etc.
- A detailed description of the methods used to make the model should be given, including theoretical and mathematical principles.
- A detailed description of how the model was tested should be given.
- A detailed list of regression coefficients/expert derived weights for each variable, along with worked examples of using these to justify the category of prediction at any location should be given, in order that non-experts can interrogate and test the model.

7. Conclusions

It is possible that some members of the archaeological community see archaeological predictive modelling as the *Philosophers stone* of cultural resource management. Simply plot any development application against an archaeological predictive model and issue a standard letter stating exactly what (if any) archaeological investigation should be carried out. Could such a system replace experienced archaeologists? It can never be that

simple - archaeological predictive modelling is a tool, applicable for use in certain (but not all) circumstances and the final decision regarding cultural heritage management should always be made by a human. It is only human to make mistakes, but you need a computer to really mess things up! Hence, we must make our models accessible for scrutiny by non-technical users by providing sufficient data and worked examples of how to interrogate and test a model at any point.

We must strive to answer all the criticisms raised against the technique, admitting that some aspects can never be fully resolved, but comparing the technique against the existing system of cultural heritage management. Cultural heritage management is about managing risk on a given budget. Does the existing system of cultural heritage management offer the best protection for undiscovered archaeological remains/artefacts or would archaeological predictive modelling do a better job for the same budget - in which case this needs to be proved!

Whilst we should support new ideas about modelling the past, it is not ever more inventive modelling techniques that we need right now. It is (more) hard proof that archaeological predictive models actually predicts new archaeological data to a high degree and that (in certain circumstances), it can do a better and cheaper job than the existing system of cultural heritage management.

References

- Altamira Consulting Ltd. Accessed 2009. http://www.archaeology.ca/featured_projects/predictivemodel.html.
- Cuming, P. 2002. "An Assessment of the SMR as a Predictive Tool for Cultural Resource Management and Academic Research." In *Contemporary Themes in Archaeological Computing, University of Southampton Department of Archaeology Monograph 3*, edited by D. Wheatley, G. Earl, and S. Poppy, 28-36. Oxford: Oxbow Books.
- Dalla Bona, L., and L. Lacombe. 1996. "Modelling Prehistoric Land use in Northern Ontario." In *New Methods, Old Problems: Geographical Information Systems in Modern Archaeological Research. Centre of Archaeological Investigations, Occasional Paper 23*,

- edited by H. Maschner, 252–71. Carondale: Southern Illinois University Press.
- Ducke, B. 2003. “Archaeological Predictive Modelling in Intelligent Network Structure.” In *The Proceedings of the 29th Conference of the Computer Applications in Archaeology*, edited by M. Doerr, and A. Sarris, 267–73. Crete: Hellenic Ministry of Culture.
- Hatzinikolaou, E. 2006. “Quantitative methods in archaeological prediction: from binary to fuzzy logic.” In *GIS and Archaeological Site Location Modelling*, edited by M. Mehrer and K. Wescott, 437–46. London: Taylor and Francis.
- Kamermans, H. 2008. “Smashing the crystal ball: a critical evaluation of the Dutch national archaeological predictive model (IKAW).” *International Journal of Humanities and Arts Computing* 1: 71–84.
- Löwenborg, Daniel. 2010. *Excavating the Digital Landscape: GIS analysis of social relations in central Sweden in the first millennium*. Uppsala: Uppsala Universitet.
- Madry, S., M. Cole, S. Gould, B. Resnick, S. Seibel, and M. Wilkerson. 2006. “A GIS based Archaeological Predictive Model and Decision Support System for the North Carolina Department of Transport.” In *GIS and Archaeological Site Location Modelling*, edited by M. Mehrer and K. Wescott, 317–34. London: Taylor and Francis.
- Veljanovski, T., and Z. Stančič. “Predictive Modelling in Archaeological Location Analysis and Archaeological Resource Management: Principles and Applications.” In *GIS and Archaeological Site Location Modelling*, edited by M. Mehrer and K. Wescott, 393–412. London: Taylor and Francis.
- Verhagen, P. 2007. *Case Studies in Archaeological Predictive Modelling*. Leiden: Leiden University Press.
- Verhagen, P. 2010. “On the road to nowhere? Least-cost paths and the predictive modelling perspective.” In *Fusion of Cultures; Abstracts of the 38th Annual Conference on Computer Applications and Quantitative Methods in Archaeology*, edited by F. Melero, P. Cano and J. Revelles, 439 – 442. Spain: Granada.
- Wheatley, D. 2004. “Making Space for an Archaeology of Place.” *Internet Archaeology* 15. <http://eprints.soton.ac.uk/28800>
- Whitley, T. 2010. “Potentially Prediction and Perception: Using caloric landscapes to reconstruct cognitive patterns of subsistence and social behaviour’.” In *Fusion of Cultures; Abstracts of the 38th Annual Conference on Computer Applications and Quantitative Methods in Archaeology*, edited by F. Melero, P. Cano and J. Revelles 214–244. Spain: Granada.
- Wilcox, W. 2012. “Archaeological Predictive Modelling used for Cultural Heritage Management’ in *Revive the Past*.” *Proceedings of the 39th Annual Conference on Computer Applications and Quantitative Methods in Archaeology*, edited by M. Zhou, I. Romanowska, Z. Wu, P. Xu and P. Verhagen, 353 - 358. Amsterdam: Amsterdam University Press.

Zooming Patterns Among the Scales: a Statistics Technique to Detect Spatial Patterns Among Settlements

Alessio Palmisano

Univeristy College London, United Kingdom

Abstract:

The present paper aims to offer an overview of the existing statistical approaches to settlement patterning in Archaeology by one multi-scalar method (Ripley's K function) and to show both problems and potentiality of such technique dealing with spatial data. I will use as case study the Iron Age I period archaeological sites located in the actual West Bank (or Cisjordan) in order to show how point pattern analysis can help us to detect spatial patterns and investigate if phenomena of attraction or repulsion among settlements are mainly related to the first or second order of effects. With the term "first order of effects" I mean that the observations throughout a study area "vary from place to place due to the changes in the underlying properties of the local environment" (O'Sullivan and Unwin, 2003, 79). The second order of effects, instead, is due to the local interaction between the observations. Therefore, I am going to investigate through this particular method how the environmental variables (first order of effects) and the direct interaction between the settlements themselves (second order of effects) can determine different spatial patterns (cluster, even or random distribution of settlements).

Keywords:

Spatial Analysis, Iron Age I Period, Southern Levant, Settlement Pattern, Landscape Archaeology, West Bank (Cisjordan)

1. Introduction

During the last years digital technologies have been used in Archaeology for the documentation, the management and the representation of archaeological data. A consequence of this phenomenon is the increasing popularity of Geographical Information Systems (GIS) as powerful tool for the organization and the visualization of archaeological data in relation with the correspondent spatial information, while less attention is paid to the application of spatial statistics for detecting specific patterns of such datasets.

Recently, however, GIS has been used by archaeologists not only for data management, but also for analysing data and their spatial references. Thus, researchers can detect particular patterns from archaeological data collected through archaeological surveys or excavations by carrying out statistical spatial analyses. It is important to point out that statistics and spatial analysis have the power to

detect and explain even the most vague and complex aspects of the world (Shennan 1997, 3). Researchers, by using statistics, can discern the presence or the absence of any pattern in the archaeological data, and detect the relationships between the spatial and the attribute datasets (Conolly and Lake 2006, 122). Many foundations of spatial analyses were established by geographers in the 1950's and 1960's, and then adopted and modified by archaeologists in the 1970's and 1980's (Hodder and Orton 1976). In the past 20 years there have been several important advances and a renewed interest in the application of spatial statistics techniques to the study of the past human behaviour (Williams 1993; Beardah 1999; Premo 2004; Crema et al. 2010; Ladefoged and Pearson 2010; Bevan 2012).

This paper focuses on a set of point pattern and process models that now put archaeologists in a position to return to the analysis of spatial pattern and process with renewed ambition, especially with regard to distribution maps. Therefore, the first section offers an overview of the existing statistical approaches to settlement patterning in Archaeology

Corresponding author: a.palmisano@ucl.ac.uk

by one multi-scalar method (Ripley's K function) and shows both problems and potentiality of such technique dealing with spatial data. The subsequent section focuses on one particular case study (the Iron Age I settlements in the actual West Bank or Cisjordan²), which highlights some important conceptual issues and new analytical opportunities.

2. Spatial Statistics and Settlement Patterns

Archaeologists frequently use points for representing the location of artefacts, sites, particular features, etc. The analysis of point distribution can be an important tool for understanding, interpreting and explaining the spatial configuration of our observations in a given study area. There are three idealised point distributions: random, clustered and regular. Nevertheless, these states rarely occur so clearly in reality and settlement patterns are more complex above all if we consider that the scale of analysis can change the spatial configuration of a spatial pattern. A regular distribution of sites can reflect a form of competition among settlements, the existence of agricultural catchments or both (Hodder and Orton 1976, 54-85). Clustering of sites may be due to a localized distribution of sources or to the presence of polities and regional centres. On the other hand, random distributions have been usually treated as null-hypotheses, although this kind of spatial pattern can be determined by specific environmental and social variables (Bevan and Conolly 2006, 218). The most known technique in Archaeology for detecting spatial patterns among settlements is nearest neighbour analysis, which was originally designed by Clark and Evans (1954) for ecological purposes. The widespread of this technique among archaeologists is due to the fact that it is simple to calculate and it provides an easy coefficient to interpret. Nevertheless, this technique considers only the first nearest neighbour, it is not multi-scalar and its result is strongly affected by the size of the analysed area.

At this point, I will propose a multi-scalar spatial statistics technique known as Ripley's

² The West Bank is a landlocked territory bordering Jordan to the east and Israel to the west, north and south. Since the Oslo peace's accords in the 1993, parts of the West Bank are under full or partial control of the Palestinian Authority. Actually, 164 nations refer to the West Bank, including East Jerusalem, as "Occupied Palestinian Territory".

K function, designed to detect spatial patterns of aggregation or segregation of point data at different scales (Ripley 1976). In this section I will try to identify the possible interactions between settlements distributed in a given study area by carrying out point pattern analyses to investigate if phenomena of attraction or repulsion are mainly related to the first or second order of effects. With the term "first order of effects" I mean that the observations throughout a study area "vary from place to place due to the changes in the underlying properties of the local environment" (O'Sullivan and Unwin 2003, 79). The second order of effects, instead, is due to the local interaction between the observations. Therefore, in the present paper, I am going to investigate how the environmental variables (first order of effects) and the direct interaction between the settlements themselves (second order of effects) can determine different spatial patterns (cluster, even or random distribution of settlements). In practice, it is important to point out that it is not easy to distinguish the first/second order effects in the point pattern simply by observing as the intensity of a process varies across space. In fact, any points distribution documented in a given study area can be thought as the result of one or more underlying processes. Moreover, in many real world examples, it is probably that different spatial patterns (cluster, even or random distribution of points) are due to multiple processes behaving differently in different parts of a specific study area. Therefore, is it important to make a distinction between the first order effects as factors and phenomena affecting the intensity of points across a region, and the second order effects that describe different patterns of interaction among points such as attraction and segregation.

In this section I will use two different point pattern analyses for investigating if the spatial patterns of the Iron Age I settlements are due to the direct interaction among them or to the environmental variables: homogenous and inhomogeneous Ripley' K functions. I will also investigate two different kinds of homogenous Ripley's K functions: global and local.

The homogeneous Global Ripley's K function describes the second order properties of a point pattern by using measures based on distances between all events in the study area (Lloyd 2007).

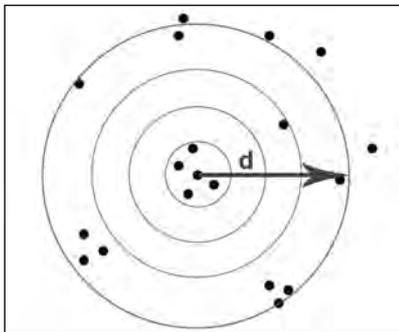


Figure 1. Number of events within the circle with radius d centred on location x .

The K function for distance d is given by the following formula:

$$K(d) = \frac{\#(C(x,d))}{\lambda}$$

where $\#(C(x,d))$ indicates the number of events in the circle $C(x,d)$, with radius d centred on location x (Fig. 1) and λ is the average intensity ($M = N(a)/A$) of the process. This global function examines spatial dependence over small spatial scales by assuming homogeneity and isotropy over the scale of analysis (stationary point process) (Lloyd 2007, 177). The K function is obtained by counting the number of point within radius d of an event and calculating the mean count for all events; then, the mean count is divided by the overall study area event density. We can repeat the same procedure at different scales by modifying the values of the radius d around each event.

The local K function is similar to the global K function, but only pairs of points that have a given point i as one of the members of the pair are included (Lloyd 2007, 186-187). A local K function of distance d is defined by the following formula:

$$K(d) = \frac{|A|}{n} \frac{N(C(x_i,d))}{\lambda}$$

where counts are of all points within distance d of point i and $|A|$ indicates the area of the window analysis. Unlike the Global Ryley's K function, the local K function allow us to show the spatial distribution of clustering at each spatial scale by plotting the K value for all artefacts at each bandwidth in turn.

Nevertheless, the homogeneous Ripley's K function has some weaknesses: it supposes homogeneous space and is characterized by a stationary point process, where the intensity

function λ^3 is constant throughout the whole study area (Marcon and Pluech 2003, 2). If an initial analysis suggests that a pattern is not characterized by a constant intensity over the study area, the homogenous K function does not allow further analyses (Baddeley et al. 2000, 330).

Inhomogeneous point pattern could arise if, for instance, we detect that our observations are distributed according to environmental variables such as soil fertility, water source proximity and so on. Therefore, the inhomogeneous K function is a non-parametric second order analysis, which allows us to analyse point patterns distributed in a heterogenic space. This is possible by creating a predictive surface including the environment variables as covariates (Baddeley and Turner 2005, 23). Therefore, an inhomogeneous K function will not only depend on the distribution of points in a study area, but also on the underlying intensity function, so that to de-trend the first order of effects (Comas et al. 2008, 390).

3. Study Area and Dataset

My study focuses on past human settlement in Southern Levant, particularly in the West Bank highlands (Cisjordan). This area (coordinates: 32°00'N 35°15'E) has a total extension of 5,860 km² and an average altitude of 600 m. The highest and best known peak is Tall Asur (1,016 m), a mount located 10 km North-East of Ramallah, while the lowest point is located in the Dead Sea (- 408 m).

The peak settlement that took place in the West Bank highlands, during the Iron Age I period (12th-11th centuries B.C.), is one of the most controversial and discussed topics in the history of the land of Israel. Many scholars have studied this stormy period with different points of view: the biblical interpretation (Albright 1939; Dever 1992; Thompson 1992), historical research and archaeology (Finkelstein 1996; Finkelstein 1998; Finkelstein and Mazar 2007). Most of them assumed, in their researches, the presence of a unique ethnical group living in this region during Iron Age I period: the proto-Israelites. Instead, in this paper I will generally speak about Iron Age I

³ The intensity is given by the following formula $\lambda=N/A$, where N indicates the number of events and A is the overall area of the window analysis.

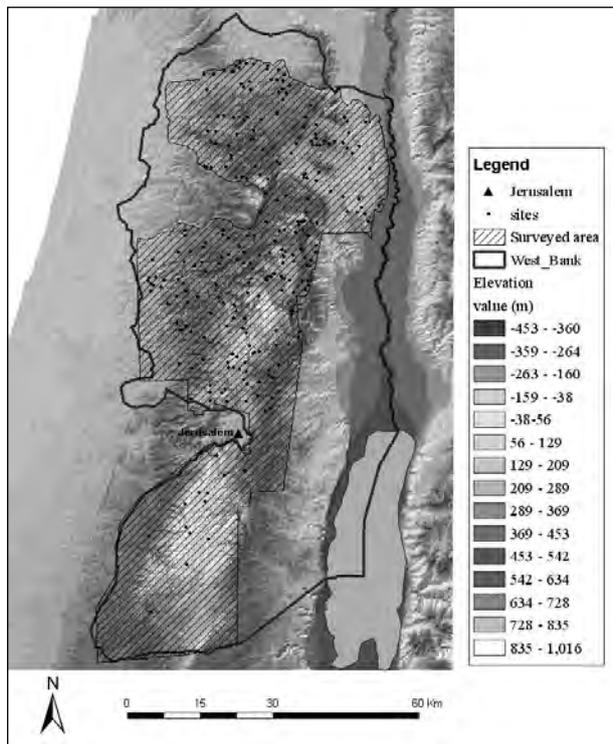


Figure 2. Map showing the West Bank boundaries (in black) and the total surveyed area.

settlers without assuming any ethnic boundary but assessing how in this region the environmental features, related to cultural-socio-economic aspects, may have affected the human decision-making process and the settlement distribution of Iron Age I people.

The principal archaeological surveys throughout the West Bank have been carried out in the north by A. Zertal between 1978 and 1988 (Zertal 2004) and by Finkelstein, Lederman and Magen between 1980 and 1988 (Finkelstein and Lederman 1997; Finkelstein and Magen 1993); in the south by M. Kochavi in 1967-68 and A. Ofer in early 90's (Kochavi 1972).

The methodology of the surveys consisted of making a grid of a selected region and then spreading out in a line and walking throughout the whole area looking for all traces of ancient remains. The archaeologists carefully recorded any natural (topography, distance from water sources, soil typology, elevation) and archaeological feature (wall lines, buildings remains, potsherds, etc.) in order to detect the nature of ancient occupation. After

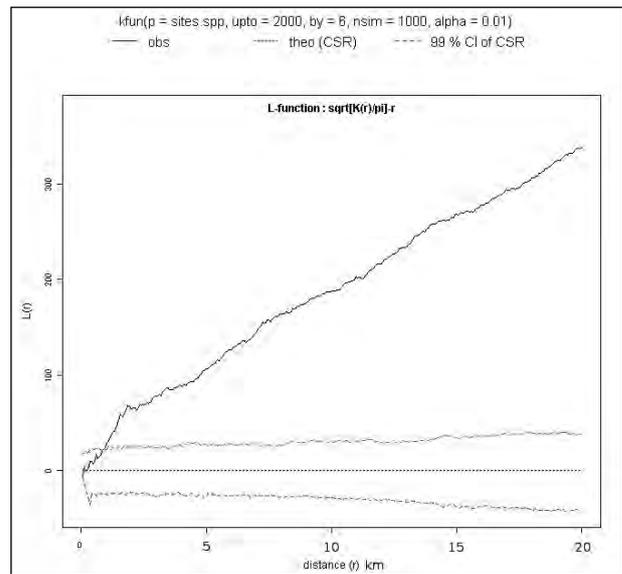


Figure 3. Global homogeneous Ripley's K function for Iron Age I settlements.

surveying all sites, maps of each archaeological period were drawn to show the location and distribution of sites recorded by size and typology. I managed to identify and locate 270 sites distributed throughout the whole area in the Iron Age I by gathering the data coming from the archaeological surveys that I have above mentioned (Fig. 2).

4. Analyses and Results

I carried out an univariate global Ripley's K function for all 270 settlements by using an interval of 60 metres (the minimum nearest neighbour distance between the settlements) and a maximum bandwidth of 20 Km.

The resultant K function plot shows the cumulative frequency distribution of average point intensity at set increments (equal to 60 m) of r. I used Monte Carlo simulations of points random distributions to estimate local confidence limits of the null hypothesis of complete spatial randomness (CSR) and obtained 99 percent confidence interval by carrying out 1000 iterations (Bevan and Conolly 2006, 220). These estimates are, then, compared to the observed values of K (L) in order to provide a statistical robust measure of a clustered or regular point distribution in our study area (Conolly and Lake 2006, 166). The Ripley's K plot (Fig.3) shows a strong positive deviation from the confidence

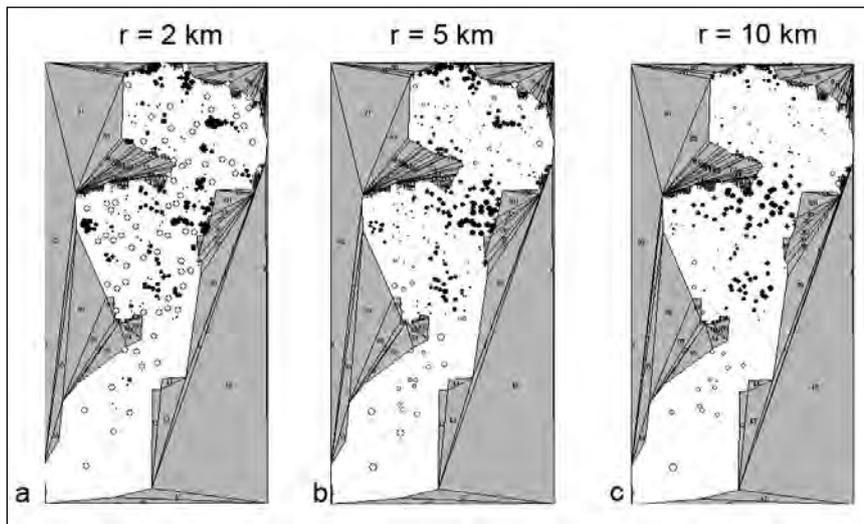


Figure 4. Local Ripley's K values for all settlements at bandwidth 2, 5 & 10 Km.

interval and indicates that from a distance (r) greater than 1 Km sites cluster into statistically significant groups ($p < 0.01$) indicating that there are more neighbours than expected at almost all distances r .

Unlike the univariate Global Ripley's K function, the local K function allow us to show the spatial distribution of clustering at each spatial scale by plotting the modified K-distribution (L) function value for all settlements types at each bandwidth in turn. At this point, I created different plots of all 270 settlements at bandwidth 2, 5 and 10 km in order to detect different local patterns. I would like to point out that there is not an ideal bandwidth's value because this kind of spatial statistics technique is multi-scalar and its results depend on the scale adopted. So, in this case I have used three different bandwidths for showing the common general trends occurring in my study area at different scales.

In my plots the dimension of black circles represents the degree of aggregation of a site, while the dimension of white circles represents the degree of segregation of each point. So, the greater is a black circle the more clustered are the corresponding sites. If we look at the local clusters of all sites at bandwidth equal to 2km (Fig. 4a) we can see that some small clusters are distributed mainly on the eastern and north-western ridges of the West Bank. Instead, at bandwidth 5 (Fig. 4b) and 10 Km (Fig. 4c), the clusters appear greater than the cluster obtained by using a bandwidth equal to 2 km. That means that the results change according our scale

and that there is not an ideal bandwidth to use for detecting the spatial distribution of clusters in our study area. Thus, we can compare the different results obtained by making use of different bandwidths either for detecting general trends occurring in any scale or for inquiring particular pattern at a given scale. The results of local Ripley's K function, by looking at the three outputs obtained with three different bandwidths, tell us that the settlements were generally clustered in the Central West Bank and in the northern edge of the study

area. The inhomogeneous K function, as we have previously said, allows us to detect the interaction between the sites by de-trending the first order of effects (environmental factors) through the creation of an-inconstant intensity surface underlying all our study area.

We can therefore build some formal point process models to consider what environmental affordances could have significantly affected the distribution of settlements in the West Bank highlands. We begin by considering, as examples, six related environmental factors: elevation, slope, aspect, ridge-top landforms, distance into/out of the dolomite geological typology and topographic wetness in a local catchment (Figures 5a-f)⁴.

This selection is prompted in part by many commentators' informal impressions that rugged topography and hydrology were important factors

4 The digital elevation model (DEM) used here is NASA's 30m ASTER dataset. Ridge-like landforms were defined from the DEM via a fuzzy feature classification across focal filter scales from 3x3 to 11x11 cells (Fisher et al. 2004). Catchment-based topographic wetness was calculated via focal filtering of a standard topographic wetness index surface (itself derived from the DEM) in a way that summed all values within a circular neighbourhood of 1.5 km radius (the average of the nearest neighbour distance of the Iron Age I sites). The slope and aspect maps were derived from the DEM. After carrying out a chi-squared test (p -value : 0.02813) I assessed that there was a spatial correlation between the settlements distribution and the geology type, and more than expected sites were located in dolomite geological category. Therefore, I created a raster surface indicating the distance of each pixel into and out of dolomite.

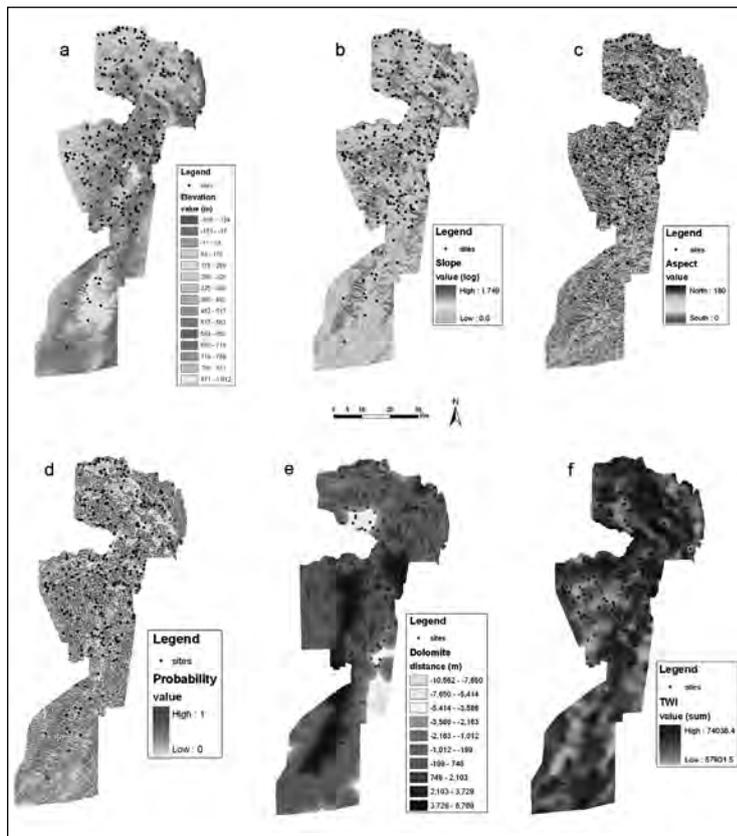


Figure 5. Iron Age I sites in the West Bank and six possible first-order covariates: (a) elevation (light to dark ranges from - 306 – 1012 m ASL), (b) slope map, (c) aspect map, (d) ridge landforms (darker is more likely to be geomorphometrically classified as a ridge), (e) Map of the study area showing the distance into dolomite (positive values) and the distance out of the dolomite (negative values), (f) topographic wetness index summed over a local neighbourhood (darker is wetter).

behind settlement locations in this region and period, for a variety of practical reasons (e.g. Zertal 1988; Gibson 2001).

If we now run a multivariate regression and select the best combination of these six variables via stepwise comparison (minimising an Akaike Information Criterion), we find that that elevation, distance into/out of the dolomite and the aspect are not particularly good predictor of the intensity of sites across the landscape, that the other three variables are all significant ($p < 0.05$ or better), and that this new model with a first-order trend is substantially more effective than a null, random hypothesis (Fig. 6). We can then create a predicted first-order intensity surface (Fig. 7) by implementing covariates that are good predictors: slope, topographic wetness index

and ridge-like landforms. I performed Monte Carlo simulations of random points distribution to estimate local confidence limits of the null hypothesis of complete spatial randomness (CSR) and I obtained 95 per cent confidence interval by carrying out 500 iterations. Unlike the homogeneous Ripley's K function, the random points created at each iteration are distributed according to the highest values of the intensity surface.

If we look at the inhomogeneous K function plot of the 270 settlements (Fig. 8), we can see that the sites are randomly spaced at small distances r (lower than 1 km) and they tend to be clustered at a distance approximately greater than 1 km. These results reflect the same pattern that we have when we perform a homogeneous Ripley's K function and they tell us that the environmental variables (first order of effects) implemented in our model could have not strongly affected the spatial distribution among the Iron Age I settlements in the West Bank highlands. So, the phenomena of aggregation or repulsion among the sites could be due either to socio-cultural factors or to other environmental factors not included in the model.

The analyses' results could be framed into a long term study including a time span from the Chalcolithic period (fourth millennium BC) to the Iron Age II (first millennium BC). It would be worth investigating the cyclic mechanism of alternating processes of sedentarization and nomadization of indigenous groups in response to the changing political, economic, and social circumstances. This could be done through the study of the faunal assemblage of the sites, where pastoral communities will yield more sheep/goats remains and sedentary society more cattle remains. In fact, there were three peaks of settlement with two intervals of collapse in the West Bank in the third and second millennium BC. The peaks of settlement occurred in the Early Bronze Age I (3300-3000 BC), in the Middle Bronze Age (ca. 1750-1550 BC) and in the Iron Age I (ca. 1200-1000 BC), while the demographic collapses occurred in the Intermediate

variables	Pr (> z)	p-value
Slope	1.21e-09	1.936202e-10
Elevation	0.267	0.2654711
Wetness index	0.0145	0.014
Distance from dolomite	0.367	0.3651802
Ridges	2.73e-14	2.109424e-15
Aspect	0.527	0.5270893

Figure 6. Top. $P(r)$ and p -values from multivariate logistic regression of the settlement distribution.

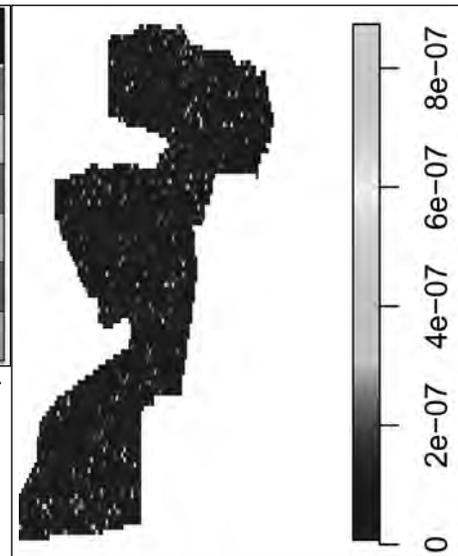


Figure 7. Right. Prediction surface based on the three significant covariates (lighter is higher point intensity).

Middle Bronze Age (ca. 2200-2000 BC) and in the Late Bronze Age (ca. 1550-1200 BC). Therefore, in a future research it would be worth investigating and comparing the different settlement patterns occurred in the area across the millennia in order to have a better understanding of the phenomena causing this cyclic settlement process.

5. Conclusions

In this paper I have analysed how the first and the second order of effects may have affected the distribution of Iron Age I settlements in the West Bank highlands. The results of the global and local univariate homogenous Ripley's K function indicate that the settlements are clustered at almost all scales. Nevertheless, the results of these analyses have to be analysed carefully because the homogeneous K function assumes the stationarity of the point process and that the space is homogeneous. These assumptions cannot be valid for the West Bank highlands, a large study area characterized by environmental heterogeneity, geological and topographical properties that can determine inhomogeneous structure. This does not mean that it is useless investigating settlement patterns regarding the second order of effects. In addition, the local Ripley's K has revealed that the settlements were mainly located in the central West Bank and along the northern edges of the study area. This could reflect spatial patterns also connected with different environmental conditions and subsistence strategies.

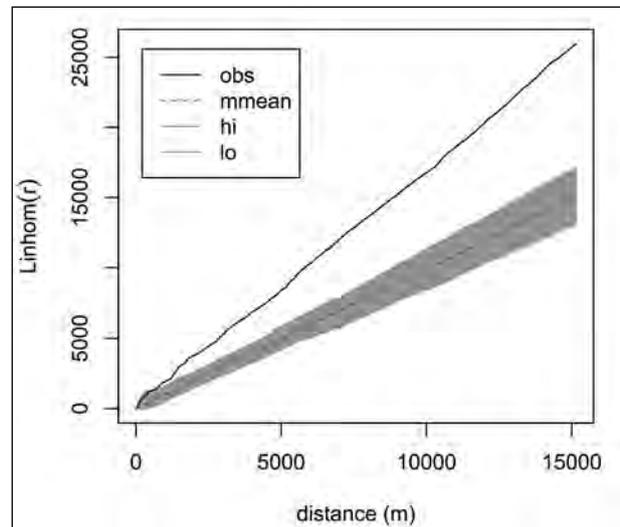


Figure 8. Global inhomogeneous Ripley's K function for Iron Age I settlements.

As I have formerly pointed out, the homogenous K function shows some weaknesses and it is not completely reliable for our purposes. For this reason I carried out the inhomogeneous Ripley's K function analysis by de-trending the first order of effects. The results of this analysis indicate that at a distance greater than 1km the Iron Age I settlements are significantly clustered, by indicating that in this particular study area the environmental factors (first order of effect) could have not affected the spatial pattern among the settlements. Therefore, the homogeneous and inhomogeneous Ripley's K analyses should be used complementarily in order to detect both the environmental and the cultural

factors that could have affected the settlements patterns in the West Bank highlands during the Iron Age I period.

It is interesting to notice that about 70 % of the settlements were located on the ridges, suggesting that this environmental feature was the privileged location where to settle. The results of the analyses make us state that the settlements were located on the top hills for a series of reasons such as the building of agricultural terraces, and the visibility. Most Iron Age I settlements were small and medium unfortified agricultural villages specialized in a horticulture-oriented economy including the industrialization of orchard's products. Considering this aspect, the agricultural terraces were the most striking man made features dominating the highland landscape of the West Bank. The terraces were built to guarantee additional ground space dedicated to the cultivation of fruit trees and other crops. The terraces were located on the slope of hills, on sloping plateau areas and over gullies and valleys. In this context the proximity of water sources and land suitable for cultivation played an important role. In fact, the sites were mainly located on the ridges bounding green and fertile intermountain valleys, where the agricultural productivity was higher. The northern and central West Bank, with vast fertile intermountain valleys and a relatively moderate topography were suitable zones for the proliferation of small sedentary rural communities.

Therefore, some archaeological features such as the plastered cisterns, the pillared room houses and the agricultural terraces mirror the socio-economic conditions of Iron Age I settlers to adapt and exploit the hilly landscape and not the cause that favoured the peak of settlements over the region.

The eastern flank of the west Bank, instead, characterized by the presence of steppes and the proximity to the desert fringes, was more suitable for animal husbandry and dry farming (in flat and open areas) undertaken by nomadic or semi-nomadic communities.

Another important aspect is that I have identified several clusters distributed on the western and eastern edges of my study area and in some intermountain valleys in the central highlands. These clusters are mostly composed by settlements

not located on the ridges but mainly distributed over intermountain valleys (central and western West Bank) and steppes (eastern West Bank). This could suggest that the position of settlements on the ridges may reflect the need of segregation of the "highlanders", while the settlements located in the valleys and in the steppes tended to be more clustered.

The vacuum of power following the Sea People advent and the end of Egyptian domination, may have turned the relationships of various groups within highlands into a form of competition among villages resultant in the raise of "totemic" and "local" identities (Faust 2006, 229). This determined the proliferation of many rural and small communities spread over large areas and divided into several local "factions".

In the end, I conclude by stating that perhaps the "proto-Israelites" were not a unique ethnical group living in the West Bank highlands, but one of the several communities that inhabited the region in the Iron Age I.

References

- Albright, W. F. 1939. "The Israelite conquest of Canaan in the light of Archaeology." *Basor* 74.
- Baddeley, A., J. Moller and R. Waagepetersen. 2000. "Non-and semi-parametric estimation of interaction in inhomogeneous point patterns." *Statistica Neerlandica* 54: 329-350.
- Beardah, C. 1999. "Uses of multivariate kernel density estimates." In *Archaeology in the age of the Internet: Computer Applications and Quantitative methods in Archaeology 1997*, edited by L. Dingwall, S. Exon, V. Gaffney, S. Laflin and M. van Leusen. Oxford: Archaeopress.
- Bevan, A. 2012. "Spatial methods for analysing large-scale artefact inventories." *Antiquity* 86(332): 492-506.
- Bevan, A., and J. Conolly. 2006. "Multiscalar Approaches to Settlement Pattern Analysis." In *Confronting Scale in Archaeology*, edited by G. Lock, and B. L. Molyneaux, 217-234. New York: Springer.
- Clark, P. J., and F. C. Evans. 1954. "Distance to Nearest

- Neighbour as a Measure of Spatial Relationships in Populations." *Ecology* 35: 444-453.
- Crema, E., A. Bevan, M. Lake. 2010. "A probabilistic framework for assessing spatio-temporal point patterns in the archaeological record." *Journal of Archaeological Science* 37 (5): 1118-1130.
- Comas C., M. Palahí T. Pukkala, and J. Mateu. 2008. "Characterising forest spatial structure through inhomogeneous second order characteristics." *Stochastic Environmental Research and Risk Assessment* 23 (3): 387-397.
- Conolly, J., and M. Lake. 2006. *Geographical Information Systems in Archaeology*. Cambridge: Cambridge University Press.
- Dever, W. G. 1992. "How to tell a Canaanite from an Israelite." In *The Rise of Ancient Israel* edited by H. Shanks. Washington: Biblical Archaeology Society.
- Faust, A. 2006. *Israel's Ethnogenesis: Settlement, Interaction, Expansion and Resistance*. London: Equinox.
- Finkelstein, I. 1996. "Ethnicity and Origin of the Iron I Settlers in the Highlands of Canaan: Can the Real Israel Stand Up?" *The Biblical Archaeologist* 59 (4): 198-212.
- Finkelstein, I. 1998. "The Rise of Early Israel Archaeology and Long-Term History." In *The Origin of Early Israel, current debate*, edited by A. Shmuel and E. D. Oren. Jerusalem: Posner & Sons Ltd.
- Finkelstein, I. and Z. Lederman. 1997. *Highlands of many cultures. The Southern Samaria Survey*. Jerusalem: Graphit Press Ltd.
- Finkelstein, I. and Y. Magen. 1993. *Archaeological Survey of the Hill Country of Benjamin*. Jerusalem.
- Finkelstein, I. and A. Mazar. 2007. *The quest for the historical Israel*. Atlanta: Society of Biblical Literature.
- Gibson, S. 2001. "Agricultural Terraces and Settlement Expansion in the Highlands of Early Iron Age Palestine: Is There Any Correlation Between The Two?" In *Studies in the Archaeology of the Iron Age in Israel and Jordan*, edited by A. Mazar. Sheffield: Sheffield Academic Press.
- Hodder, I., and C. Orton. 1976. *Spatial Analysis in Archaeology*. Cambridge: Cambridge University Press.
- Kochavi, M. 1972. *Juda, Samaria and Golan. Archaeological Survey 1967-68*. Jerusalem.
- Ladefoged, T. N. and R. Pearson. 2000. "Fortified castles on Okinawa Island during the Gusuku Period, AD 1200-1600." *Antiquity* 74: 404-412.
- Lloyd, C. D. 2007. *Local Models for Spatial Analysis*. London: Taylor & Francis Group.
- Marcon, E. and F. Puech. 2003. *Generalizing Ripley's K function to Inhomogeneous Populations*.
- O'Sullivan, D. and D. J. Unwin. 2003. *Geographic Information Analysis*. Hoboken: John Wiley & Sons.
- Premo, L. 2004. "Local spatial autocorrelation statistics quantify multi-scale patterns in distributional data: an example from the Maya Lowlands." *Journal of Archaeological Science* 31: 855-866.
- Ripley, B. D. 1976. "The Second-Order Analysis of Stationary Point Process." *Journal of Applied Probability* 13: 255-266.
- Shennan, S. 1997. *Quantifying Archaeology*. Edinburgh: Edinburgh University Press.
- Thompson, T. L. 1992. *Early History of the Israelite People From the Written and Archaeological Sources*. Leiden: Brill.
- Williams, J. T. 1993. "Spatial autocorrelation and the Classic Maya collapse: one technique, one conclusion." *Journal of Archaeological Science* 20: 705-709.
- Zertal, A. 1988. "The water factor during the Israelite Settlement Process in Canaan". In *Society and Economy in the Eastern Mediterranean (c. 1500-1000 BC)*, edited by M. Heltzer and, E. Lipinski, 341-52. Leuven: Peeters.
- Zertal, A. 2004. *The Manasseh hill country survey. Vol. I-II*. Leiden, Boston: Brill.

The Long and Winding Road: Combining Least Cost Paths and Network Analysis Techniques for Settlement Location Analysis and Predictive Modelling

Philip Verhagen

VU University Amsterdam, The Netherlands

Tom Brughmans

University of Southampton, United Kingdom

Laure Nuninger

CNRS, UMR6249, Besançon, France

Frédérique Bertonecello

CNRS, UMR 7264, Nice

Abstract:

In this paper, we describe an exploratory analysis of the possibilities of combining least cost path analysis and network analysis techniques. Accessibility is a potentially important site location factor. So far, the definition of accessibility has been approached through the creation of accumulated cost surfaces and least cost paths. However, these methods do not provide direct information on the foci of movement. Starting from networks created from least cost paths, network analysis and space syntax were used to obtain additional information on the structural features of the network. It is concluded that both techniques can be used with least cost path-based networks, and will provide new insights into the characteristics of the network. For most applications however the space syntax measures that take the geographical dimension into account seem to be preferable to the simple network analysis measures used here.

Keywords:

Network Analysis, Space Syntax, Least Cost Paths, Accessibility, Movement

1. Introduction

When thinking about socio-cultural factors influencing settlement location choice, the accessibility of places in the landscape is a potentially important variable to take into account. Among the possible factors determining settlement location, access to resources and ease of movement in the landscape may have been important elements. For example, settlements might be preferentially located in areas that offer good access to prime agricultural land, and that allow them to interact easily with neighbouring settlements. How to define accessibility in such a way that it might be used as a variable for site location analysis and predictive modelling is however still very much open to debate. Most published research considering landscape accessibility limits it to the ease with which humans can reach a certain location. So-called hiking

Corresponding author: j.w.h.p.verhagen@vu.nl

equations are often used to obtain cost surfaces of accessibility, and accumulative cost surfaces are then applied to find the travel time or energy expenditure needed to reach a single destination from all points (pixels) in the area studied. By adding up these accumulated cost surfaces for each and every pixel, a map of differential accessibility of the landscape can be obtained (total path costs; Llobera 2000). This accessibility can also be analyzed for different travel times (short/medium/long distance; see also Mlekuž and Vermeulen in press).

However, these methods do not provide much information on the possible foci of movement in the landscape. We argue that to this purpose some additional steps are needed. We depart from the creation and addition of least cost paths (LCPs) from and to multiple locations in the landscape (see e.g. Zakšek et al. 2008; Whitley and Burns 2010; Murrieta Flores 2012; Verhagen in press). This will

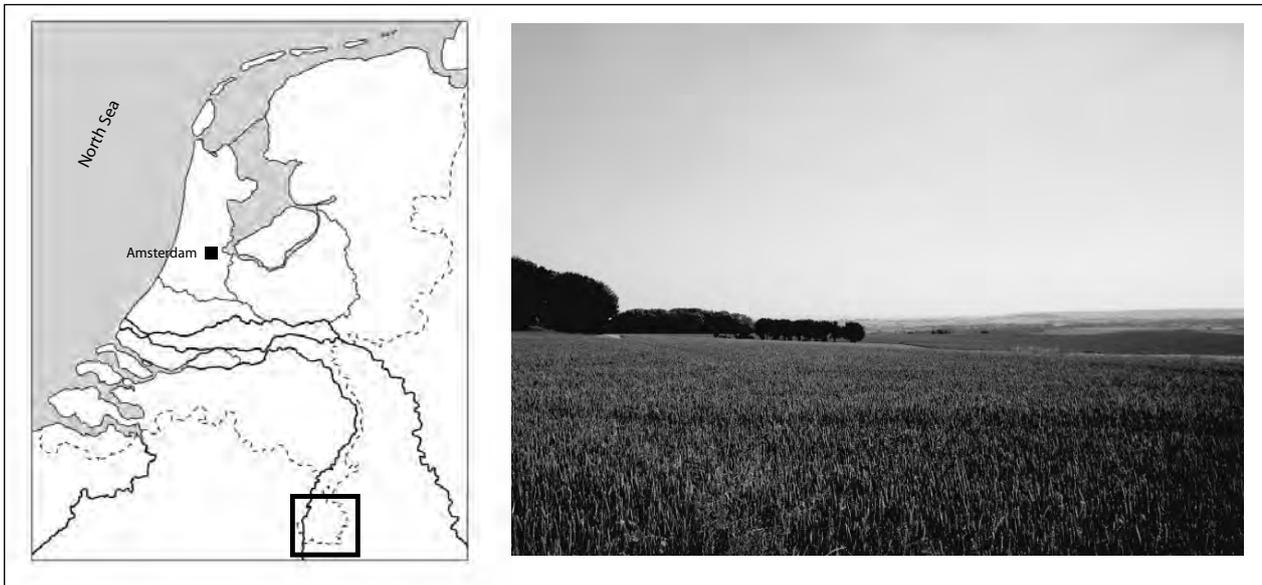


Figure 1. Location of the study area in the Netherlands, and a general impression of the landscape in the area.

result in structures of cumulative least cost paths that resemble networks with different weights attached to the edges, but that do not have any real 'nodes'. We can then assume that edges with a high weight will have been most attractive to travel, and thus may have been more attractive to settlement as well. However, since these 'networks' are raster-based, they do not allow us to explore their topological characteristics, as can be done through network analysis. The analysis of the connections and distances between network nodes can tell us something about their potential for interactions. We used simple exploratory network analysis measures and space syntax to examine a range of structural features of the junctions in the landscape that were identified by the multiple LCPs.

2. Case Study

As a case study, we have taken the region of Zuid-Limburg in the Netherlands (Fig. 1). This area, roughly measuring 32 by 35 kms, is characterized by undulating hills, dissected by valleys. Elevation values range between approx. 35 m a.s.l. in the Meuse Valley to more than 300 m in the southeast corner of the region (Fig. 2). Accessibility of the landscape is not greatly hampered by terrain, but the differences in elevation and slope are large enough to suppose that terrain conditions will have influenced movement preferences, as is in fact witnessed by the location of roads on historical maps. The elevation

model used was created using the ArcGIS 9.3 *Topo to Raster* tool² on the basis of contour lines digitized from 1920s topographical maps, scale 1: 25 000³. The vertical accuracy of the contour lines is 2.5 meters; elevation point measurements were added where available. The horizontal resolution used is 50 x 50 m.

A cumulative cost path map was then created using the method described in Verhagen (in press). Cost surfaces were calculated for sample points 250 m apart, using Tobler's (1993) hiking equation to specify the costs of movement.⁴ For each cost surface, LCPs were then created departing from 72 radially distributed points at a distance of 5 km from the sample point.⁵ The LCPs for each sample point were then added to create a cumulative cost path map, that represents the attractiveness of the landscape for movement within a 5 km radius (Fig. 3). It shows a dense network of paths, with a

² This is a discretized thin plate spline technique, for which the roughness penalty has been modified to allow the fitted DEM to follow abrupt changes in terrain, such as streams and ridges.

³ Chromotopografische Kaart des Rijks or Bonnebladen, the first detailed topographical maps of the Netherlands, made between the 1890s and 1930

⁴ $W = 6e^{3.5|s|+0.05|}$ where
W = walking speed in km/h; e = the base of natural logarithms;
s = slope in m/m

⁵ These were created using the ArcGIS (anisotropic) *PathDistance* and *CostPath* modules

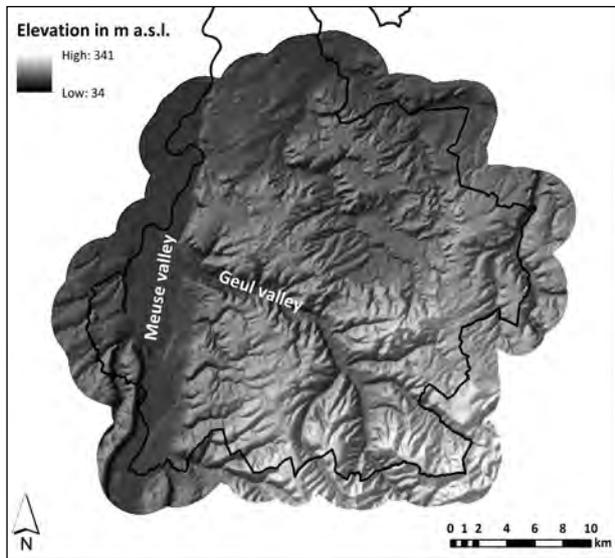


Figure 2. Digital elevation map of the study area.

concentration of movement potential in the valleys and on ridges, i.e. the areas with the gentlest slopes.

In order to investigate the potential of network analysis techniques, two approaches have been tested. The first one is based on simple node-based network analytical tools on undirected and unweighted edges (Newman 2010). The measures used include the number of nodes, connected components, clustering coefficient, density, heterogeneity, diameter, average shortest path length, average degree, closeness centrality, and betweenness centrality, they are all described in the next section. We decided to use these simple network measures because they reflect and summarise key structural features of networks (e.g. the ratio of links over nodes, the distribution of paths, the level of clustering), and they are therefore most commonly used to compare different network structures (Newman 2003). This case study aims to evaluate whether these measures can also be used for exploring network structures where geographical distance between nodes plays an important role. The open source software platform Cytoscape⁶ was used for this (Smoot et al. 2011), but other network analysis packages could be applied as well.⁷ This node-based approach was therefore confronted with a second approach that focuses on the connections. For this we applied measures commonly used in space syntax (Hillier and Hanson 1984, Bafna 2003)

⁶ www.cytoscape.org

⁷ Alternatives include Pajek (de Nooy et al. 2005), UCINET (Borgatti et al. 2002), and Gephi (Bastian et al. 2009).

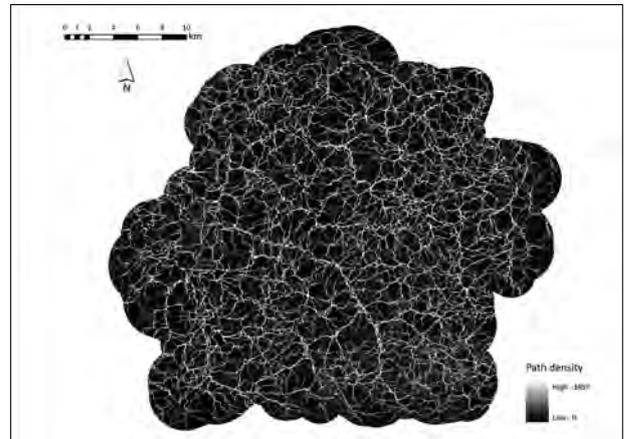


Figure 3. Cumulative cost path (or path density) map of the study region.

and implemented using Depthmap⁸ (freeware; Turner 2004); no competing products are available.

In order to prepare the raster-based cumulative cost paths for analysis in the vector-based network analysis packages, some pre-processing had to be done. We assumed that only the paths that are most frequently chosen in the LCP modelling would have been capable of supporting long-distance connections. We therefore extracted the upper 10% and 20% of most frequently chosen paths from the cumulative cost paths map. These needed to be thinned and converted to polylines, before exporting the resulting edges to Cytoscape and Depthmap.

3. Cytoscape Results

Only the most common network analysis measures provided in Cytoscape were used for this case study. On the one hand, the software offers a number of global measures that hint at structural characteristics of the total network. On the other hand, local measures are available that provide information on the structural position of individual nodes in the network. Definitions of measures and terms in network analysis can differ between software packages, so the definitions given here are specific to Cytoscape.

The data included a lot of loops (i.e. edges from one node to itself). These were removed before any analysis took place. A few edges are also duplicated because they appear in both directions (e.g. from A to B and from B to A). Since the LCPs do not

⁸ www.vr.ucl.ac.uk/depthmap

	Network 10%	Network 10% largest connected component	Network 20%	Network 20% largest connected component
Number of nodes	1605	1417	6525	6495
Connected components	74	1	16	1
Density	0.001	0.002	0.000	0.000
Heterogeneity	0.405	0.357	0.158	0.152
Clustering coefficient	0.045	0.049	0.059	0.059
Diameter	67	67	130	130
Average shortest path length	28.787	28.793	52.840	52.840
Average degree	2.314	2.456	2.886	2.895

Table 1. Results global measures Cytoscape.

represent directed networks these multiple edges are meaningless and they were removed before performing any analysis.

3.1 Results global measures

The following definitions apply to the global measures calculated by the Network Analyzer plugin in Cytoscape:

- A **component** is a connected subnetwork, i.e. nodes in one component are not connected to nodes in other components.
- The **average degree** is the average of the number of neighbours of all nodes.
- The **density** is a normalized version of the average degree that represents the fraction of all possible connections that are actually present (self-loops and duplicated edges are ignored).
- The **network heterogeneity** reflects the tendency of a network to contain hub nodes that are linked to many other nodes.
- The **clustering coefficient** calculates the average probability that two neighbours of a vertex are themselves neighbours, as a ratio of the number of edges between the neighbours of a given node and the maximum number of edges that could possibly exist between these neighbours (Watts and Strogatz 1998). The network clustering coefficient used here is the average of the clustering coefficients for all nodes in the network.
- The **network diameter** is the largest path

length distance between two nodes. If a network is disconnected, its diameter is the maximum of all diameters of its connected components.

- The **average shortest path length** is the average of all shortest path scores between all possible pairs of vertices in the network.

The analysis results are shown in Table 1. Both the 10% and 20% networks consist of one big connected component that includes over 80% of all nodes. All other nodes are part of a large number of small components or are isolated. It was decided to only analyze the largest connected component of each network. The networks are very linear, i.e. nodes tend to follow one after the other and rarely branch off. This is reflected in the low average degree, and is a direct result of using the cumulative cost path approach. This is also the reason why the density score and clustering coefficients are low, the diameter high and the average shortest path length very high. The overall structure does not seem to be very much affected by the network size, since both the 10% and 20% networks show similar results. This network structure means that it is highly likely for a number of local measures (like node centrality) to emphasize one or two paths. Also, the geographical and linear nature of this network makes it extremely susceptible to edge-effects: the peripheral nodes will almost always be underrepresented and will contribute to the overrepresentation of the core nodes.

3.2 Results local measures

A number of local measures were used but only two of them provided interesting results. The following definitions apply to the local measures

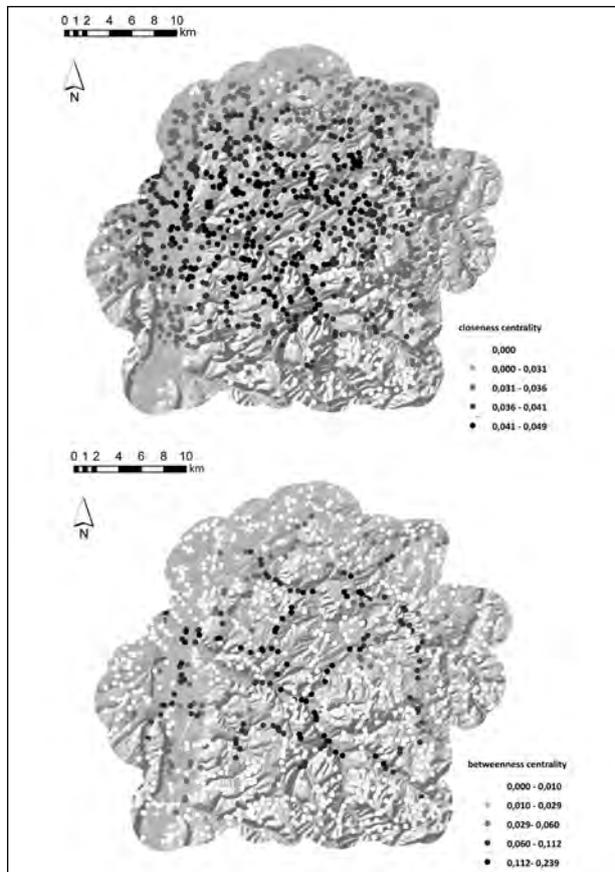


Figure 4. Closeness and betweenness centrality for the 10% network, obtained from Cytoscape and plotted on a hillshaded relief map of the study region.

calculated by the Network Analyzer plugin in Cytoscape:

The **closeness centrality** of a node is “the number of other vertices divided by the sum of all [topological] distances between the vertex and all others” (de Nooy et al. 2005, 127). Closeness centrality is often considered a measure of how fast information spreads from a given node to other reachable nodes in the network.

The **betweenness centrality** of a node is the proportion of all shortest (topological) paths between pairs of other vertices that include this vertex (de Nooy et al. 2005, 131). The betweenness centrality of a node is often taken to reflect the amount of control that this node exerts over the interactions of other nodes in the network. This measure favours nodes that join communities (dense subnetworks), rather than nodes that lie inside a community.

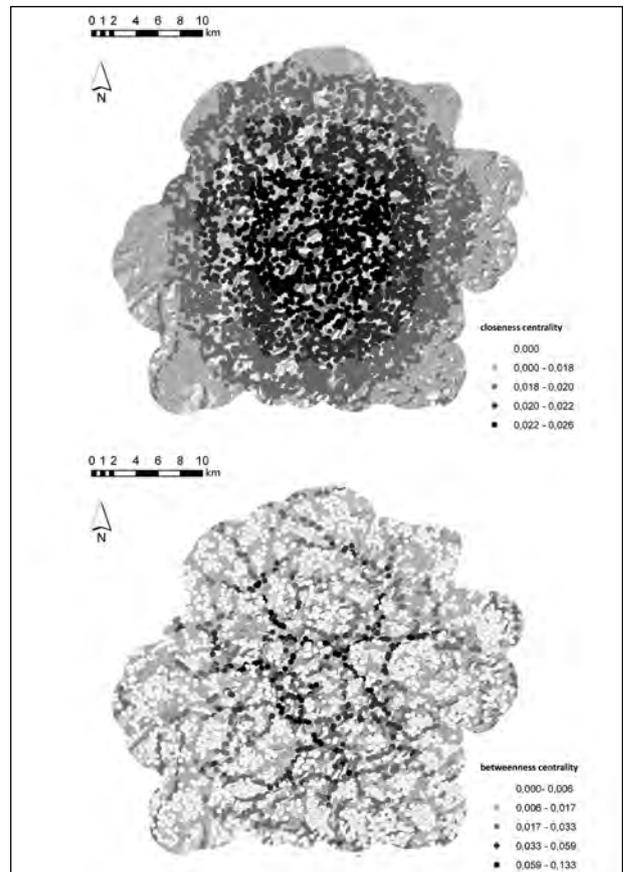


Figure 5. Closeness and betweenness centrality for the 20% network, obtained from Cytoscape and plotted on a hillshaded relief map of the study region.

The results of the analysis are displayed in Figures 4 and 5. For the 10% network the closeness centrality scores are generally low, and their frequency follows a normal distribution. A substantial number of nodes has a higher score, these seem to lie at the centre of the region and partly in the NW corner. The lowest scores are for nodes at the periphery of the network, particularly those at the eastern and southern edges. This entire pattern is a result of the sensitivity to edge-effects of this measure.

The frequencies of the betweenness centrality scores are more extremely distributed than for the closeness scores. Many nodes have a score of 0 or an extremely low score. Some have a relatively high score, which are the nodes that are frequently traversed. These higher scores seem to follow a number of distinct paths picked up by the least cost path analysis.

For the 20% network the core area with the highest closeness scores is still situated at the centre of the region, but no longer in the NW corner. The results are very similar to those of the 10% network, again stressing the influence of edge-effects on this measure.

The betweenness centrality scores show a similar distribution of scores as for the 10% network. The analysis again seems to pick up a number of more prominent paths. Although they are more or less in the same area as those of the previous network, the paths do seem to be different. It seems then that this measure is quite sensitive to network size and most importantly to the presence or absence of possible shortcuts which cause key paths to be diverted.

4. Depthmap Results

Space syntax applies a different terminology than many network analysis approaches, although there is strong similarity in the measures used since both approaches are rooted in graph theory. The Depthmap measures used here are calculated for the edges of networks and do not provide any direct information on the characteristics of the nodes themselves. Some of the measures are very similar to those used in Cytoscape, however, but applied to edges rather than nodes. It therefore becomes an interesting exercise to compare results of a node-based network analysis in Cytoscape with an edge-based space syntax analysis in Depthmap. This will allow us to evaluate which approach using freely available user-friendly software is most suitable for exploring geographical LCP networks. We have calculated the following measures in Depthmap:

- **Step depth** is the topological distance covered between two nodes, when passing through all the nodes between them, similar to calculating a path in Cytoscape.
- **Mean depth** is the topological distance from a node to all other nodes summed and divided by the number of nodes minus one. This is the inverse of closeness centrality in Cytoscape, where the number of nodes is divided by the sum of all paths.
- **Relative/relativised asymmetry** is the normalized mean depth on a scale of 0-1.

Mean depth and relative asymmetry are used to analyze the **integration** of the network. A highly integrated network has a large number of direct connections, it is symmetric. A segregated network has few direct connections, it is asymmetric.

- **Control** is the number of nodes that are connected to a node relative to the number of nodes that these have connected to themselves.
- **Controllability** is the ratio of nodes with depth 2 and nodes with depth 1 from a node

Control is used to calculate the **distributedness** of the network, i.e. if nodes dominate the structure of the network locally. If the structure depends on a relatively large number of nodes, then the network is distributed. If few nodes dominate the network, then it is non-distributed.

- **Degree centrality or connectivity** is the number of neighbours of a node (i.e. the degree of a node).
- **Closeness centrality** is the average topological distance from one node to all other, like the same measure in Cytoscape.
- **Choice** is number of shortest paths that passes through a node (i.e. the same as betweenness centrality in Cytoscape).

Note that the definitions are given here as measures for nodes. From the documentation it is not immediately clear how the measures are calculated, i.e. if measures derived from nodes are attached to edges, or calculated directly on the edges; the results obtained however suggest that the measures are calculated on the edges. Somewhat confusingly, the term 'node count' in Depthmap refers to the number of edges.

It is not possible to analyze subnetworks separately in Depthmap without breaking up the network, and running a new analysis on the subnetwork. Since Depthmap operates on edges that are placed in a (geographical) coordinate grid, rather than on nodes that only have topological connections, it can also include node-to-node metric distances in its analysis. Topological distance only takes into account the number of steps it needs

	Network 10%	Network 20%
Node count (number of edges)	5728	17703
Average connectivity	2.47727	3.05293
Maximum connectivity	5	6
Average step depth	67.8382	96.1734
Maximum step depth	133	203
Average mean depth	82.0956	98.7524
Maximum mean depth	133.388	147.438
Average integration	0.143313	0.120811
Average control	1	1
Average controllability	0.460255	0.407049

Table 2. Results global measures Depthmap.

to move from node A to node B. Depthmap can also calculate the actual metric distance it takes to move from A to B when passing through all the nodes between them, and in this way obtain metric measures as well. This leads to some differences in the analysis results. Depthmap can also perform analyses at different scale levels, by setting the radius of analysis to a maximum topological or metric distance.

4.1 Global results

The global measures obtained with Depthmap are shown in Table 2. The results are in line with those obtained through Cytoscape. The connectivity of the network is low, nodes will have few direct connections, resulting in a low average integration. In order to reach nodes in the network, a large number of other nodes has to be passed. This is the consequence of using the cumulative cost path approach.

4.2 Local results

The results for closeness centrality are comparable to those obtained with Cytoscape. More interesting are the results for the metric choice measure (Fig. 6). It shows very clear preferred pathways for the 10% network; for the 20% network patterns are less clear with a relatively even distribution of preferred pathways over the area. In both cases, the centrally located areas are given a higher weight, as a consequence of edge effects. For the 10% network, it is very clear that the preferred paths follow the valleys. A preferred route is also present on the north side of the region; this

is related to the effect discussed in 3.2, with the concentration of high closeness values in the NW corner of the area. On the basis of the Depthmap results, we can conclude that this is due to the lack of a direct connection in the central valley of the river Geul, forcing routes to the north instead of following the more direct east-west connection. In the 20% network, this connection is present and therefore the high closeness and metric choice values are concentrated in the centre of the area.

The differences between the choice (not shown in the figure) and metric choice parameters are not very conspicuous when visually comparing the

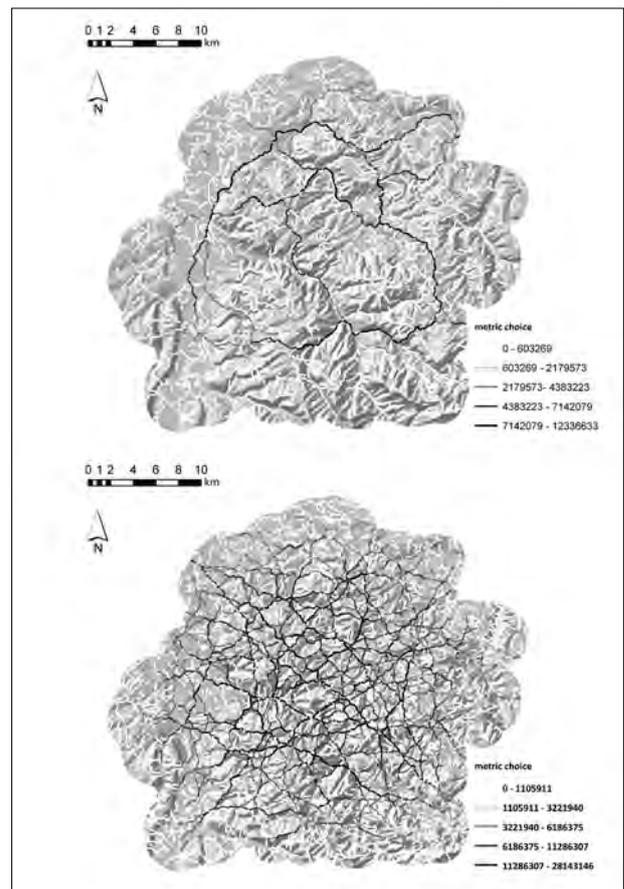


Figure 6. Metric choice for the 10% and 20% network, obtained from Depthmap and plotted on a hillshaded relief map of the study region.

results, although they are more striking for the 20% than for the 10% network. This is partly because of the presence of artefacts of the LCP modelling, that creates tightly packed clusters of nodes in places where many LCPs meet. This effect is stronger for the 20% network. Passing these clusters of nodes will increase the topological distance much quicker than metric distance.

5. Discussion

There are some clear differences between the two approaches used in this case study.

Cytoscape was used to only calculate measures for nodes, Depthmap for edges. It depends on the application whether one is preferable to the other. In archaeology, the characteristics of settlements in a network might be interesting to analyze, but it can also be useful to see where the major connections are. In the first case, a network package like Cytoscape is to be preferred, in the second case it is better to use space syntax. Exporting data from GIS to Depthmap is easier since it only needs an export to MIF/MID format. For use in Cytoscape, a list of connections in the format of source nodes and target nodes has to be produced.

Visualisation in both packages is completely different. Depthmap is very much comparable to a GIS, and retains the original spatial configuration of the edges, which makes interpretation of the patterns easier (at least for the purposes of this case study). The lay-out algorithms embedded in Cytoscape can display the nodes in a wide array of configurations, none of which take the spatial configuration into account. This has the advantage of emphasizing and visually identifying topological features through different lay-outs. In this case study, however, the networks were too large and the network analysis results not striking enough to motivate a visual topological exploration. In order to visually compare the analysis results from both packages, the data has to be exported back to GIS.

Both Cytoscape and Depthmap offer global and local network analysis measures. These are not fully comparable, even when strong similarities are found. We have not undertaken an exhaustive survey of all the differences and similarities between the techniques, even when this would be a useful

exercise in itself. It seems that Depthmap is less well documented; measures in the manual (Turner 2004) are usually described in fairly general terms, with references to the corresponding scientific papers. The analysis of general network structure and the identification of the most important connections can equally well be done in both packages. Cytoscape is somewhat stronger on the global measures, they are easier to extract using the Network Analyzer plugin and offer a wider range of options. The calculation of the control measure however is specific to Depthmap. The clustering coefficient of nodes calculated in Cytoscape is similar in that it reflects local differences in degree, but this measure did not provide interesting results for this case study.

The strongest point of Depthmap is that it offers the opportunity to calculate metric measures alongside the topological ones. In geographical space, real distance is an important factor to take into account when calculating network characteristics. Network analysis software packages do not offer many spatial analysis techniques as a standard feature. An alternative approach would be to add geographical distance to edges, but since these distances are relative to coordinate systems and projections a GIS import tool would make more sense, or a combination of a spatial database with a network analysis extension.

Furthermore, Depthmap can limit the calculation of centrality measures to a certain distance, so centrality will then be measured within a set radius, instead of the whole region. This reduces edge effects, and emphasizes the local areas of through-movement, offering a wider range of network features to explore.

The results obtained show that network analysis techniques offer some additional information on the network structure of the cumulative cost path maps that cannot be obtained from the raster-based maps. The networks created from the cumulative cost paths show a weak integration and high distributedness; many nodes have to be passed to go from one location to the other, and junctions usually connect only three edges. This is a consequence of using the cumulative cost path approach; using a different set of connections, e.g. between settlement locations, would lead to different results.

The major connections that are obtained through the calculation of betweenness centrality (choice) follow the valleys rather than the ridges. This is because the pathways created on the ridges are not very well connected to the rest of the paths. The steep slopes on the valley sides prevent those routes of becoming well-connected to other paths, and thus through-movement is concentrated in the valleys instead. However, the centrality measures calculated for the whole region also tend to emphasize the centrally located nodes; they are very sensitive to edge effects. Furthermore, the choice of network size (10% or 20% most frequently chosen paths from the cumulative cost path map) has a strong influence on the betweenness centrality. When less frequently used paths are included, the general structure of the network becomes less clear.

The results also show, however, that network measures incorporating geographical space provide the most interesting results. The global non-geographical network analysis results helped to identify the general structure of the dataset, the local network measures that would be most suitable and the kinds of results we could expect to emerge. All of the results of local measures we derived with Cytoscape, however, found an equivalent in results derived with Depthmap. It seems then that the spatial nature of networks derived by the LCPs is not trivial and significantly affects network analysis results. In order to attain non-trivial network analysis results the networks and measures need to address this spatial nature explicitly.

Acknowledgements

Part of this study was made possible through a VENI-grant awarded to Philip Verhagen by NWO ("Introducing the human (f)actor in predictive modelling for archaeology"). Travel funds were made available to Philip Verhagen and Laure Nuninger by the Réseau Franco-Néerlandais (Van Gogh-grant) and the French Ministry of Foreign Affairs (PHC Van Gogh).

References

Bafna, S. 2003. "Space Syntax: A Brief Introduction to Its Logic and Analytical Techniques." *Environment and Behavior* 35:17-29.

Bastian, M., S. Heymann, and M. Jacomy. 2009. "Gephi: an open source software for exploring and manipulating networks." Paper presented at the *International AAAI conference on weblogs and social media*.

Borgatti, S.P., M.G. Everett, and L.C. Freeman. 2002. *Ucinet 6 for Windows: Software for Social Network Analysis, Analytic Technologies*. Cambridge: Harvard University.

Hillier, B., and J. Hanson. 1984. *The Social Logic of Space*. Cambridge: Cambridge University Press.

Llobera, M. 2000. "Understanding movement: a pilot model towards the sociology of movement." In *Beyond the Map. Archaeology and Spatial Technologies*, edited by Gary Lock, 65-84. Amsterdam: IOS Press/Ohmsha.

Mlekuž, D., and F. Vermeulen. *In press*. "Time Geography, GIS and archaeology." In *Fusion of Culture. Proceedings of the XXXVIII Conference on Computer Applications and Quantitative Methods in Archaeology, 6-9 April 2010, Granada*, edited by F. Contreras, M. Farjas and F. J. Melero. Oxford: Archaeopress.

Murrieta Flores, P. 2012. "Understanding human movement through spatial technologies. The role of natural areas of transit in the Late Prehistory of South-western Iberia." *Trabajos de Prehistoria* 69:103-22.

Newman, M.E.J. 2003. "The structure and function of complex networks." *SIAM review* 45:167-256.

Newman, M.E.J. 2010. *Networks: an introduction*. Oxford: Oxford University Press.

Nooy, W. de, A. Mrvar, and V. Batagelj. 2005. *Exploratory social network analysis with Pajek*. Cambridge: Cambridge University Press.

Smoot, M., K. Ono, J. Ruschinski, P-L. Wang, and T. Ideker. 2011. "Cytoscape 2.8: new features for data integration and network visualization." *Bioinformatics* 27:431-32.

Tobler, W. 1993. *Three presentations on geographical analysis and modelling. 1) Non-isotropic modelling 2) Speculations on the geometry of geography 3) Global spatial analysis*. Santa Barbara: National Center for Geographic Information and Analysis.

- Turner, A. 2004. *Depthmap 4. A Researcher's Handbook*. London: Bartlett School of Graduate Studies, University College London.
- Verhagen, P. In press. "On the road to nowhere? Least cost paths, accessibility and the predictive modelling perspective." In *Fusion of Culture. Proceedings of the XXXVIII Conference on Computer Applications and Quantitative Methods in Archaeology, 6-9 April 2010, Granada*, edited by F. Contreras, M. Farjas and F.J. Melero. Oxford: Archaeopress.
- Watts, D.J., and S.H. Strogatz. 1998. "Collective dynamics of 'small-world' networks." *Nature* 393 (6684): 440-2. doi:10.1038/30918.
- Whitley, T.G., and G. Burns. 2008. "Conditional GIS Surfaces and their Potential for Archaeological Predictive Modelling." In *Layers of Perception. Proceedings of the 35th International Conference on Computer Applications and Quantitative Methods in Archaeology (CAA), Berlin, Germany, April 2-6, 2007*, edited by A. Posluschny, K. Lambers and I. Herzog, 292-98. Bonn: Dr. Rudolf Habelt.
- Zakšek, K., É. Fovet, L. Nuninger, and T. Pobodnikar. 2008. "Path Modelling and Settlement Pattern." In *Layers of Perception. Proceedings of the 35th International Conference on Computer Applications and Quantitative Methods in Archaeology (CAA), Berlin, Germany, April 2-6, 2007*, edited by A. Posluschny, K. Lambers and I. Herzog, 309-15. Bonn: Dr Rudolf Habelt.

Can Infovis Tools Support the Analysis of Spatio-Temporal Diffusion Patterns in Historic Architecture?

Jean-Yves Blaise and Iwona Dudek

CNRS, UMR CNRS/MCC 3495 MAP

Abstract:

This paper's main claim is that analytical reasoning on spatio-temporal diffusion patterns requires a step into abstraction that traditional figurative solutions like maps or 3D virtual models do not encourage. Accordingly, we investigate alternative research practices, namely infovis (Information visualisation) and visual analytics, where the focus is put on revealing patterns of change, and more generally on gaining insight on individuals and collections through visual means. We introduce four graphic combinations implemented on a test case – Zbigniew Dmochowski's architecture of Poland, a respected and comprehensive classification of architectural facts & trends in Poland over a millennium that combines morphological, stylistic and functional division lines. The contribution presents the pluses and minuses of these combinations, the arguments behind their making, how they have shed (or not) a new light on the test case.

Keywords:

Architectural heritage, Information Visualisation, Spatio-Temporal Data, Classification

1. Introduction

There are books analysts can rely on in order to analyse pieces of historic architecture in space (dictionaries, maps, etc.) or to spot the time slots they can be connected with (literature about History and History of Architecture). There are also some books that relate changes in shapes over time to a context, and thereby provide means for relations assessment: a typical example is Viollet Le Duc's encyclopaedia of architecture.

But in all three cases, graphics used almost never fully integrate the time parameter and the spatial features. They rather put them side by side, like in the classic timeline+ cartography paradigm, or provide low-level indications (a limited number of variables taken into consideration) like in the typical map+arrows or map+layers paradigms. Now maybe this was before the computer age? Well, of course not. What computer solutions can we today rely on if we need to understand and represent the patterns of diffusion of an architectural trend, in time and space? Animation techniques, where time is mapped by time? These techniques are well suited for following movements, but they also have proven inefficient in supporting notably explorative tasks.

Corresponding author: jean-yves.blaise@map.archi.fr

In this contribution, we investigate how solutions stemming from the fields of Information visualisation (*infovis*) can reveal patterns of change, and more generally can support background tasks researchers need to carry out when analysing the evolution of historic architecture (relate findings on individual cases to general knowledge, underline contradictions, local architectural inventions, foster the understanding of the individual's position with regards to trends at that time and that place, renew the interpretation of a stylistic classification; etc).

In a state of the art preliminary section, the contribution positions the above *infovis* field in terms of scientific legacies, and with regards to spatio-temporal datasets. In this same section we also shortly comment on the visual solutions that will be tested and then focus on the four combinations (Fig. 1) that we have implemented on a test case – Z.Dmochowski's *architecture of Poland* (Dmochowski, 1956).

However it has to be said that we shall make no general claim on the history of Polish architecture: Dmochowski's classification – although interesting by many aspects – relies on a subjective, partial selection of edifices. Our intent is to use his classification as a mean to investigate potential

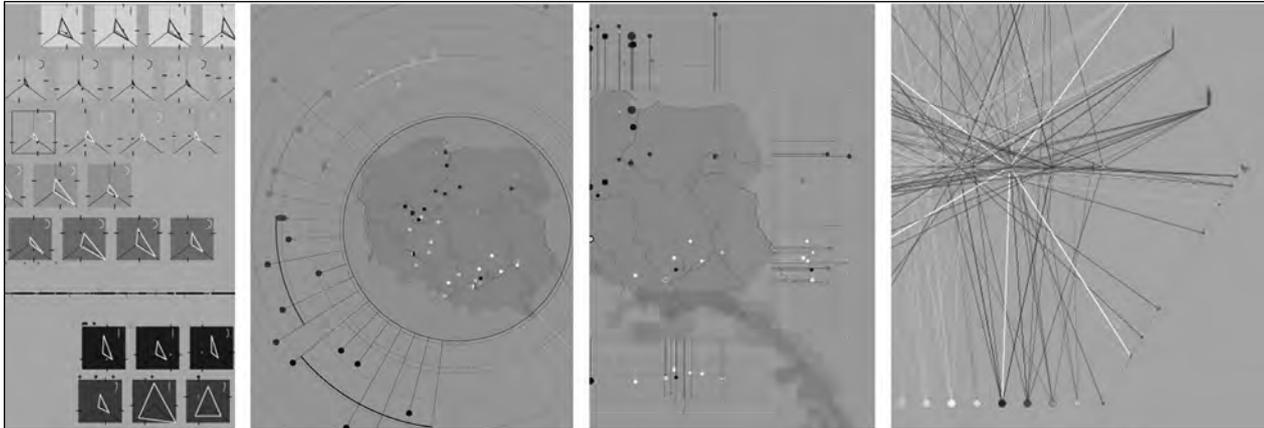


Figure 1. The four combinations, from left to right, Timeline + small multiples, concentric time, cartography + multidimensional icons, timeWheel (Aigner et al. 2008).

benefits of Infovis solutions:

- in uncovering unsaid biases,
- in uncovering spatial and temporal patterns,
- in underlining exceptions, unexpected behaviours,
- in supporting lacking / uncertain / imprecise data.

In the following sections we first come back to Bertin's vision of "*graphics as an answer to a question*" (Bertin 2005) and propose a short introduction to Z.Dmochowski's choices so as to underline *what we wish to understand*. We then present the pluses and minuses of our combinations and how they have shed (or not) a new light on the test case, before concluding on how the specificity of the data handled in historic sciences can be taken into consideration in the visual analytics process.

2. Conceptual Background

In his clear-cut "brief history of data visualization" Michael Friendly (Friendly 2006) identifies cartography and statistics as the two main scientific legacies of infovis, quoting for instance Galton's time series or Minard's figurative maps. By choosing these examples, Michael Friendly somehow already makes the point we wish to make: at the origin of infovis is a concern for spatio-temporal data sets, a need to foster a better understanding of spatio-temporal patterns, and at the end of the day the idea that graphics can help us conduct reasoning

tasks on such data sets so as to uncover unthought-of explanations. All through E.R Tufte's stunning written works further proofs are given that graphics can usefully back up analyses of trends in space and time (Tufte 2001; Tufte 1990; Tufte 1997; Tufte 2006). A number of ground-breaking contributions still emerge today on applications of infovis and related fields to spatio-temporal datasets, as exemplified in (Keim et al. 2011).

In the typical contexts of geovisualisation or GeoSpatial Visual Analytics fields, applications range for instance from analyses of moving objects (Biadgilgn et al. 2011) (Zhao et al. 2008) (Kapler and Wright 2005) – a revival of the time geography paradigm (Chardonnel 2007) - to more abstract developments like the use of visual metaphors for the assessment of semantic relations (Sabol and Scharl 2008).

In parallel, a number of contributions focus on the temporal aspects – may it be in terms of visual solution like (Havre et al. 2002)'s Theme River, or in terms of time-oriented data modelling challenges like (Aigner et al. 2008). Issues related to the production and readability of graphics – in the shadow of Bertin's graphic semiology- are also addressed such as the classic 2D vs. 3D debate (Vrotsou et al. 2010).

Yet it has to be said that infovis covers a wide range of application fields, far beyond spatio-temporal data sets – and a comprehensive overview of related works would be, if not unreachable, for sure irrelevant here. Moreover, the term *infovis* itself

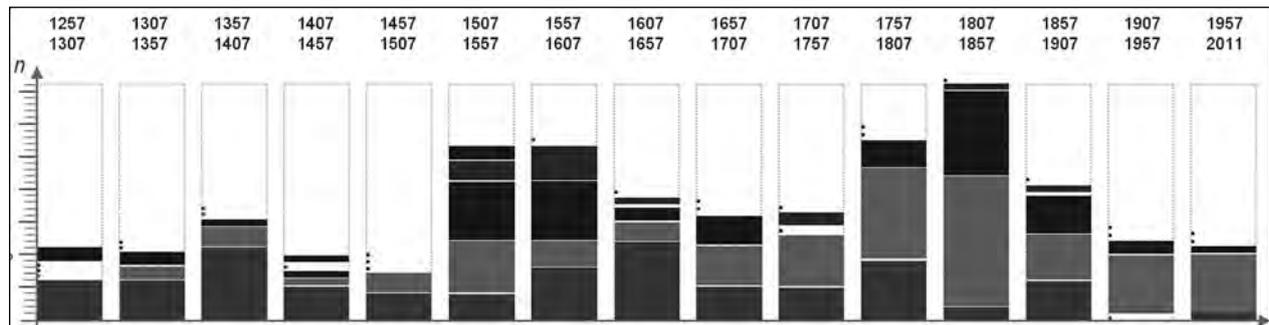


Figure 2. Represented as a classic histogram, a time series (50 years granularity) showing density of changes on Kraków's market square (Dudek and Blaise 2011). On the y axis, we represent the overall amount of changes and alternatives for the whole collection of 27 artefacts under scrutiny, counted every fifty years. Reddish rectangles represent, from bottom to top, morphological changes (a storey added to the artefact for instances), episodic changes (basically maintenance, however can include major changes that have no impact on the artefact's morphology such as change of roof material), destructive changes (partial or total destruction). Greenish rectangles represent confirmed (bottom) and unconfirmed (top, darker) alternatives. Confirmed alternatives mean we deal with possible divergent options on what did occur, basing on indications that are duly established and related to the artefact under scrutiny. Unconfirmed alternatives correspond to indications that may be taken into consideration or may be not – because the information is questionable, or more often because it is not directly related to the artefact under scrutiny but to a neighbour, or any semantic group it may belong to (typically useful when considering possible (but unaccounted for) consequences of a fire on a neighbouring artefact). In this visualisation note for instance for the first half of the XIXth century (1807-1857) that episodic changes (modifications - orange) widely outnumber destructive changes (brownish red). The visualisation shows the “pattern of destruction” usually associated with this period should rather be called “pattern of renunciation” in front of maintenance and repair difficulties.

has here and there been complemented or replaced by other terms such as *Knowledge Visualisation* or *Visual analytics*. So we consider more fruitful to briefly introduce in the next sub-section some definitions and pinpoint some differences, before coming back to this contribution's storyline.

2.1 Infovis and visual analytics

To start with a definition has to be given for the term “visualisation”, a term that today can be found in literature as a sort of flabby synonym for representation, and used in the context of numerous practices (abstract diagrams, cartography, CAD and GIS, realistic rendering, etc.). As far as information and data *visualisation* are concerned, visualisation is not about *drawing* but about *thinking* – in the words of (Card et al. 1999) it is about *using vision to think*. In his overview of the infovis field, R.Spence (Spence 2001) defines visualisation as a *cognitive activity*, underlining the idea that we rephrased in the informative modelling 15th rule: *If your graphics did not lead to new insight on your data, consider it useless* (Dudek and Blaise 2006). In other words, whereas representation is an end – the end of a

cognitive process; visualisation is a mean – a mean to perform reasoning tasks all along that process.

Infovis is defined by W. Kienreich (Kienreich 2006) as *the use of computer-supported, interactive, visual representations of abstract data to amplify cognition*. W. Kienreich shows that an infovis solution is most often composed of three fundamental units combined inside a master visualisation: visual formalisms (using a diagrammatic visual language to convey info in an abstract way – bar charts, timelines, histograms - see Fig. 2), visual metaphors (basing on real world equivalents – a city layout for a bibliography for instance), and visual models (in cases when the information is itself a real-world equivalent – typically cartography).

Infovis is an expanding field, with a number of branches we are not particularly interested in (in the context of this contribution). What makes an infovis solution distinguishable is the info (abstract information, large and/or complex information spaces) and the services expected (information seeking tasks, information discovery tasks).

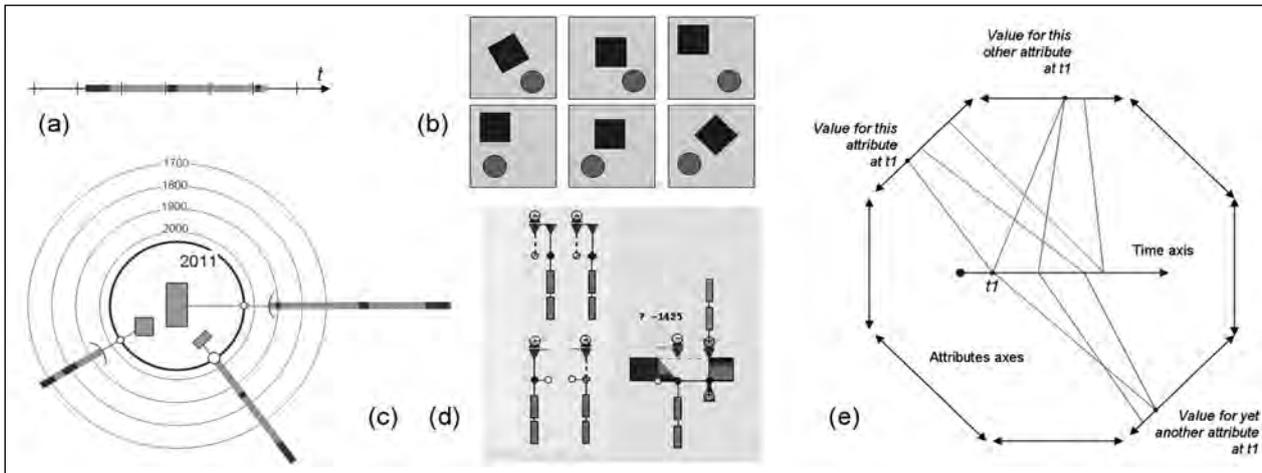


Figure 3. Principles of the five solutions combined.

By contrast, the sometimes disputed field of knowledge visualisation often works on smaller, but highly organized sets of information, and focuses on the transfer of knowledge among persons – may the information be abstract or not (Kienreich 2006).

Finally, the recent field of *visual analytics* is defined as an outgrowth of the fields of information visualization and scientific visualization: it focuses on “analytical reasoning facilitated by interactive visual interfaces” (Thomas and Cook 2005). Obviously the dispute on “which is which” is of poor relevance here. What is key to mention though is that these fields share on one hand overlapping legacies and on the other hand common techniques, goals, and sometimes methods.

2.2 The visual solutions tested

Over the years, a large number of visual solutions have been introduced to cope with spatio-temporal or time oriented data sets – an overview of this latter topic can be found in (Aigner et al. 2011). In this paper we shall analyse combinations of the following:

Timeline (Fig. 3, a) this is a concept we do not need to comment on very long, as its use is rather common (in particular in Western cultures where time is seen as linear). Naturally this is its major strong point: it is easily understood. In fact when analysing it from closer a timeline is strongly restrictive: it implies a de-facto discrete time model, needs continuity in the time scale to be readable, it

is poor in assessing rhythms and frequencies, and furthermore it better adapted to long walls than to screens where it requires tedious zoom and pan interactions (see on this Mostern and Johnson 2008 or Blaise and Dudek 2008).

Small multiples (Fig. 3, b): this expression was introduced by E.R. Tufte (Tufte, 1990), but the concept itself is attributed by M. Friendly (Friendly, 2006) to Christopher Scheiner’s 1611 solution to show the changing configuration of sunspots over time. The idea is to show in a grid-like layout successive spatial distributions of this or that object, of this or that phenomenon, thereby allowing a visual analysis of changes. This solution was for instance used by Galton in his XIXth century multivariate weather charts to uncover meteorological trends (Friendly, 2006), and applied on historical data sets in (Dudek and Blaise 2011).

Concentric time (Fig. 3, c): this is a visualisation we introduced in a recent experiment on the development of the market square in Kraków, that combines time and space. It is composed of a central 2D map, with all around it expanding concentric circles that represent a move towards the past. To each artefact positioned inside the 2D map corresponds a classic timeline, developing radially, along which events connected to the artefact’s evolution can be localised in time. Various interactions between the temporal indications along the timelines and the spatial indications inside the central 2D map are then implemented to facilitate the analysis (Blaise and Dudek 2011).

Multidimensional icons (Fig. 3, d): there are a number of striking examples of how multidimensional icons can help fostering insight “at a glance” in (Spence 2001) or in (McCandless 2010). The idea is to summarise the values of several variables for a given item in a collection through a simple 2D icon – often relying on analogies or metaphors. In the context of spatio-temporal datasets this solution however has drawbacks. Multidimensional icons are abstract representations (spatial features are hard to convey), and the number of variables that can be compared, in a readable way, is an issue (Chan 2006). Multidimensional icons are usually extremely efficient in depicting an individual item’s feature – they are less efficient in underling similarities and trends.

TimeWheel (Fig. 3, e): this concept, first introduced in (Tominski et al, 2004), combines a classic horizontal timeline with segments distributed around the timeline in the shape of a polygon. Each segment of the polygon represents the range of values of a given attribute: for instance a length, an age, a price, etc. An item is localised in time on the central timeline, and then lines connect it to values of its attributes on each segment of the polygon. Although at first glance it may appear complex, the concept is definitely efficient in underlining trends inside a collection. In return, one to one comparisons are not that readable and naturally spatial distribution is clearly hard to convey.

3. The Test Case

In the introduction of his *historical survey*, Z.Dmochowski, a Polish architect exiled in Great Britain during WWII, defines his objective as such: “*show the development of architecture within the polish State from the Xth to the XIXth century*” (Dmochowski, 1956, pp. xvii). Yet Dmochowski’s book is more than a survey, it is a de-facto classification - yet with little indications on the division lines used. Some hints are given in the introduction: “*the book divides the material according to style, but the distinctions should not be pressed too far*”. Dmochowski pushes this doubt further by adding “*It is a great help, in studying the history of architecture of any country, to know something of its political [...] background*”: a way to put stylistic affiliations into a wider perspective. Finally, his definition of architecture - “[...]

architecture, which of all the arts is the most socially conditioned” – further questions classic stylistic division lines. In short, Dmochowski acknowledges that styles are a key marker of trends, but in the same time that there are other aspects – context in a broad sense – that are equally key in a sensemaking classification effort.

Dmochowski’s survey is basically a description and a classification of individual edifices, with varying amounts of information given depending on the edifice. For a majority, elements of chronology and a spatial layout (2D plan) are given. Other pieces of information on the morphology are more unevenly available. Edifices are segregated by groups (one group one chapter) that correspond to the third step in a three-levels classification (Fig. 4):

- At step 1, styles act as the main division line (*Romanesque, Gothic, Renaissance, etc.*)
- At step 2, use and/or destination of the edifice refine the classification (*Romanesque secular architecture, Romanesque ecclesiastical architecture, etc.*)
- At step 3 – groups - architectural composition and choices are introduced in order to further segregate the collection (*Basilican churches with transept, Basilican churches without transept, Single cell circular churches, etc.*).

Beyond these three division steps there are other unsaid steps: space and time. Indications given here by Z.Dmochowski are quite hereterogeneous (dates like “XIIth century, or “1086”) and furthermore this variability differs from group to group.

Now at this stage it has to be made clear that our goal was not to come out with an exhaustive analysis of Dmochowski’s survey, but to conduct something like a like a proof of feasibility experiment. So the experiment we report was carried out on a subset of Dmochowski’s survey: 50 ecclesiastic edifices, 2 styles (Romanesque /Gothic) 10 groups (Fig. 4), time and space, plus a subjective selection of features (*our* interpretation step : dimensional and architectural features, corresponding to heterogeneous variables, not systematically present in the survey). In short, from Dmochowski’s survey

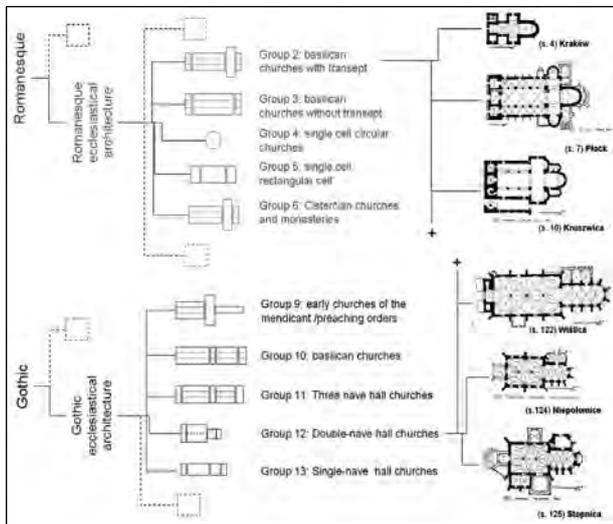


Figure 4. Classification levels, groups considered in this experiment, examples of edifices.

we extracted an information set that can appear as rather simple, with eight variables to combine (Fig. 5):

- Style,
- Group,
- Foundation date,
- Localisation (2D coordinates),
- Length,
- Shape of the Apse,
- Number of chapels,
- Number of Interior volumes.

Yet it might not be *that* simple –even with such a limited test case there are some clear difficulties:

- Unsaid division lines in Dmochowski’s classification,
- Heterogeneous variables,
- Lacking / imprecise pieces of data,
- Handling of BOTH time and space,
- Need to depict each item, its relation to others, and the global features of the collection.

Well these difficulties are precisely what makes his input a great test case: our objective will be to try and see, given these difficulties, whether

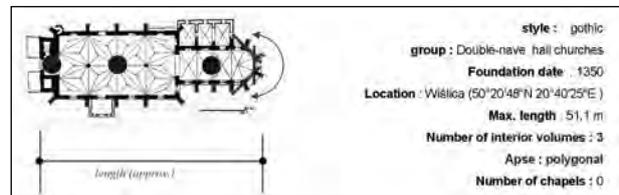


Figure 5. "ID card" of an edifice.

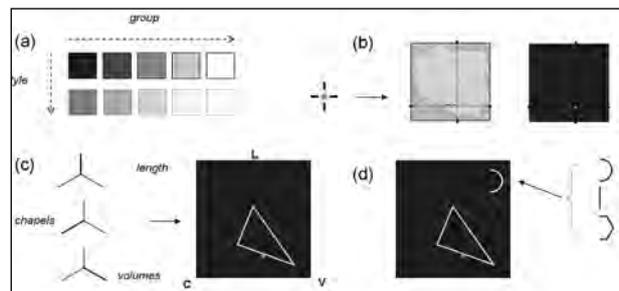


Figure 6. Visual encoding of an edifice's eight variables.

or not visual tools can help us uncover information and knowledge, whether they can help us gain some insight on his classification effort, on each of the 50 edifices, and on the 50 of them as a collection.

3.1 Timeline + small multiples

In this first combination a point corresponding to the foundation date of the edifice is localised along a classic timeline. In parallel, inspired by the small multiples precedent, we sum up inside a square the seven remaining variables through colour, value, position, shape. The square's background colour and its value are used to convey the style and group variables (Fig. 6, a). The square acts as an equivalent to the map of Poland – more precisely a map of Poland is inscribed inside the square. As a consequence the localisation (2D geographic coordinates) of the edifice can be interpreted as (x,y) coordinates inside the square (Fig. 6, b). Values for length, number of chapels and number of interior volumes are reported on three axes, forming a triangle (Fig. 6, c). Finally, a small icon on the top right corner identifies the shape of the apse (Fig. 6, d).

But then how can these squares be connected to a chronology, to the timeline? If we just align them with the timeline there will be either clumsy overlapping when edifices are close in time – or a lengthy visualisation unsuited to a context view (the whole collection within the eyespan). In fact

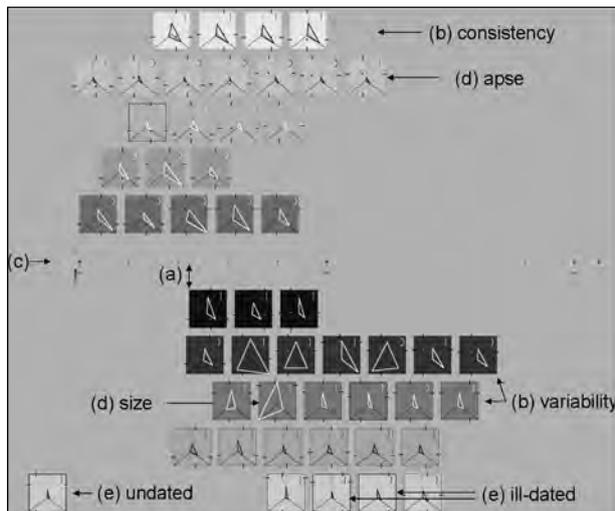


Figure 7. Combining multiples in ordered time.

this combination is not well suited to reading the chronology in discrete or continuous time. Instead, we use it to read the chronology in ordinal time (only the order is known): each group is represented with its oldest item duly aligned on the timeline (Fig. 7, a), and the rest of the group simply positioned horizontally by order of appearance. Interactions on each small multiple do allow the reader to precisely date it on the timeline (Fig. 7, c).

The result is a clear insight gained on the variability of items inside each group, and accordingly on groups consistencies (Fig. 7, b). Exceptional behaviours for one specific variable are

also underlined (Fig. 7, d), as well as cases when the dating is questioned or imprecise (Fig. 7, e). A point can for instance be made on this subset of Dmochowski's survey: the most "recent" group the least accurate in terms of dating...

A projection of each square along the timeline at a position corresponding to the actual foundation dates for each edifice can be added so as to highlight densities of change, or to correlate changes to other factors.

This combination does have its weaknesses: ordinal time is convenient when wanting to reveal group behaviours, but costly in terms of readability for individual items. Furthermore, as it is, from this visualisation only a rather vague understanding of the consistency of a group in space is within reach.

3.2 Concentric time + small multiples

This combination reuses the small multiple introduced above, but with the ambition to better handle and visualise the space parameter. A map of Poland is drawn in the central spatial layout of the visualisation, with coloured dots that position edifices (colour and value are used to convey the style and group variables). In cases when several edifices are localised in the same city rectangles are used instead of a dot. To each edifice also corresponds a dot on the concentric circles – the older the farther from the central spatial layout. Each item in the

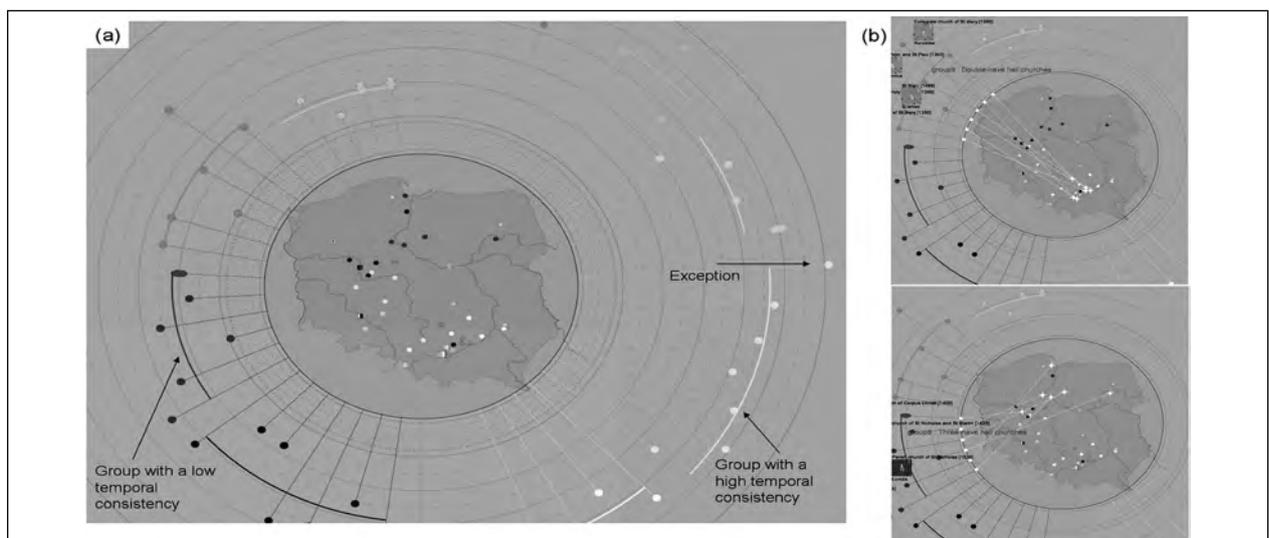


Figure 8. Assessing group behaviours with the concentric time visualisation.

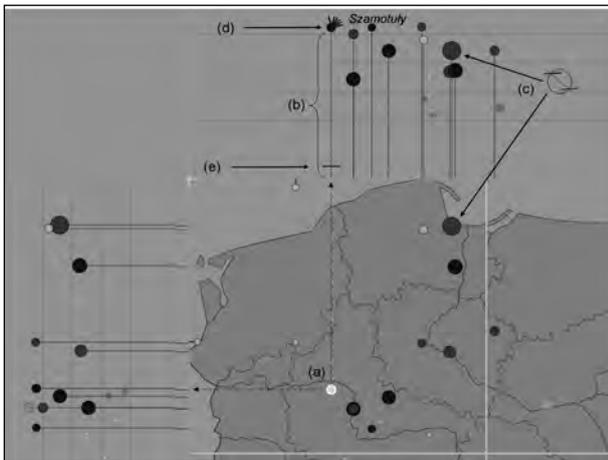


Figure 9. Combining a basic cartography with multidimensional icons.

collection, each edifice, is thereby localised in time and positioned in space. On user demand items can be browsed one by one, opening the small multiple and a plan view. This modality corresponds to a focus view: it allows one to one feature comparisons inside a group, between neighbouring edifices, etc. Alternatively the whole collection of small multiples can be shown together. This modality corresponds to a context view: it delivers an overview of spatial and temporal features across the collection.

At this stage the visualisation does allow a better analysis of spatial features, but groups are not as readable. In addition, by scattering squares all around the visual display, it in fact has little to do anymore with the very concept of small multiples. This combination can however be noticeably ameliorated when representing a group by an arc located at a position corresponding to the average foundation date for edifices of the group.

When doing so, temporal consistency inside the group is clearly highlighted (Fig. 8, a) – a point that could not be made with the previous solution. Furthermore, if interaction is implemented between the arc representing the group and the dots on the map, clear (yet unsaid in Dmochowski's survey) spatial patterns are revealed. For instance, group 9 "Double nave hall churches" appears typical of Małopolska (south-east, around the former Polish capital Kraków) whereas group 8 "Three nave hall churches" is widespread rather in the northern half of Poland (Fig. 8, b).

In short, this solution probably has a higher potential in rendering both spatial and temporal aspects. Yet it implies a steeper learning curve, and needs stronger interactions.

3.3 Cartography + multidimensional icons

When handling geo-related pieces of data one of the first reaction analysts tend to have is to produce some kind of cartography. And there is no reason to believe cartography can't be used in the context of infovis solutions. In this third combination we start from a basic 2D cartography of Poland on which we position edifices as we did in the previous solution (coloured dots). Each edifice is identified by the system as belonging either to the north or to the south part of Poland, and then either to the east or to west part of Poland. As a result each edifice on the map can be connected to two multidimensional icons located on the sides of the map (Fig. 9, a). The remaining variables are summarised in a multidimensional "flower-like" icon composed as follows:

- Length of the stalk corresponds to the distance, in time, of the current edifice to the oldest inside the collection (here 990). The longer the stalk, the younger the edifice (Fig. 9, b).
- Diameter of the corolla's centre: length of the edifice (Fig. 9, c).
- Number of petals: number of interior volumes (Fig. 9, d).
- Shape of bottom leaf : shape of the apse (Fig. 9, e).
- Number of leaves: number of chapels (displayed on user demand).
- Ill-defined positions and dating are represented by line styles and colour.

Used in context view (Fig. 10, a), the visualisation allows east/west; north/south comparisons. It helps for instance uncovering different distribution patterns between the west (regular distribution in space) and the east (imbalanced distribution in space), or between the south (high diversity in groups present, consistent

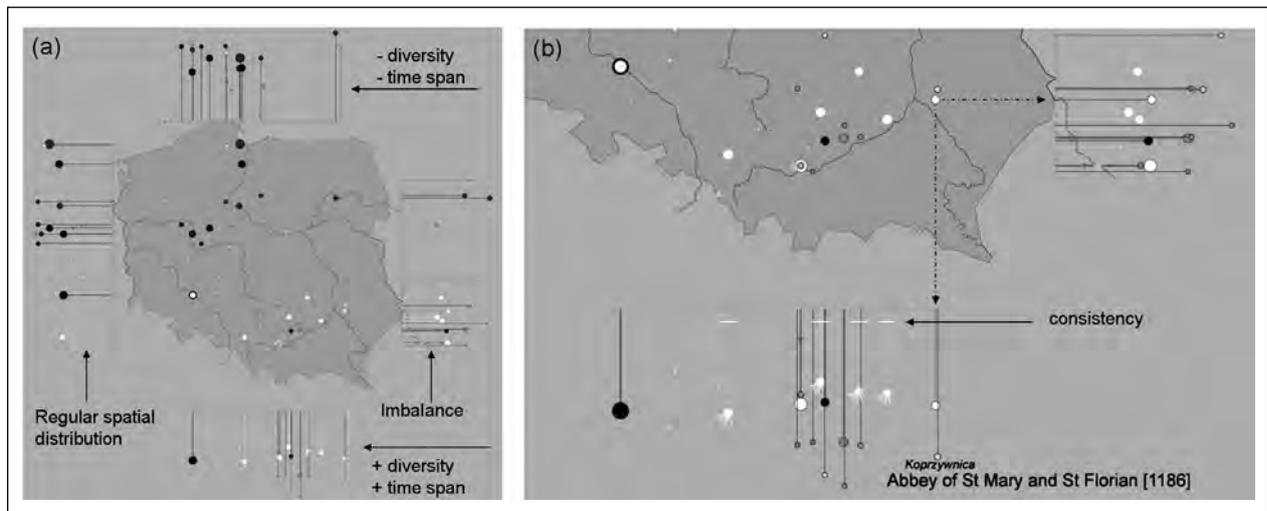


Figure 10. Pattern identification in context and focus views.

and long lasting time distribution) and the north (less diversity in groups, smaller temporal coverage).

Used in focus view (Fig. 10, b), the visualisation allows item to item comparisons among neighbours and helps for instance reading spatially consistent group behaviours.

On the overall, this solution over-performs the previous ones on one aspect mainly: its capacity to superimpose some layer of spatial analysis – exemplified here by the four sectors tested.

However the solution also has its drawbacks –the spatial analysis layer requires a high level of interaction, and temporal aspects are seemingly rendered with less efficiency. In detail, some choices on the making of the multidimensional icon are

questionable – typically representing a 1D value by a 2D object (mapping the length to the diameter of a circle) is a notably risky strategy.

Before developing the fourth solution it is important to make two intermediary statements. Whereas with cartography one tends to pack everything up inside a unique system of representation – in this approach – infovis applied to historic data sets - we rather make use various systems of representation. Up to here the three solutions proposed remain item-oriented: each edifice is represented separately, here by a small multiple, there by a multidimensional icon, etc. Accordingly these solutions allow one-to-one comparisons, but are not necessarily well suited to analyses of the collection as a whole. Typically, if my question is “are apses in majority polygonal” the above graphics can give the answer, but with a boresome decoding effort from the user.

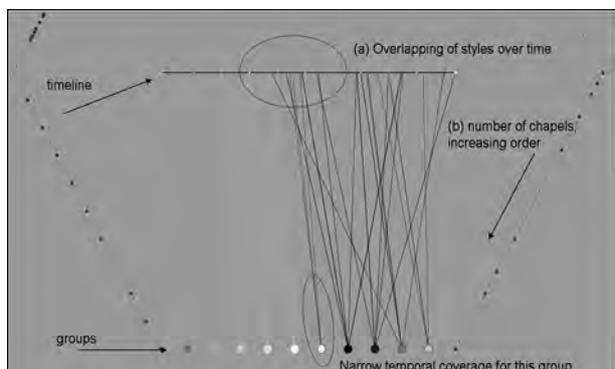


Figure 11. The TimeWheel concept, all variables values hidden except for groups.

3.4 TimeWheel

The TimeWheel concept (Aigner et al, 2008) combines a horizontal timeline with segments distributed it, along which values of various attributes can be reported. In our experiment foundation dates are distributed along the central timeline, the length of which being this of the time interval between the “oldest” and the “youngest” edifice in our collection. The polygon used is a hexagon, with a bottom side used to convey two variables: groups and style, reusing colour and value

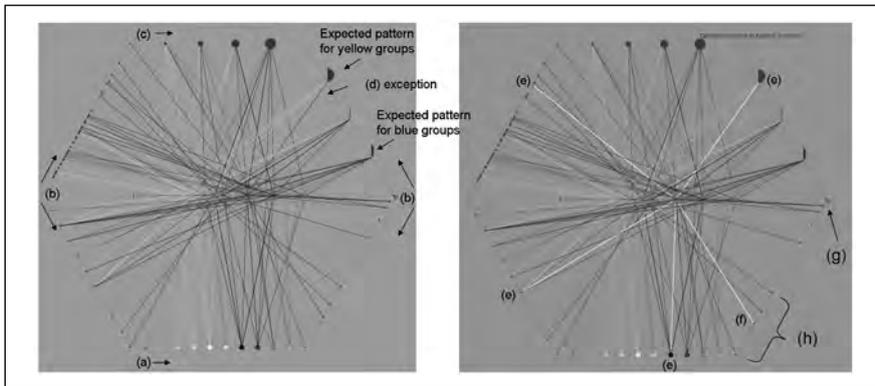


Figure 12. Distribution and cross examination of the variables.

as for the previous combination (Fig. 12, a). On either side of the hexagon are four segments used to convey the variables length, shape of the apse, number of chapels, and number of interior volumes (Fig. 12, b). Finally, we chose here to substitute to the geographical coordinates of the edifice another indication: the size (number of inhabitants) of its host city today (Fig. 12, c). At first glance the visualisation may appear rather visually crowded – but with some interaction it reveals its full potential (the time overlapping of styles is for instance clear-cut to read Fig. 11, a). We represent on each segment all the values for the corresponding variable inside the collection by a point (for instance, all lengths are aligned in increasing order). As a result densities are straightforward to read (Fig. 11, b).

Probably more interesting is the ease with which this graphics helps spotting expected patterns and exceptional behaviours. In the case we illustrate hereafter the visualisation shows that only one edifice classified by Dmochowski as “gothic” has a round apse (Fig. 12, d). When interactively highlighting this particular edifice (Fig. 12, e) further indications are given that can help the analyst get an overall idea of where this particular edifice stands with regards to other features (exceptional high amount of chapels in Fig. 12, f for instance).

Finally, this visualisation proves efficient in analysing how the value of a given variable varies over the whole collection. In the example above (number of chapels) value “zero” – no chapels – occurs almost in every group and style (Fig. 12, g). By contrast, high values always occur inside the same 2 groups (Fig. 12, h).

However, the “spatial distribution” feature is noticeably absent from this visualisation. This calls for a general remark: when dealing with spatio-temporal data sets in the context of history, one probably needs to come out with a specific, fine-tuned blend between abstraction – the infovis legacy – and figuration – the storytelling legacy.

3.5 Interaction and implementation

In the context of this written contribution we basically show screenshots of a set of computer-supported, interactive graphics. These graphics are created dynamically, at query time, from the reading of an RDBMS system storing the actual data. They are written in the standard SVG (Scalable Vector Graphics) format (ASCII format). The whole process is supported by ad-hoc programming, avoiding dependence on any piece of commercial software. Accordingly the can be easily duplicated, provided some knowledge of RDBMS and basic programming skills are at hand.

Graphics can be read on standard web browsers, and freely distributed. Interactions can be attached to each geometric object inside the SVG file, may it be a simple background line or a group – a multidimensional icon for instance.

We do acknowledge that a number of issues are left uncommented in this paper (possible data pre-processing here and there, escaping from two-dimensional visualisations, cognitive load and evaluation in general, etc.). The fact is that they would drive us too far from the paper’s main claim, and require quite extensive (and unplanned by the editors) explanations since visualisations we tried out would each call for specific comments.

4. Conclusions

The first conclusion we draw can be seen as a tribute to Bertin’s view: since efficient graphics are those designed as answers to a question, it is a necessity when facing several questions, to

promote several graphics. Indeed, the four solutions we have tested shed different lights on our data set. Accordingly there is nothing like a universal visualisation – there are only visualisations that do help uncovering information, and the others - useless visualisations.

Coming back to the test case, the visual solutions we have proposed did help us point out some interesting figures from Dmochowski's survey:

- basic spatio-temporal patterns, like a move northbound as time and styles pass by,
- architectural patterns, like higher or lower consistency in the architectural composition features depending on the groups,
- unsaid biases, like a relative (and questionable) homogeneity in the number of items presented in each group.

But it would take more than a sub-set of the survey, more detailed architectural analysis of the edifices chosen, more cross-examination with historic factors, to really question Dmochowski's classification.

At first glance, what we have proposed in this contribution is a method to re-read, re-interpret existing spatio-temporal data sets. From closer, we have exemplified this approach through a set of graphics that each with its pluses and minuses help uncovering patterns of the underlying data set, and have shed a new light on Dmochowski's classification effort.

Infovis and its outgrowth visual analytics are precisely about that: re-investigating a data set through visual means, and thereby giving ourselves chances to reach to some new conclusion, or at least to renew the way we picture the data set to ourselves.

However methods, tools and practises from these fields rely on human reasoning capabilities: the graphics by themselves are of no use if :

- They are not the result of adequate modelling and filtering choices (for instance keeping tracks of doubts),

- They are dedicated to communication rather than to perform reasoning tasks.

We hope this contribution showed infovis / visual analytics solutions can apply to test cases like ours, and can be fruitful even when dealing with history-oriented data sets. In the same time it has to be stressed not all of the infovis / visual analytics methodology - solutions from visual analytics are *not* solutions unless we fully master the process from data modelling to the production and coding of graphics. In general scientists dealing with history-related data sets should be cautious with statistics-oriented formalisms and/or automatic methods, like for instance clustering (possibly deceptive).

To conclude, we wish to add that infovis solutions should not be seen as not yet another technological magic wand – but they can , and probably should, be seen as yet another mean for us to re-examine data sets in the light of our knowledge of historical data sets.

References

- Aigner, W., S. Miksch, W. Müller, H. Schumann, C. Tominski. 2008. "Visual methods for analyzing time-oriented data." *Transactions on Visualization and Computer Graphics* 14: 47-60.
- Aigner, W., S. Miksch, H. Schumann, C. Tominski. 2011. *Visualization of Time-Oriented Data*. New York: Springer.
- Bertin, J. 2005. *Sémiologie graphique Les diagrammes - Les réseaux - Les cartes (Semiology of Graphics Diagrams, Networks, Maps)*. Paris: EHESS.
- Biadgilgn, D. M., C. A. Blok, O. Huisman. 2011. "Assessing the Cartographic Visualization of Moving Objects." *Momona Ethiopian Journal of Science* 3(1): 80-104.
- Blaise, J. Y., and I. Dudek. 2008. "Experimenting timelines for artefacts analysis: from time distribution to information visualization." In *Proceedings VSMM 2008 Digital Heritage, 197-204*. Archeolingua: Budapest.
- Blaise, J. Y., and I. Dudek. 2011. "Concentric Time: Enabling Context+Focus Visual Analysis of Architectural changes." In *Foundations of Intelligent Systems*, edited by M. Kryszkiewicz, H. Rybinski, A. Skowron, W. Raś, 632-641. Berlin, Heidelberg: Springer-Verlag.

- Card, S.K., J. D. Mackinlay, and B. Schneiderman. 1999. *Information visualization: Using vision to think*. San Francisco: Morgan-Kaufmann.
- Chan, W. W-Y. 2006. "A Survey on Multivariate Data Visualization." Accessed June 2012. <http://www.cse.ust.hk/~wallacem/winchan/research/multivis-report-winnie.pdf>.
- Chardonnel, S. 2007. "Time-geography: individuals in time and space." In *Models in Spatial Analysis*, edited by Lena Sanders, 97-122, London: ISTE
- Dmochowski, Z. 1956. *The architecture of Poland – an historical survey*. London: The Polish research centre.
- Dudek, I., and J. Y. Blaise. 2006. "Informative Modelling." *Mia Journal* 1. Accessed June 22, 2012. <http://www.map.archi.fr/mia/journal/>.
- Dudek, I., and J. Y. Blaise. 2011 "Visualizing alternative scenarios of evolution in heritage architecture." In *Proceedings of the 11th International Conference on Knowledge Management and Knowledge Technologies, ICPS*. New York: ACM.
- Friendly, M. 2006. "A brief history of data visualization." In *Handbook Of Computational Statistics: Data Visualization* edited by C. Chen W. Hardle A. Unwin, 15-56. Heidelberg: Springer-Verlag.
- Havre, S., E. Hetzler, P. Whitney, and L. Nowell. 2002. "ThemeRiver: Visualizing Thematic Changes in Large Document Collections." *IEEE Transactions on Visualization and Computer Graphics* 8(1): 9-20.
- Kapler, K., and W. Wright. 2005. "GeoTime Information Visualization." *Information Visualization archive* 4 (2): 136 - 146.
- Keim, D., J. Kohlhammer, G. Ellis, and F. Mansmann. 2011. *Mastering The Information Age – Solving Problems with Visual Analytics*. Konstanz: University of Konstanz
- Kienreich, W. 2006. "Information and Knowledge Visualisation - An Oblique View." *Mia Journal* 0(1): 7-17.
- McCandless, D. 2010. *Information is Beautiful*. London: HarperCollins.
- Mostern, R., and I. Johnson. 2008. "From named place to naming event: creating gazetteers for history." *International Journal of Geographical Information Science* 22(10): 1091 – 1108.
- Sabol, V., and A. Scharl. 2008. "Visualizing Temporal-Semantic Relations in Dynamic Information Landscapes." In *Proceedings of the 11th International Conference on Geographic Information Science AGILE-2008, Girona*, edited by G. and N. Andrienko, J. Dykes, S. Fabrikant and M. Wachowicz, 1-6. ICA Commission on GeoVisualization. <http://geoanalytics.net/GeoVis08/a15.pdf>
- Spence, R. 2001. *Information Visualisation*. Boston: Addison-Wesley/ACMPress.
- Thomas, J. J., and K. A. Cook. 2005. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. Washington: IEEE CS Press.
- Tominski, C., J. Abello, and H. Schumann. 2004. "Axes-Based Visualizations with Radial Layouts." In *Proceedings ACM Symposium on Applied Computing*, 1242–1247. New York: ACM Press.
- Tufte, E. R. 1990. *Envisioning Information*. Cheshire: Graphics Press.
- Tufte, E. R. 1997. *Visual Explanations: Images and Quantities, Evidence and Narrative*. Cheshire: Graphics Press.
- Tufte, E. R. 2001. *The Visual Display of Quantitative Information*. Cheshire: Graphics Press.
- Tufte, E.R. 2006. *Beautiful Evidence*. Cheshire: Graphics Press.
- Vrotsou, K., C. Forsell, and M. Cooper. 2010. "2D and 3D representations for feature recognition in time geographical diary data." *Information Visualization* 9(12): 263-276.
- Zhao, J., P. Forer, A. S. Harvey. 2008. "Activities, ringmaps and geovisualization of large human movement fields." *Information visualization* 7 (3-4): 198-209.

Introducing the Human Factor in Predictive Modelling: a Work in Progress

Philip Verhagen

VU University Amsterdam, The Netherlands

Laure Nuninger

CNRS, UMR6249, Besançon, France

François-Pierre Tourneux

Université de Franche-Comté, Besançon, France

Frédérique Bertonecello

CNRS, UMR 7264, Nice, France

Karen Jenson

Thermenmuseum, Heerlen, The Netherlands

Abstract:

In this paper we present the results of a study aiming at integrating socio-cultural factors into predictive modelling. So far, predictive modelling has largely neglected the social and cultural dimensions of past landscapes. To maintain its value for archaeological research, therefore, it needs new methodologies, concepts and theories. For this study, we have departed from the methodology developed in the 1990s during the Archaeomedes Project. In this project, cross-regional comparisons of settlement location factors were made by analyzing the environmental context of Roman settlements in the French Rhône Valley. For the current research, we expanded the set of variables with 'socio-cultural' factors, in particular accessibility, visibility, and the effect of previous occupation, and created predictive models from this. In this way, we have developed a protocol for predictive modelling using both environmental and socio-cultural factors that can easily be implemented for different regions and time periods.

Keywords:

Predictive Modelling, Socio-Cultural Factors, Regional Comparison, Diachronic Comparison, Roman Period

1. Introduction

Archaeological predictive modelling has a long history of application, especially in cultural resources management (see Judge and Sebastian 1988; Verhagen 2007; Kamermans et al. 2009). Despite its popularity for archaeological heritage management, it has also been the subject of substantial criticism from academic researchers (van Leusen 1996; Wheatley 2004; van Leusen and Kamermans 2005; Kamermans 2007). The goals of predictive modelling in heritage management are the accurate and cost-effective prediction of the location of archaeological remains within a limited region. However, academic researchers are usually more interested in finding explanations of why

archaeological remains are concentrated in particular parts of the landscape. Predictive modelling can be used as a tool for this purpose as well, but should be used with caution. Little attention is paid to the role of socio-cultural factors in prehistoric and historical site location choice (Verhagen et al. 2010). The result is a rather static way of modelling, in which the human factor remains unexplored. Furthermore, issues of temporality have been addressed uncritically or insufficiently. To maintain its value for archaeological research, therefore, predictive modelling needs new methodologies, concepts and theories.

For this study, we have tried to address this issue by developing a protocol for predictive modelling, using both environmental and socio-cultural factors, that can easily be implemented for

Corresponding author: j.w.h.p.verhagen@vu.nl

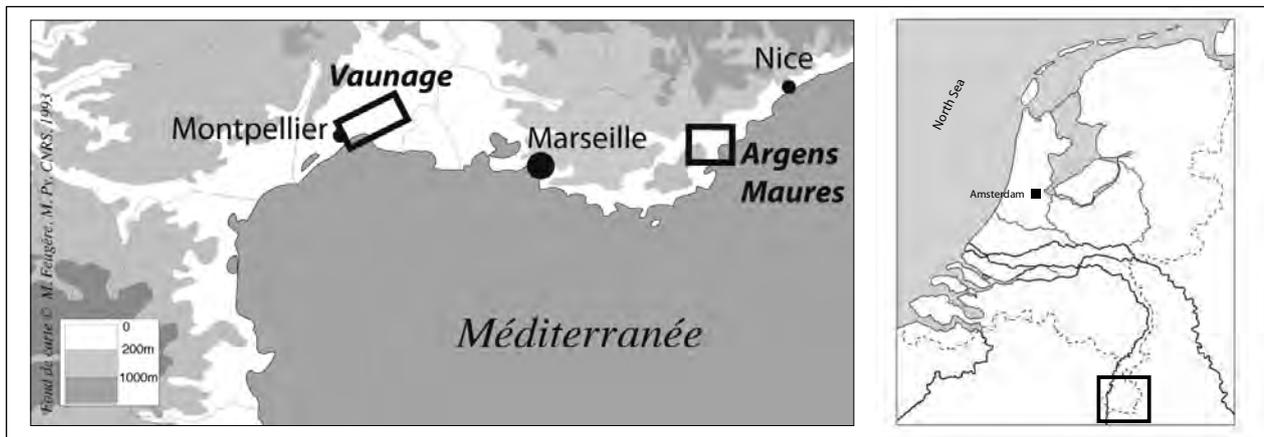


Figure 1. Location of the study areas in France and the Netherlands.

different regions and time periods. The methodology used was originally developed in the 1990s during the Archaeomedes Project (Favory et al. 1999; van der Leeuw 1998; van der Leeuw et al. 2003; Nuninger et al. 2008). In this project, the surroundings of Roman settlements in 8 study regions in the French Rhône Valley were analyzed in order to make cross-regional comparisons of potential settlement location factors, like slope, aspect, solar radiation and soil type. However, at the time this method was not used for predictive modelling purposes, and did not include an analysis of non-environmental factors. For the current research, we have adapted this approach and applied it in three different study regions, the Vaunage (Languedoc, France), Argens-Maures (Provence, France) and Zuid-Limburg (the Netherlands; Fig. 1).

When we are speaking of 'socio-cultural' variables, we can distinguish two different classes. On the one hand, we have variables that can be described as measurable attributes of the archaeological site sample itself; they are not related to any environmental factor. These include properties like site location, size, functional type or duration of occupation. They can be considered as expressions of forms of social behaviour, even when the interpretation of the specific behaviour involved may be subject to discussion. In themselves, these variables are not extremely difficult to obtain, but the problems of analysing and interpreting archaeological site databases are manifold and must be addressed before these properties can actually be used for predictive modelling. The second class of socio-cultural variables concerns features of the

landscape itself that can be interpreted as having socio-cultural significance, for example visibility, topographical prominence, or accessibility. These could be described as socio-cultural landscape variables, and are not necessarily excluded from 'traditional' predictive modelling. It can, in fact, be argued that all environmental variables have a cultural component, even though for most variables used in traditional predictive modelling this cultural aspect is usually related to subsistence economy, and not to e.g. ritual practices.

When thinking about ways to include socio-cultural factors in predictive modelling, two main approaches can be followed, that can be described as theory-driven (deductive) and data-driven (inductive) modelling (see Verhagen 2007). Both approaches have been applied to predictive modelling ever since the technique was developed in the 1970s, but data-driven modelling has long been dominant (see also Judge and Sebastian 1988; van Leusen and Kamermans 2005). Data-driven models take a set of input variables, usually representing natural landscape features, and use statistical methods to find correlations between archaeological site location and the environment. These correlations are then extrapolated to areas where no archaeological sites have been found. Theory-driven models on the other hand, depart from hypotheses about site location preferences, and combine the variables involved, usually by some form of multi-criteria analysis, into a predictive model. Theory-driven models have clear advantages over data-driven models: they involve the perspective of human decision making in



Figure 2. *The landscape characteristics of the study areas.*

modelling, and there is no need to develop elaborate archaeological site databases in order to run and test the models (Verhagen and Whitley 2012). There are however also disadvantages to this type of modelling, because it needs a well-defined theoretical framework for settlement location choice. Whitley et al. (2010), for example, developed an elaborate theory-driven predictive model for settlement location choice in the coastal areas of Georgia (USA), based on detailed historical and ecological information on subsistence strategies of the 16th century native American population in the area. In many cases, these frameworks are not available or insufficiently developed, or the necessary data may not be available. So, in practice, many predictive models will continue to be made using a data-driven approach, trying to get the most out of the available environmental and archaeological datasets. In those cases, the inclusion of socio-cultural factors can still be an option, but very few studies have actually tried to do this (but see Ridges 2006). The current paper takes a somewhat intermediate position, and uses theoretically informed data-driven modelling to better understand the role of socio-cultural variables for Roman rural settlement patterns.

2. Study Areas

The study areas were chosen because they all have good quality archaeological data of rural settlements in the Roman period. This allows us to make cross-regional comparisons and test the quality of the predictive models made. Two of those areas, the Vaunage and Argens-Maures, are regions that were already analyzed during the Archaeomedes Project. The data for Zuid-Limburg are part of a larger dataset stretching into adjacent areas of Belgium and Germany that was analyzed by Karen Jeneson for her PhD-research (Jenson in prep.).

The study regions show considerable differences in landscape (Fig. 2). The Vaunage region is a basin surrounded by low hills in the Languedoc west of the town of Nîmes, with a total area of 204 km². The Argens-Maures region consists of a mountain range (Massif des Maures) on the Mediterranean coast near Fréjus, with the Argens river valley running on its north side from east to west. The total area is 986 km². Zuid-Limburg is a hilly region of 914 km², with the river

Meuse running from south to north on its western margin. The major towns in it are Maastricht and Heerlen. The original research area stretches across the Dutch border into Belgium and Germany, but it was decided to restrict the study region to the Dutch part, since it was impossible to obtain environmental datasets at the same level of resolution for all three countries involved.

The archaeological data collected in the French study regions using systematic survey during the Archaeomedes Project have gone through a process of careful inspection and classification into functional categories on the basis of the available finds information (predominantly building material and site size). It allows for a chronological resolution of 100 years for most settlements. The dataset includes sites dating from 800 BC until 800 AD, so it covers a much longer time span than the Roman period. For Zuid-Limburg, the available information on site dating and function is much less detailed, no systematic survey was done, and in many cases the chronological resolution cannot be made more specific than 'Roman period' (12BC – 450AD). Nevertheless, within the Netherlands it is probably the most detailed and complete dataset available. We did not analyze geographical research biases for the three study regions, basically because of a lack of reliable evidence concerning this. The available evidence however does not indicate serious biases in data collection (Nuninger 2002), so we expect the data to be representative for all landscape contexts, except locally as in the Argens alluvial plain (Bertoncello 2011) - but this judgement is based on expert opinion.

3. Modelling Approach

Our main goal was to establish whether including socio-cultural factors actually makes a difference for the interpretation of site location patterns and predictive model quality. The modelling was set up using a restricted set of variables based on the available digital elevation models and archaeological settlement data. While soil type is known to have been a factor for Roman rural settlement location choice as well (Favory and

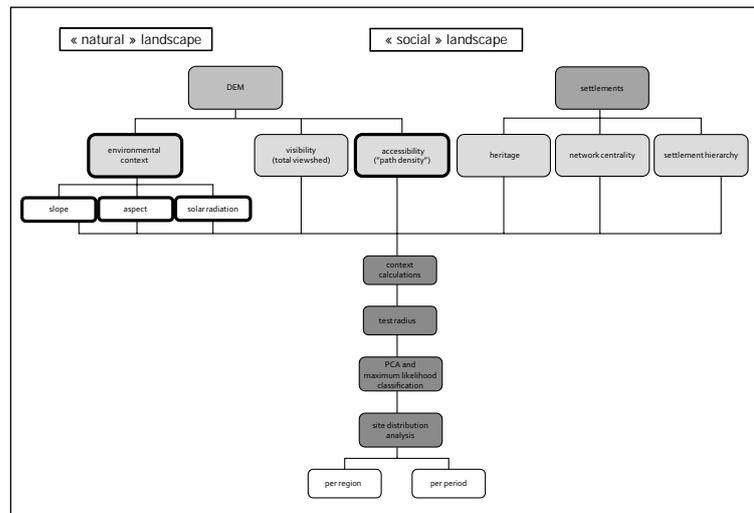


Figure 3. The modelling protocol. Outlined in bold are the factors that were actually analyzed for this paper.

van der Leeuw 1998), synchronizing soil mapping systems between different areas of France and the Netherlands would have been too complicated. The 50 x 50 m DEMs used for the French regions (originally obtained from the IGN) only contained integer values for elevation, and were therefore re-interpolated to floating point to get rid of the artificial terraces. For Zuid-Limburg, the impact of modern urbanisation and mining is much more evident than for the more rural French study regions. The modern-day disturbances contained in the available 5 x 5 m LiDAR-based DEM² proved to be too disruptive for the analysis, so a new DEM was created by digitizing and interpolating the 2.5 m interval contours from 1920s topographical maps scale 1:25,000³. These contours represent the situation just before large scale mining and urbanisation in the area started.

From the digital elevation models, three basic 'environmental factors' were derived: slope, aspect and solar radiation. These are thought to have been important for site location primarily because of their interest for agricultural production, and were already shown to have influenced site location patterns in the Archaeomedes Project. Furthermore, the DEMs were used to extract two additional factors that have a stronger socio-cultural connotation:

2 Actueel Hoogtebestand Nederland or AHN (www.ahn.nl)

3 Chromotopografische Kaart des Rijks or Bonnebladen, the first detailed topographical maps of the Netherlands, made between the 1890s and 1930

the accessibility of the landscape (approached through the calculation of path density maps; see section 4), and visibility (through the calculation of total viewsheds). From the settlement data, three additional factors were derived that are potentially of interest to settlement location choice: the impact of previous settlement ('heritage'), the position of sites in the settlement network, and the hierarchical position of the settlements (Fig. 3). In this paper however, we will only focus on the combination of the accessibility factor with the environmental factors, since the other analyses are still in progress. It illustrates the general process of model building, and gives some preliminary answers to the question if socio-cultural factors actually improve the predictive models and offer a new perspective on site location patterns in these particular archaeological and environmental settings.

An important element of the modelling is the concept of context: instead of just analyzing the landscape characteristics at the location of an individual grid cell, a radius around each cell is used. This method was originally developed in the Archaeomedes Project, and originally only involved the analysis of radii around settlements (Favory et al. 1999; van der Leeuw 1998; van der Leeuw et al. 2003; Nuninger et al. 2008). For predictive modelling purposes however, it is also important to know the characteristics of locations where no archaeological evidence is found, so for the current study all grid cells in the study regions were taken into account. First, it involves the categorization of the variables used into discrete classes. For example, the slope map was reclassified into 5 classes (0-2%, 2-4%, 4-8%, 8-15% and > 15%) that are roughly equally distributed over the total of the three study areas. For each cell in the study areas, the proportions of these discrete classes within a certain radius are then calculated.

It should be stressed that the choice of the size of the analysis radius is not based on theoretical considerations about site catchment sizes. Instead, the appropriate radius for analysis was defined as the one that provided most statistical contrast in the context profiles. If a category was present in more than 50% of the cases (i.e. context units centred on each cell), and if the standard deviation of the proportions for this category was > 15%, then the variable was considered significant. The radius with

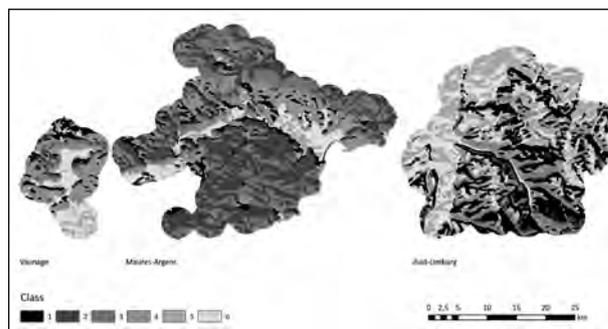


Figure 4. Maximum Likelihood Classification of the landscape into environmental factors.

the largest number of significant variables was then considered the most appropriate one. For each of the three study regions the most appropriate radius obtained was 250 m, corresponding to a circular region around each cell location containing 81 grid cells of 50 x 50 m.

In the first run of the model, only the environmental factors slope, aspect and solar radiation were analyzed, to provide a purely environmentally based model as a baseline. The context profiles were calculated for 16 variables for each cell in each region, a total of 819,398 grid cells. These profiles were then used to classify the landscape into broad, discrete regions sharing similar environmental characteristics. For this, first a Principal Components Analysis (PCA) was done, to identify and isolate the main statistical trends. On the basis of inspection of the scree plot of the PCA, it was decided to keep the 5 most important factors explaining almost 80% of the variation. A Maximum Likelihood Classification⁴ was done on these 5 factors, resulting in a map with 6 different landscape classes (Fig. 4). The resulting classes can be described in general terms as follows:

Class 1: predominantly north-facing medium slopes, low solar radiation (west- and east-facing slopes represented as well).

Class 2: steep slopes, predominantly low solar radiation, but sometimes with high solar radiation as well.

⁴ The PCA and Maximum Likelihood Classification were done in ArcGIS 9.3, using the Isocluster command to create signatures, and inspecting the resulting dendrogram to decide into how many classes the landscape should be classified.

Region	Period	χ^2	p	max. gain	# settl.	% settl.	# contexts (cells)	% contexts (cells)
All	800 BC - 800 AD	57.98	0.0000	9.7%	1178	100%	819277	100%
Limburg	Roman (12 BC - 450 AD)	7.97	0.1578	4.4%	378	100%	352971	43%
	Early Roman (12 BC - 70 AD)	5.29	0.3100	20.1%	26	7%		
	Middle Roman (70 - 270 AD)	4.35	0.5000	6.9%	54	14%		
Vaunage	800 BC - 800 AD	17.84	0.0030	8.0%	336	100%	81418	10%
	200 - 101 BC	11.59	0.0400	27.7%	29	9%		
	100 - 1 BC	7.49	0.1868	10.2%	75	22%		
	1 - 99 AD	22.11	0.0005	14.5%	181	54%		
	100 - 199 AD	20.57	0.0010	14.9%	160	48%		
Argens	800 BC - 800 AD	24.47	0.0002	10.4%	464	100%	384888	47%
	200 - 101 BC	6.86	0.2311	13.5%	76	16%		
	100 - 1 BC	1.85	0.8697	3.1%	134	29%		
	1 - 99 AD	59.43	0.0000	23.7%	205	44%		
	100 - 199 AD	62.68	0.0000	25.4%	189	41%		

Table 1. Site distribution and predictive value for the model based on environmental factors.

Class 3: south-facing steep slopes, very high solar radiation.

Class 4: predominantly south-facing gentle slopes, high solar radiation (west- and east-facing slopes represented as well).

Class 5: flat areas or gentle slopes, low to medium solar radiation.

Class 6: flat areas or gentle slopes, medium to high solar radiation.

Importantly, these landscape classes are based on the total number of context profiles over all three study regions. All classes occur in all three regions, but not in equal amounts, since the landscapes characteristics are different. Since the Vaunage region is much smaller than the other two regions, its contribution to the final classification is of course less important.

The archaeological sites were then overlain on the landscape classification map, and a χ^2 -test was employed to see if any significant site location preferences could be established. Relative gains (Wansleben and Verhart 1992; Verhagen 2007) were calculated to see how strong site location

preferences are (Table 1).

For Zuid-Limburg, the results are not statistically significant, and do not provide any evidence for a preference of settlement in a specific environmental context. For the Vaunage, the results are statistically significant and indicate an absence of preference as well. For the Argens-Maures region, a stronger patterning can be discerned. This is probably due to the more rugged relief in this region. The steeper slopes in this region will have made settlement more difficult than in the other two study areas. When looking at a more detailed chronological scale, we can see that for the French study regions the predictive power of the model is higher for the 1st century AD than for the 1st century BC, the first period of Roman settlement. It seems that during the initial agrarian colonization of the areas, environmental factors did not play an important role for choosing settlement locations. Later however, it seems that the environmental context became more important and settlements were located in more specific locations. This phenomenon was already observed in the Archaeomedes Project (Favory et al. 1999; Tourneux 2000), as well as in the Combas region north of the Vaunage by Fovet (2005). For Zuid-Limburg, this tendency seems to be reversed, with a more specific location choice in the first phase of

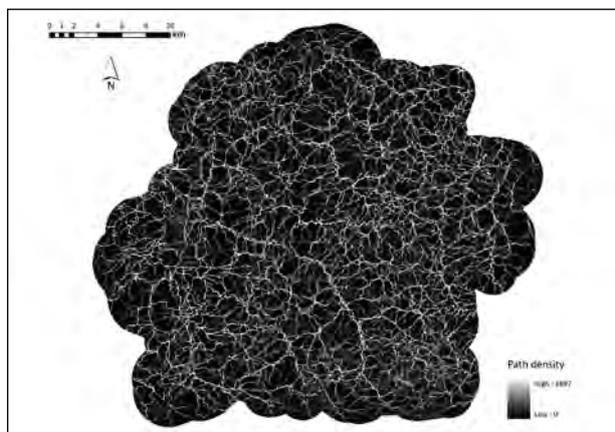


Figure 5. Path density map for Zuid-Limburg.

colonization, and an occupation of all environmental niches in the Middle Roman period. However, given the low number of accurately dated sites for this region, the reliability of this interpretation may be questioned.

4. Including Accessibility

Accessibility of the landscape is a variable that can be defined in a number of ways. For this study, we used the method developed by Verhagen (in press). For a number of sample points that are distributed systematically at a distance of 250 m in each study region, accumulative cost surfaces are calculated that give the cost distance to each sample point for each grid cell. The costs used are based on Tobler's (1993) hiking equation, in which slope is determining the speed of movement by foot⁵. We then used the accumulative cost surfaces to calculate least cost paths to the sample points, starting from 72 locations radially distributed on the edge of a circle with a 5000 m radius from the sample point. We then added the resulting least cost paths for each sample point to obtain a cumulative cost path, or **path density** grid, for the whole area. These cumulative cost paths represent the preferred movement axes in the landscape when moving over a distance of 5000 m, given the assumption of least effort (Fig. 5). However, alternative methods could be used as well, like the total path costs developed by Llobera (2000), or the potential path fields by Mlekuž (in prep.).

⁵ $W = 6e^{3.5|s+0.05|}$, where

W = walking speed in km/h; e = the base of natural logarithms;
s = slope in m/m

Due to the limitations of available computing power, the path density maps have not been calculated for the Argens-Maures region yet, and we can therefore only present the results for the Vaunage and Zuid-Limburg. The path density maps were classified into 5 discrete categories that are roughly equally distributed over both regions, and the context profiles of these categories were calculated within a radius of 250 m for each grid cell, in the same way as it was done for the environmental variables. The path density context profiles were then added as extra variables, and a new PCA and Maximum Likelihood Classification were performed on the extended set of variables. After inspection of the scree plot, the first 6 components of the PCA were retained, and the classification resulted in 5 classes.

Class 1: predominantly north-facing medium slopes, low solar radiation (west- and east-facing slopes represented as well).

Class 2: flat areas or gentle slopes, low to medium solar radiation.

Class 3: flat areas or gentle slopes, medium to high solar radiation.

Class 4: predominantly south-facing medium slopes, high solar radiation (west- and east-facing slopes represented as well).

Class 5: north-facing medium slopes, low to very low solar radiation.

The characteristics and geographical distribution of these classes are not very different from the original, environmentally based classification. The distinction between steep and medium slopes that is quite clear in the environmental classification however is less obvious in the new classification, and class 5 represents a specific category of cold, north- to northwest facing slopes. The differences in path density are not clearly reflected in the classes. This is probably due to the fact that the cumulative cost paths are relatively evenly spread over the study region. There are no distinct zones that are clearly avoided or preferred for movement. Despite the modest changes in class characteristics, a slight but significant improvement in predictive power can be observed for both regions

Region	Period	χ^2	p	max. gain	change in p	change in max. gain
All	800 BC - 800 AD	43.27	0.0000	11.9%	0.0000	2.1%
Limburg	Roman (12 BC - 450 AD)	13.05	0.0110	8.3%	0.1468	3.9%
	Early Roman (12 BC - 70 AD)	6.93	0.1400	22.1%	0.1700	2.0%
	Middle Roman (70 - 270 AD)	4.06	0.4000	11.5%	0.1000	4.6%
Vaunage	800 BC - 800 AD	26.00	0.0000	11.9%	0.0030	3.9%
	200 - 101 BC	2.80	0.5926	14.8%	-0.5526	-12.8%
	100 - 1 BC	7.59	0.1079	12.2%	0.0790	2.0%
	1 - 99 AD	22.13	0.0002	12.3%	0.0003	-2.2%
	100 - 199 AD	23.45	0.0001	13.7%	0.0009	-1.2%
Argens	800 BC - 800 AD	24.47	0.0002	10.4%	464	100%
	200 - 101 BC	6.86	0.2311	13.5%	76	16%
	100 - 1 BC	1.85	0.8697	3.1%	134	29%
	1 - 99 AD	59.43	0.0000	23.7%	205	44%
	100 - 199 AD	62.68	0.0000	25.4%	189	41%

Table 2. Site distribution and predictive value for the model including path density.

(Table 2). How to interpret this slight increase in predictive power is for the moment still open to debate, especially since it is not reflected when we look at the relative gains per period, and we have not looked at the individual contribution of path density to site location preference.

5. Conclusions

The results presented are preliminary, and we still need to improve on it, by including the path densities for the Argens-Maures region, and by experimenting with other approaches to accessibility that may be less demanding in terms of computing resources. Since the modelling protocol is clearly defined, it is relatively easy to include and analyze new variables. The models provide information on the relevance of variables for site location choice, and allow us to make cross-regional and diachronic comparisons of settlement pattern development. The prediction of potential settlement locations is implied, but is not the main goal of the exercise.

We want to stress that in this modelling protocol, the choice of settlement location factors for the modelling is not governed by considerations of optimal model performance, such as would be the goal of standard statistical approaches like logistic regression; the ‘non-performance’ of a variable is

an equally important result. Neither is this method based on an elaborate theoretical framework of Roman agricultural production, including all the possible factors that might have played a role in choosing a settlement location. Instead, we use the protocol to extract the main factors that influence settlement location over the longer term, and that have relevance for not just one study region, but for three regions in different geographical contexts. Evidently, there are limits to this approach: since the environmental side of the model is developed using elevation data, it will not work for areas that are completely flat. Similarly, certain social variables may not be relevant for other archaeological settings, or cannot even be modelled in all situations because of the limitations of the available archaeological data. For example, including the hierarchical status of settlements in the model implies that we can actually discern a hierarchy between settlements. Furthermore, the availability of good quality archaeological data sets is essential to this approach. In areas with little reliable archaeological data, a purely theory-driven approach will be more appropriate.

Acknowledgements

The research for this paper was partly made possible through a VENI-grant awarded to Philip

Verhagen by NWO (“Introducing the human (f)actor in predictive modelling for archaeology”). Travel funds were made available to Philip Verhagen and Laure Nuninger by the Réseau Franco-Néerlandais (Van Gogh-grant) and the French Ministry of Foreign Affairs (PHC Van Gogh).

The archaeological data for Zuid-Limburg were elaborated by Karen Jeneson for her PhD-research within the framework of the research programme “Roman villa landscapes in the north: Economy, Culture, Life-Styles” (financed by NWO). The archaeological data for Argens-Maures were prepared by Frédérique Bertoncello, and for the Vaunage by François Favory, Claude Raynaud and Laure Nuninger within the frameworks of the programmes ATP-Fréjus-Argens (financed by CNRS), Archaeomedes (financed by the European Union) and Archaeodyn (financed by the Ministère de la Recherche et des Nouvelles Technologies, and later by the Agence National de la Recherche).

References

Bertoncello, F. 2011. “Dynamique du paysage et centuriation dans le territoire de *Forum Iulii*, Fréjus (Var, France).” In *Sistemi centuriali e opere di assetto agrario tra età romana e primo medioevo, Atti del convegno Borgoricco (Padova), Lugo (Ravenna), 10-12 settembre 2009*, edited by Pier Luigi Dall’Aglia and Guido Rosada, 75-91. Pisa: Fabrizio Serra.

Favory, F., J.-J. Girardot, L. Nuninger, and F.-P. Tourneux. 1999. “ARCHAEOMEDES II : une étude de la dynamique de l’habitat rural en France méridionale, dans la longue durée (800 av. J.-C.-1600 ap. J.-C.)” *AGER* 9:15-35.

Favory F., and S. van der Leeuw. 1998. “ARCHAEOMEDES, la dynamique spatio-temporelle de l’habitat antique dans la vallée du Rhône: bilan et perspectives.” *Revue Archéologique de Narbonnaise* 31:257-98.

Fovet, É. 2005. “Dynamique spatiale du peuplement et analyse des ressources agro-pastorales dans le bassin de Combas et le vallon de l’Aigalade (Gard, France), du Bronze final au bas Moyen Age.” In *Temps et espaces de l’Homme en société, analyses et modèles spatiaux en archéologie : Actes des XXVe Rencontres internationales d’archéologie et d’histoire d’Antibes, 21-23 Octobre 2004*, edited by Jean-François Berger, Frédérique Bertoncello, Franck Braemer, Gourguen Davtian and Michiel Gazenbeek, 499-

503. Sophia-Antipolis: Éditions APDCA.

Jeneson, K. In prep. *Exploring the Roman rural villa world between Tongres and Cologne. A landscape archaeology approach*. Amsterdam: Vrije Universiteit Amsterdam.

Judge, W.J., and L. Sebastian, eds. 1988. *Quantifying the Present and Predicting the Past. Theory, Method and Application of Archaeological Predictive Modeling*. Denver: U.S. Department of the Interior, Bureau of Land Management.

Kamermans, H. 2007. “Smashing the Crystal Ball. A Critical Evaluation of the Dutch National Archaeological Predictive Model (IKAW).” *International Journal of Humanities and Arts Computing* 1:71-84.

Kamermans, H., M. van Leusen, and P. Verhagen, eds. 2009. *Archaeological Prediction and Risk Management*. Leiden: Leiden University Press.

Leeuw, S.E. van der, ed. 1998. *The Archaeomedes Project - Understanding the natural and anthropogenic causes of land degradation and desertification in the Mediterranean*. Luxembourg: Office for Official Publications of the European Union.

Leeuw, S. van der, F. Favory, and J.-L. Fiches, eds. 2003. *Archéologie et systèmes socio-environnementaux: études multiscalaires sur la vallée du Rhône dans le programme ARCHAEOMEDES*. Valbonne: Presses du CNRS.

Leusen, M. van, and H. Kamermans, eds. 2005. *Predictive Modelling for Archaeological Heritage Management: a research agenda*. Amersfoort: Rijksdienst voor het Oudheidkundig Bodemonderzoek.

Leusen, P. M. van. 1996. “GIS and Location Modelling in Dutch Archaeology, A Review of Current Approaches.” In *New Methods, Old Problems: Geographic Information Systems in Modern Archaeological Research*, edited by H.D.G. Maschner, 177-97. Carbondale: Southern Illinois University.

Llobera, M. 2000. “Understanding movement: a pilot model towards the sociology of movement.” In *Beyond the Map. Archaeology and Spatial Technologies*, edited by G. Lock, 65-84. Amsterdam: IOS Press/Ohmsha.

Mlekuž, D. In prep. “Exploring the topography of

- movement." In *Computational Approaches to Movement in Archaeology*, edited by S. Polla and P. Verhagen. Berlin: De Gruyter.
- Nuninger, L. 2002. *Peuplement et territoires protohistoriques du VIIIe au Ier s. av. J.-C. en Languedoc oriental (Gard-Hérault)*. Besançon: Université de Franche-Comté.
- Nuninger, L., F.-P. Tourneux, and F. Favory. 2008. "From Archaeomedes to Archaedyn." In *Layers of Perception, Proceedings of the 35th Computer Applications and Quantitative Methods in Archaeology Conference, Berlin, Germany, April 2-6, 2007*, edited by A. Posluschny, K. Lambers and I. Herzog, 278 and CD-ROM. Bonn: Dr. Rudolf Habelt GmbH.
- Ridges, M. 2006. "Regional Dynamics of Hunting and Gathering: An Australian Case Study Using Archaeological Predictive Modelling." In *GIS And Archaeological Site Location Modelling*, edited by M. W. Mehrer and K.L. Wescott, 123-43. Boca Raton: CRC Press.
- Tobler, W. 1993. *Three presentations on geographical analysis and modelling. 1) Non-isotropic modelling 2) Speculations on the geometry of geography 3) Global spatial analysis*. Santa Barbara: National Center for Geographic Information and Analysis.
- Tourneux, F.-P. 2000. *Modes de représentation des paysages*. Besançon: Université de Franche-Comté.
- Verhagen, P. 2007. *Case Studies in Archaeological Predictive Modelling*. Leiden: Leiden University Press.
- Verhagen, P. *In press*. "On the road to nowhere? Least cost paths, accessibility and the predictive modelling perspective." In *Fusion of Culture. Proceedings of the XXXVIII Conference on Computer Applications and Quantitative Methods in Archaeology, 6-9 April 2010, Granada*, edited by F. Contreras, M. Farjas and F.J. Melero. Oxford: Archaeopress.
- Verhagen, P., and T.G. Whitley. 2012. "Integrating Archaeological Theory and Predictive Modelling. A Live Report from the Scene." *Journal of Archaeological Method and Theory* 19:49-100.
- Verhagen, P., H. Kamermans, M. van Leusen, J. Deeben, D. Hallewas, and P. Zoetbrood. 2010. "First thoughts on the incorporation of cultural variables into predictive modelling." In *Beyond the artefact - Digital Interpretation of the Past - Proceedings of CAA2004 - Prato 13-17 April 2004*, edited by F. Niccolucci and S. Hermon, 307-11. Budapest: Archaeolingua.
- Wansleben, M., and L.B.M. Verhart. 1992. "The Meuse Valley Project: GIS and site location statistics." *Analecta Praehistorica Leidensia* 25:99-108.
- Wheatley, D. 2004. "Making Space for an Archaeology of Place." *Internet Archaeology* 15. intarch.ac.uk/journal/issue15/wheatley_index.html.
- Whitley, T. G., G. Moore, G. Goel, and D. Jackson. 2010. "Beyond the Marsh: Settlement Choice, Perception and Spatial Decision-Making on the Georgia Coastal Plain." In *Making History Interactive. Computer Applications and Quantitative Methods in Archaeology (CAA). Proceedings of the 37th International Conference, Williamsburg, Virginia, United States of America, March 22-26, 2009*, edited by B. Frischer, J. Webb Crawford and D. Koller, 380-90. Oxford: Archaeopress.

Changing Settlement Patterns in the Mediterranean Context: a Case Study of Menorca (Balearic Islands) from Prehistory to the 19th Century AD

Monica De Cet, Rainer Duttmann
Christian-Albrechts-Universität zu Kiel, Germany

Simón Gornés
Universitat de les Illes Balears, Spain

Joana Gual
Consell Insular de Menorca, Spain

Johannes Müller
Christian-Albrechts-Universität zu Kiel, Germany

Roberto Risch, Elena Sintes
Universitat Autònoma de Barcelona, Spain

Bianca Willié
Christian-Albrechts-Universität zu Kiel, Germany

Abstract:

Menorca (Balearic Islands, Spain) offers exceptional geographical conditions for examining complex human development in the Mediterranean. This paper investigates the settlement development of Menorca over the last four thousand years, beginning with the first occupation of the island. Using a well-established deep chronological resolution, a GIS-based approach has been applied to archaeologically assembled settlement data. The application of spatial analysis techniques provides significant results regarding site density, settlement pattern, and site location preferences. The usage of a deep time-span converts this research into the first long-term spatial analysis of the settlement record in a Menorcan framework.

Keywords:

Domestic Settlement, Site Density, Settlement Pattern, GIS-Based Approach

1. Introduction

The application of geographic information systems (GIS) is a powerful tool for investigating the complexity of socio-archaeological entities. Recent developments in the application of GIS software in archaeology have, in fact, led to the creation of a reference framework incorporating the methodological insights achieved in spatial archaeology during the last twenty years (Lock and Stančić 1995; Lock 2000; Wheatley and Gillings 2002; Conolly and Lake 2006; Verhagen et al. 2007). Spatial analysis, network analysis, map algebra, location models and prediction are some of

the most important advancements and means used by archaeologists in the spatial decoding of past human behaviour.

Menorca has received special archaeological interest for more than one hundred years, due to the exceptional monumentality of its prehistoric architecture (e.g. Ramis and Ramis 1818; Cartailhac 1892/1991). More recent research has promoted a growing interest in understanding the archaeological sequence, socio-economic development, and identification of socio-ideological transformations in the Balearic communities during later prehistory (Lull et al. 1999 a, b, c, 2002, 2005, 2008; Guerrero and Gornés 2000; Micó 2005 a, 2006; Gili et al. 2006).

Corresponding author: mdecet@gshdl.uni-kiel.de

The use of GIS approaches to understand Menorcan prehistory is relatively new. Initial analyses have been conducted on burial and domestic sites on the island using a GIS framework for the visualization of the prehistoric spatial trajectory and possible catchment areas (Gornés et al. 2004: 327-350). In addition, the general distribution of the prehistoric sites has been examined (Benejam 1993), together with the investigation of possible occupational strategies (García-Argüelles et al. 1994) and population sizes (Aramburu Zabala 2002).

The remarkable number of radiocarbon dates (ca. 800) supporting a solid chronological framework (for a recent overview, see Lull et al. 2008), and the substantial volume of archaeological, anthropological and palaeoecological data concerning Balearic prehistory, provides an exceptional context for a systematic spatial analysis of human development in the western Mediterranean. Using a GIS-based approach, a large array of data was collected and systematized over the past years, which highlights interesting patterns of site density, settlement pattern and location preferences since prehistory and up to the 19th century AD. While this research is part of a broader, multidisciplinary project² that aims to create a comprehensive model of the long-term demographic and environmental dynamics of Menorca, the present paper specifically intends to investigate the variability of the settlement evidence from the island.

2. Study Area

Menorca is situated at the easternmost point of the Balearic Island archipelago, in the centre of the western Mediterranean basin. It is the most remote island from the Iberian Peninsula, located between 200 and 400 km from eastern Iberia, northern Africa, southern France and Sardinia (Fig. 1).

With a total area of ca. 700 km² and 216 km of coastline, Menorca has been a Biosphere Reserve since 1993 and presents a variety of landscapes, but in particular two distinct geological and geographical



Figure 1. The Island of Menorca (Balearic Islands, Spain) in the Mediterranean Sea.

settings: The northern region with a rugged terrain called *Tramuntana*, consisted of geological material from the Palaeozoic, Mesozoic and Palaeogene, and the southern region, *Migjorn*, which is much less mountainous, mainly composed of Tertiary calcarenitic materials within a flat terrain. The central area of *Migjorn*, also known as *Barrancs* (ravines) is characterized by parallel, steep gullies running more or less perpendicular to the southern coast. Finally, there is also a north-eastern wetland area in the *Albufera des Grau* nature park (Didac 1997; Cardona et al. 1998; Fórnos et al. 2004). Despite the diverse elevation levels, Menorca has an overall flat surface and is often considered an “invisible island”. The highest elevation is a minor rise in the centre of the island, *El Toro*, which is 357 m above sea level (Fig. 2).

Archaeological records show that the Balearic Islands were permanently occupied since the second half of the third millennium cal BC (Alcover et al.

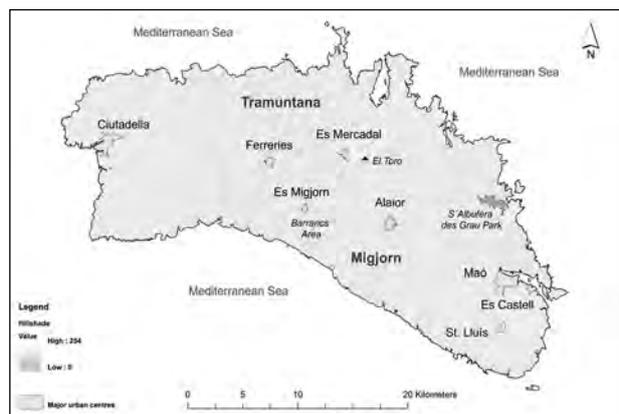


Figure 2. The Island of Menorca (Balearic Islands, Spain). Hillshade map obtained from Digital Elevation Model, courtesy of Ordenació del Territori Department - CIM Menorca, Spain. Software: ©ArcGIS 10.0.

² On-going PhD project with the working title “The Prehistory of Menorca: Environment and Demography”; PhD Student: Monica De Cet; scientific supervisors: Prof. J. Müller (Christian-Albrechts-Universität zu Kiel), Prof. R. Risch (Universitat Autònoma de Barcelona), and Prof. R. Duttman (Christian-Albrechts-Universität zu Kiel).

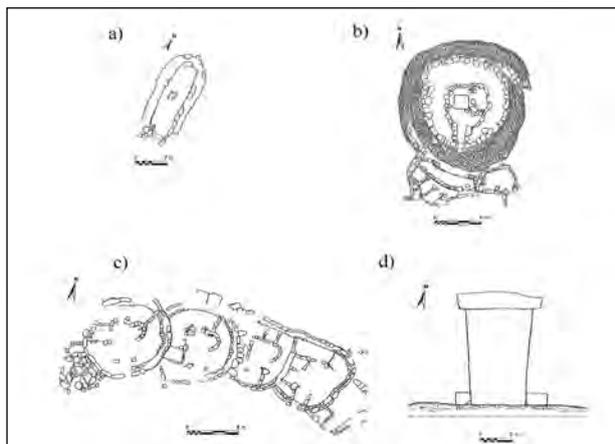


Figure 3. Architectural structural units used as indicators of chronology for Menorcan prehistory (Micó 2005 b, 2006). a) Naviform boat-shaped habitation structure: Naviform period c. 1600- 1100/1000 cal BC (Clariana, in Plantalamor 1991: fig. 1); b) Talaiot monumental tower: Talaiotic period c. 850- 550/500 cal BC (Torrelló, in Plantalamor 1991: plate IX); c) Casas Circulares circular-shaped habitation: Post-talaiotic period c. 550/500- 123 cal BC (S. Vincenç de Alcaldís, in Plantalamor and Rita 1979: fig. 119); d) Taula T-shaped monument: Post-talaiotic period c. 550/500- 123 cal BC (Torreta de Tramuntana, in Plantalamor 1991: fig. 198).

2001; Lull et al. 2004; Micó 2005). The earliest dates correspond to Majorcan sites and can be associated with the Bell Beaker Culture (2500/2300-2000 cal BC). In Menorca, human occupation began around 2100-2000 cal BC, with evidence showing shelter habitats and open-air settlement structures with perishable materials. During 1600-1500 cal BC, open-air settlements were abandoned and the first reliable domestic features appeared: Large, boat-shaped stone structures called *naviformes*. In the 9th century cal BC, a series of significant changes took place and settlements organized around one or several monumental towers called *talaiots* expanded across the island. The mid-6th century cal BC saw the end of the Talaiotic sites and the start of the Post-talaiotic settlements. The construction of unique buildings with centrally placed large T-shaped stone slabs, called *taules* (table) (Micó 2005 b, 2006), is representative of the architecture of this period. Moreover, Post-talaiotic communities maintained intense contact with the Punic, Greek and Roman world. Most notable is the participation of important detachments of sling-fighters in the Punic, and later also the Roman, armies. In 123 BC Menorca became



Figure 4. Historic pictures of Llocs, the agrarian economic structure taken into account as an indicator of chronology for the 19th century AD. Ses Cases in the Fornells area (Menorca 2012).

part of the Roman Empire. When this fell apart, the island successively shifted from the Vandal, to the Byzantine and finally to the Andalusian political spheres. Today, the historical and rural countryside persist over a large part of Menorca. Examples of this socio-natural heritage are the *Barrancs* with their hydraulic system originating from the Andalusian times and the *Llocs*, an agrarian economic structure from the 19th century AD. Further detailed chronology will be discussed in the next section.

3. Methodological Framework

Geo-referenced archaeological data collected since 2010 was the primary source for this study. We used a substantial database composed of 708 domestic sites based on the Menorcan Heritage database³, map archives, survey campaign, and detailed bibliographical analyses. The entire dataset has been linked to the geographical information regarding UTM coordinates and topography. Moreover, evidence from architecture, ceramic artefacts, numismatics, bibliography, epigraphy, and settlement extension was described for each record.

In order to effectively categorize each site, records were classified according to their chronological resolution. We have selected distinctive architectural units as indicators of chronology, chosen from time periods for which

³ Data obtained with the courtesy of Consell Insular de Menorca-Heritage Department - Pl. Biosfera, 5 - 07703 Maó.

reliable settlement information is available, based on the established framework of different archaeological structures used throughout the prehistory of Menorca (Micó 2005 b: 521-545; Micó 2006: 428); (Fig. 3). As evidence from the first settlement phase is very scant, the prehistoric domestic record was classified into the Naviform (c. 1600- 1100/1000 cal BC), Talaiotic (c. 850-550/500 cal BC) and Post-talaiotic (c. 550/500- 123 cal BC) periods. For the historical periods, historical sources and classical archaeological records provided a chronological time depth for the Roman-Republican (c. 123 cal BC-27 BC), the Early Roman Empire (c. 27 BC-284 AD), Late Antiquity and Dark Age (c. 425-903 AD) up to the Andalusian period (c. 903-1287 AD). Meanwhile, the Proto-talaiotic period (1100/1000-850 cal BC) and the Late Roman Empire (c. 284-425 AD) are hardly recognisable from survey or architectural information. Finally, the extraordinary conservation of agricultural, livestock, and demographic census data from the last six centuries and ethnographic heritage records⁴ allowed us to expand the chronological extent of this research to the 19th century AD (Fig. 4).

With the aim of analysing the long-term settlement behaviour through this dataset, we consider that the GIS-based approach needs to be focused on the social sphere of the communities and their interaction with the landscape. Accordingly, we developed a methodological strategy based on the analysis of site density and location pattern. Site density, through Kernel Density analysis (KDE), informs us about the dispersed or nucleated character of the settlement pattern and represents an indirect measure of demographic pressure on the island. On the other hand, the location pattern, determined by means of Nearest Neighbor analysis (NN), provides us with further details about the spatial organization of human communities. Both parameters do not necessarily need to correlate with each other. When they do, it implies that the population is dispersed across the island in a rather homogenous pattern.

All the methodological steps were carried out by applying standard functions within @ArcGIS 10.0 spatial analysis. First, in order to investigate

⁴ Ethnographic geo-referenced data of 948 "Llocs" from the 19th century AD obtained with the courtesy of Consell Insular de Menorca-Territori Department - Pl. Biosfera, 5 - 07703 Maó.

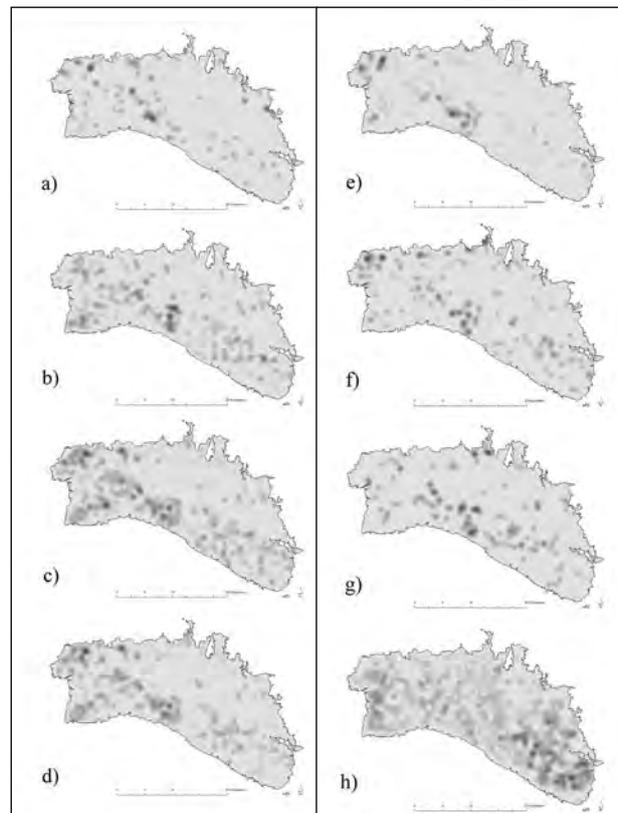


Figure 5. Left. Kernel density KDE maps output. Time periods: a) Naviform (c. 1600- 1100/1000 cal BC); b) Talaiotic (c. 850- 550/500 cal BC); c) Post-talaiotic (c. 550/500- 123 cal BC); and d) Roman-Republican (123 cal BC-27 BC). Software: @ArcGIS 10.0.

Figure 6. Right. Kernel density KDE maps output. Time periods: e) Early Roman Empire (27 BC-284 AD); f) Late Antiquity and Dark Age (425-903 AD); g) Andalusian period (903-1287 AD); and h) Llocs from 19th century AD. Software: @ArcGIS 10.0.

the density of archaeological sites, we used Kernel Density in the Spatial Analyst toolbox, which calculates the density of each site point feature around each output raster cell (Esri 2010 a; see also Beardah and Baxter 1996; Beardah 1999). Specifically, we used the 1000-metre circle radius, since it appeared as the most apt radius to choose considering the relatively small spatial scale that this island implies, together with the intensity of the settlement record. A larger radius would have implied considerable overlap between sites. Second, the mean distances were calculated through the function of Near (Esri 2010b) in the Proximity toolset. Moreover, with the aim of analysing the general location patterns in our

	Mean Distance
NAV	1193.94
TAL	738.38
PST	525.16
RRE	620.88
ERE	934.9
LAN	919.12
ISL	925.34
LLOCS	417.37

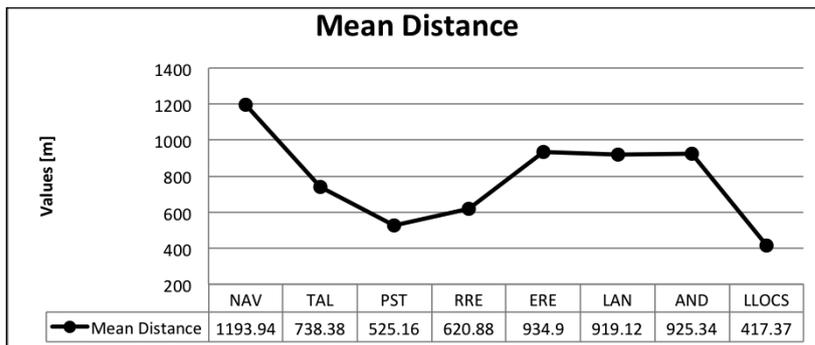


Figure 7. Mean distance values between settlements of the same time-period. NAV: Naviform (c. 1600- 1100/1000 cal BC); TAL: Talaïotic (c. 850- 550/500 cal BC), PST: Post-talaïotic (c. 550/500- 123 cal BC), RRE: Roman-Republican (123 cal BC-27 BC); ERE: Early Roman Empire (27 BC-284 AD); LAN: Late Antiquity and Dark Age (425-903 AD); AND: Andalusian period (903-1287 AD); and LLOCS: Llocs from 19th century AD.

dataset, we used Average Nearest Neighbor in the Spatial Statistics toolbox, which compares the observed average distance between points and their nearest neighbour with the distance that is expected between neighbours in a random pattern (Esri 2010 c; see also Hodder and Orton 1976; Ives 1985). Finally, preliminary site location preferences were estimated by combining settlement points with the geological map⁵, using the Intersect function in the Overlay (Esri 2010 d) toolset, and through statistical Chi-square tests.

4. Results and Discussion

During the last decades, the Consell Insular de Menorca and its Heritage Department have actively documented the archaeological sites of the island, through systematic revision of the abundant bibliographic information and field walking. The island of Menorca has suffered relatively little urban pressure compared to the other Balearic Islands. Hence, the exceptional preservation of the island's archaeological heritage allows Menorca to be considered as an "archaeological paradise". The quality of the spatial record for this case study can be expressed quantitatively in terms of a mean site density of 1.04 sites per square km; an exceptionally high figure by western Mediterranean standards. While former studies have often focused the attention on specific areas, our dataset covers the entire island of Menorca and has a more accurate chronological resolution.

⁵ Data obtained with the courtesy of Consell Insular de Menorca-Territory Department - Pl. Biosfera, 5 - 07703 Maó.

4.1 Site density

Our results using KDE show an increase in the number and density of settlements from Naviform to the Post-talaïotic period (Fig. 5). On the contrary, from the Roman Republican period towards the Early Roman Empire there is a decrease in density (Fig. 6). Specifically, this change in density is illustrated by the variation in the mean distance between domestic sites of the same time period. Figure 7 indicates a clear increase in the values from the Post-talaïotic period (738.35 m) towards the Early Roman Empire (934.9 m). A return to short mean distance values is clearly detected from the 19th century AD data (417.37 m).

The fluctuations in site density across the island also suggest a gradual shift in occupation density from the west of the island towards the east. As clearly visible in Figures 5-6, the prehistoric population is denser in the west and southwest areas of Menorca. Meanwhile, during the Andalusian period, human occupation seems to have been more intense at the centre of the island, specifically in the area of *Barrancs*. In contrast, 19th century AD occupation is denser in the eastern part of Menorca where the city of Mao, the capital and main demographic centre and market for agricultural products, is located.

An unexpected observation from the site density maps is the intense prehistoric occupation of the northwest of Menorca, as this area has a very rugged landscape with little soil cover on the sandstone bedrock in the present day, due

	Nearest Neighbor
NAV	0.9
TAL	0.86
PST	0.84
RRE	0.89
ERE	0.84
LAN	0.86
AND	0.77
LLOCS	0.97

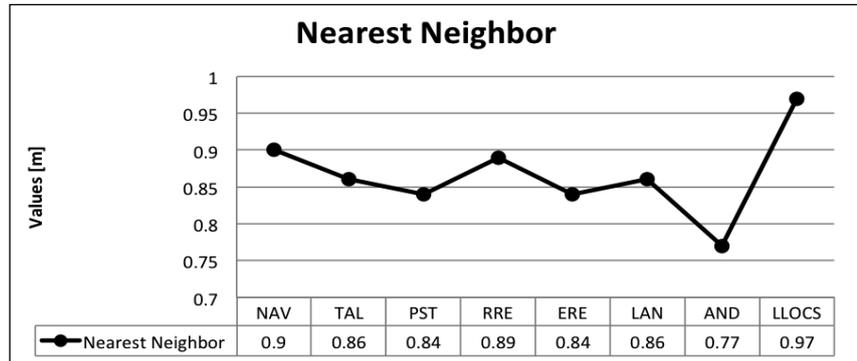


Figure 8. Nearest Neighbor values. NAV: Naviform (c. 1600- 1100/1000 cal BC); TAL: Talaiotic (c. 850- 550/500 cal BC), PST: Post-talaiotic (c. 550/500- 123 cal BC), RRE: Roman-Republican (123 cal BC-27 BC); ERE: Early Roman Empire (27 BC-284 AD); LAN: Late Antiquity and Dark Age (425-903 AD); AND: Andalusian period (903-1287 AD); LLOCS: Llocs from 19th century AD.

	Mean Distance to Llocs
NAV	1641.66
TAL	915.13
PST	761.25
RRE	819.23
ERE	1444.08
LAN	1152.23
AND	1354.72
LLOCS	0.97

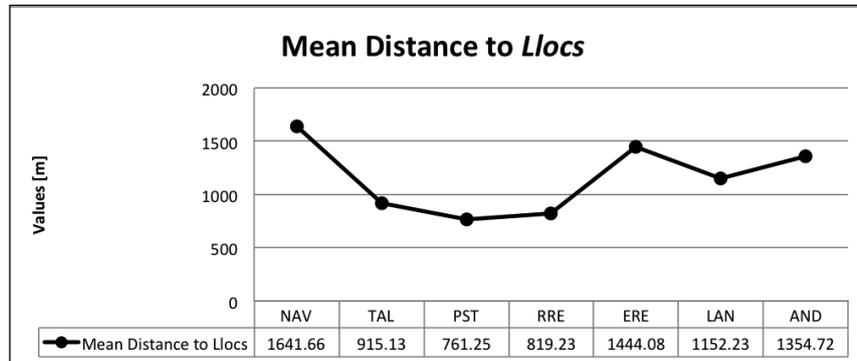


Figure 9. Mean distance values to Llocs of 19th century AD from Prehistoric settlements. NAV: Naviform (c. 1600-1100/1000 cal BC); TAL: Talaiotic (c. 850- 550/500 cal BC), PST: Post-talaiotic (c. 550/500- 123 cal BC), RRE: Roman-Republican (123 cal BC-27 BC); ERE: Early Roman Empire (27 BC-284 AD); LAN: Late Antiquity and Dark Age (425-903 AD); AND: Andalusian period (903-1287 AD); and LLOCS: Llocs from 19th century AD.

to intense Eolic erosive processes (Bourrouilh 1983: 359). Its gradual abandonment since the Early Roman Empire is particularly interesting and implies that this environmental degradation is a product of the last two thousand years, whose causes need to be further explored. In order to prevent further soil degradation and to preserve the extraordinary richness of Menorcan heritage, this clearly vulnerable area of the island needs special protection.

Our analysis of the archaeological record shows a non-linear change in human occupation from prehistoric times to the 19th century AD, with periods of settlement increase and moments of apparent depopulation. This phenomenon can be associated with the concept of “social insularity”

(Lull et al. 2002: 119). According to this theory, the Balearic Islands underwent intense immigration waves and interaction with the mainland continent that suddenly triggered very dense settlement patterns in conjunction with phases of significant decrease in the number of archaeological sites. The causes of these demographic changes can be associated with the changing situation of the mainland continent, the development of the island itself, and the relationship between both factors. For example, the first immigrants who occupied the island produced a significant demographic increase during the transition from the Early to Middle Bronze Age (17th-16th century BC) and during the Final Bronze Age (12th-9th centuries BC). The first significant break in this population development occurred at the beginning of the Roman Empire.

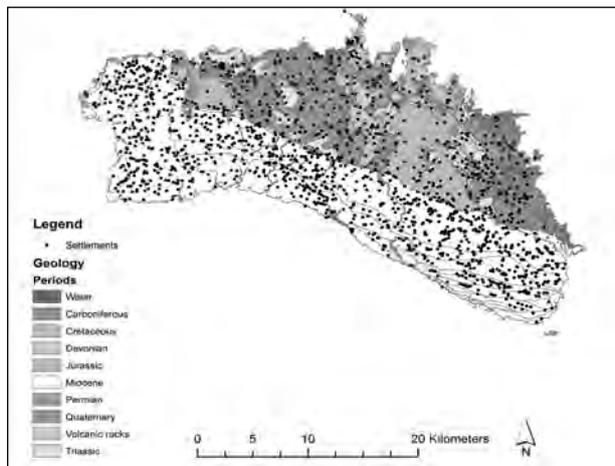


Figure 10. Geological periods and settlement records from prehistory to the 19th century AD. Map obtained from geological layer, courtesy of Ordenació del Territori Department - CIM Menorca, Spain. Software: ©ArcGIS 10.0.

The settlement decrease can be closely related to the nucleation of population in centres and the decline in agricultural production. After the Late Roman Empire, when the countryside seems to have been sparsely populated, in Late Antiquity the settlement density rapidly recovered. The highest settlement density corresponds to the *Llocs* habitation pattern of the 19th century AD. The calculated values suggest a similarly dense scenario to that during the Post-talaiotic occupation, especially if we consider that not all prehistoric sites will have been identified archaeologically.

4.2 General settlement pattern

The obtained output from Nearest Neighbor analysis (NN) indicates the possible existence of closely nucleated settlements throughout all periods, since all the NN values are quite similar and within a narrow range of 0.78-0.96 (Fig. 8). This continuity of settlement distribution structure is also supported by the geographical proximity between the prehistoric sites and historic locations. In fact, the mean distance output shows striking location proximity between prehistoric settlement sites, in particular the Post-talaiotic period, and the domestic record from the 19th century AD *Llocs* (Fig. 9).

Furthermore, the correlation of the geological map and the domestic records show a similarity

in location preferences in relation to geology throughout the last 3600 years (Fig. 10). As visible from the Chi-square test results (Fig. 11), a continued preference for occupying the Miocene platform is observed for all the periods. However, this tendency is significantly more marked in the time span from the Naviform period to the Early Roman Empire (p-value: 0.0280). On the other hand, the 19th century AD *Llocs*, although still located preferentially on the Tertiary terrains, occupy the Tramuntana region more intensively (p-value: 0.0001). Late Antiquity and Andalusian occupations show an intermediate position between the prehistoric and the modern settlement pattern (Fig. 12).

Overall, these results seem to reveal an apparent similarity in settlement pattern that continued from the Naviform period up to the 19th century AD. In fact, the long-term common thread between different chronologies is the continuity of a clustered settlement pattern correlated with similar NN values and settlement location preferences associated with an intensive occupation of the Miocene platform in the *Migjorn* area. This last evidence is particularly interesting since it implies possible long-term factors in the location strategies. Furthermore, of considerable importance is the more intense occupation of the area of Tramuntana in the 19th century AD. This fact can probably be related to the construction of drainage systems in order to expand the agricultural potential in the northern part of the island. Wetlands seem to have been a common feature in these valleys due to impermeable soils and geology (Casanovas et al. 1993: 44-45), and caused conditions of scant salubrity for human settlement (Vidal 1969: 9). Consequently, further analyses are needed to better explore how other environmental variables such as soil, land-use, hydrology, and topography conditioned the settlement structure. Furthermore, as seen in the mean distance values, the 19th century AD *Llocs* and the prehistoric sites are strikingly the closest neighbours, suggesting that similar criteria in deciding where to settle prevailed through time.

5. Conclusions

Given the quantity of spatial information, the island of Menorca offers unique archaeological conditions for an analysis of socio-environmental interaction during the last four thousand years in a

Naviform to Early Roman Empire						
Geology	Hectares	% Hectares	Observed	Expected	Chi ² values	p-values
MIOCENE	36038.6	51.9%	986	670.6	148.4	0.0000
OTHER	33449.9	48.1%	307	622.4	159.8	0.0000
Total	69488.5	100%	1293	1293	308.2	0.0000
Late Antiquity						
Geology	Hectares	% Hectares	Observed	Expected	Chi ² values	p-values
MIOCENE	36038.6	51.9%	105	79.3	8.3	0.0040
OTHER	33449.9	48.1%	48	73.7	8.9	0.0028
Total	69488.5	100%	153	153	17.2	0.0068
Andalusian						
Geology	Hectares	% Hectares	Observed	Expected	Chi ² values	p-values
MIOCENE	36038.6	51.9%	87	63.3	8.9	0.0029
OTHER	33449.9	48.1%	35	58.7	9.6	0.0020
Total	69488.5	100%	122	122	18.5	0.0049
Llocs						
Geology	Hectares	% Hectares	Observed	Expected	Chi ² values	p-values
MIOCENE	36038.6	51.9%	611	491.7	29.0	0.0000
OTHER	33449.9	48.1%	337	456.3	31.2	0.0000
Total	69488.5	100%	948	948	60.2	0.0000

Figure 11. Left. Chi-square test results table. Time periods: Naviform to Early Roman Empire (c. 1600 cal BC- 284 AD); Late Antiquity and Dark Age (425-903 AD); Andalusian (903-1287 AD); Llocs from 19th century AD.

Figure 12. Bottom. Chi-square test results table. Time periods: Naviform to Early Roman Empire (c. 1600 cal BC- 284 AD); Late Antiquity and Dark Age (425-903 AD); Andalusian (903-1287 AD); Llocs from 19th century AD.

Sites on Miocene geology						
Time periods	Total sites periods	% Total sites periods	Observed	Expected	Chi ² values	p-values
Naviform to Early Roman Empire	1293	51.4%	986	919.4	4.8	0.0280
Late Antiquity	153	6.1%	105	108.8	0.1	0.7163
Andalusian	122	4.8%	87	86.7	0.0	0.9784
Llocs	948	37.7%	611	674.1	5.9	0.0151
Total	2516	100%	1789	1789	10.9	1.7378
Sites not on Miocene geology						
Time periods	Total sites periods	% Total sites periods	Observed	Expected	Chi ² values	p-values
Naviform to Early Roman Empire	1293	51.4%	307	373.6	11.9	0.0006
Late Antiquity	153	6.1%	48	44.2	0.3	0.5686
Andalusian	122	4.8%	35	35.3	0.0	0.9661
Llocs	948	37.7%	337	273.9	14.52	0.0001
Total	2516	100%	727	727	26.7	1.5354

ca. 700 km² region of the Mediterranean. The results obtained through this preliminary exploratory analysis have revealed a noteworthy pattern of occupational behaviour in a long-term perspective. Specifically, our results highlight:

Fluctuation in the density of the archaeological domestic sites from prehistory to the 19th century AD. This non-linear development is accompanied by a gradual shift in the main occupation density from west to east across the island, possibly as a consequence of the growing importance of the harbour and city of Maó in historical times. Striking similarities are observed in the general settlement pattern. The values reflect similar Nearest Neighbor ratios over time and clear location proximity between the prehistoric and 19th century AD *Llocs* sites. Preliminary tendencies in settlement location. The correlation of prehistoric and historic domestic site records with the geological map shows a continued and significant preference for occupying the Miocene platform, in the southern *Migjorn* area of Menorca. The northern *Tramuntana* region started to be more intensely occupied during the 19th century, probably when the wetland areas existing on this impermeable geological ground could gradually be drained.

These results clearly demonstrate a complex domestic trajectory on Menorca, where gradual changes in site density were combined with important similarities in settlement structure from prehistory to the 19th century AD. Moreover, our study supports the utility of using historical data as an interpretative framework for archaeological reconstruction. The copious data from agricultural practices and subsistence patterns from the 19th century AD can be used as a comparative framework in addressing the question of sustainability, availability, and management of resources over time. In particular, these data seem to reveal an economic continuity since prehistoric times, such as the importance of agriculture and small-scale husbandry as a basic resource for population subsistence and the usage of traditional technology with no chemical fertilizers (Bisson 1966: 37; Vidal 1969: 108-153; Casanovas 2006: 286-301). Accordingly, these data can help us to test the archaeological methodology and develop a more solid model of socio-environmental interaction.

The intense occupation of the Miocene platform foreshadows more complex location trends in relation to the environmental variables. Hence, further analyses will be conducted to assess the long-term agricultural potential and land-use on Menorca. For instance, recent studies indicate that Naviform communities were possibly composed of extended families with little or no economic differentiation (Lull et al. 1999 a; Rihuete 2003; Van Strydonck et al. 2002; Gili et al. 2006). Meanwhile, in the Talaiotic settlements, we observe a probable concentration of social forces and economic resources in larger political entities (Lull et al. 1999 a; Gili et al. 2006), and lastly, Post-talaiotic communities seem to have reoccupied previous settlements in accordance with a new socio-economic tendency (Micó 2006: 431). While future investigations are necessary to shed additional light on the palaeo-economic sphere of the communities, this study shows how geo-spatial techniques can be an informative tool for identifying the underlying natural constraints and the long term social organisation of land use and settlement in the Mediterranean islands.

Acknowledgments

We wish to thank the Consell Insular de Menorca, Departments of “Cultura, Patrimoni i Educació” and “Ordenació del territori”, Spain, for the heritage and environmental data access, and in particular thanks to Ricard Cots Torrelles and Josep Florit Gomila for their assistance and availability. For their constructive and helpful comments special thanks are furthermore due to Paula Amengual, Jordi Hernández and Cristina Rihuete Herrada of the Museu de Son Fornés (Montuïri, Mallorca) and to Philip Verhagen (VU, University of Amsterdam). Finally, the authors are extremely grateful to Esther Lee (GS “Human Development in Landscapes”, Kiel) and Nicole Taylor (GS “Human Development in Landscapes”, Kiel) for their assistance in editing the manuscript. This investigation forms part of the research on socio-economic and environmental change in the Mediterranean, supported by the Spanish Ministry of Science and Innovation (HUM2006-04610, AIB2010DE-00200) and the Research Board of the Generalitat de Catalunya (2009SGR778).

References

- Alcover, J.A., D. Ramis, J. Coll, and M. Trías. 2001. "Bases per al coneixement del contacte entre els primers colonitzadors humans i la naturalesa de les Balears." *Endins* 24: 5-57.
- Aramburu Zabala, J. 2002. "Análisis comparativo del patrón de asentamiento de la cultura Baleárica." In *World Islands in Prehistory. International Insular Investigations Series* edited by W. Waldren and J. A. Ensenyat, 517-572. Oxford: Archaeopress.
- Beardah, C. 1999. "Uses of multivariate kernel density estimates." In *Archaeology in the Age of the Internet: CAA 97*, edited by L. Dingwall, S. Exon, V. Gaffney, S. Laflin and M. van Leusen, 107. Oxford: Archaeopress.
- Beardah, C., and M. Baxter. 1996. "The archaeological use of Kernel Density Estimates." *Internet Archaeology*, 1. <http://intarch.ac.uk/journal/issue1/beardah/toc.html>.
- Benejam, G.J. 1993. *El poblament de Menorca: de la prehistòria a la baixa romanitat. Aproximació a una proposta d'anàlisi de distribució espacial*. Maó: Museu de Menorca.
- Bisson, J. 1966/ 2007. *La tierra y el hombre en Menorca*. Maó: Institut Menorquí d'Estudis.
- Borrouhil, R. 1983. *Estratigrafía, sedimentología y tectónica de la isla de Menorca y del noreste de Mallorca (Balears): la terminación nororiental de las cordilleras béticas en el mediterráneo occidental*. Madrid : Servicio de Publicaciones, Ministerio de Industria y Energía.
- Cardona, L., F. Juaneda, and F. de Pablo. 1999. *Guia dels Espais Naturals de Menorca*. Maó: Institut Menorquí d'Estudis.
- Cartailhac, E. 1892/1991. *Los monumentos primitivos de las Islas Baleares*. Facsimile edition from 1892 original text published in Toulouse by Édouard Privat. Palma de Mallorca: Govern Balear.
- Casanovas, M.A. 2006. *Historia Econòmica de Menorca. La transformació d'una economia insular (1300-2000)*. Palma de Mallorca: Editorial Moll.
- Casanovas, M.A., and M. del Pilar Florit. 1993. *Una societat pagesa en una època de transició*. Ajuntament des Mercadal: Institut Menorquí d'Estudis.
- Conolly, J., and M. Lake. 2006. *GIS in Archaeology*. Cambridge: Cambridge University Press.
- Didac, F. 1997. *Geografia de les Illes Balears*. Palma de Mallorca: Editorial Lleonard Muntaner.
- Esri. 2010a. "Kernel density (Spatial Analyst)." Last modified June 2011. http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/Kernel_Density/009z0000000s000000.
- Esri. 2010b. "Near (Analysis)." Last modified April 2012. <http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/00080000001q000000>.
- Esri. 2010c. "Average Nearest Neighbor (Spatial Statistics)." Last modified March 2012. <http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/005p00000008000000>.
- Esri. 2010d. "Intersect (Analysis)." Last modified April 2012. <http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/00080000000p000000>.
- Fornós, J.J., A. Obrador, V.M. Rosselló. eds. 2004. *Historia Natural del Migjorn de Menorca: el medi físic i l'influx humà*. Palma de Mallorca: Fundació Sa Nostra.
- García-Argüelles, A., J. Gual, and A. López. 1994. "Aproximació a l'ocupació del territori a l'Antiguitat: el terme municipal d'Alaior." *Meloussa* 3: 9-24.
- Gili, S., V. Lull, R. Micó, C. Rihuete, and R. Risch. 2006. "An island decides: Megalithic Burial Sites on Menorca." *Antiquity* 80: 829-942.
- Gornés, S., J. Gual, A. López, J. de Nicolás, and A. Roca. 2004. "L'assentament humà des de la prehistòria fins al baix Imperi". In *Historia Natural del Migjorn de Menorca: el medi físic i l'influx humà* edited by J. Fornós, A. Obrador, and V. M. Rosselló, 328- 350. Palma de Mallorca: Institut Menorquí d'Estudis Fundació Sa Nostra.
- Guerrero, V.M., and S. Gornés. eds. 2000. *Colonització humana en ambients insulars. Interacció amb el medi i adaptació cultural*. Palma: Ed. UIB.
- Hodder, I., and C. Orton. 1976. *Spatial Analysis in*

- Archaeology. Cambridge: Cambridge University Press.
- Ives, J.W. 1985. *A Spatial Analysis of Artifact Distribution on a Boreal Forest Archaeological Site*. Edmonton: Alberta Culture Historical Resources Division.
- Lock, G. ed. 2000. *Beyond the Map: Archaeology and Spatial Technologies*. Amsterdam: IOS Press.
- Lock, G., and S. Zoran. eds. 1995. *Archaeology and Geographical Information Systems: A European Perspective*. London: Taylor and Francis.
- Lull, V., S. Gili, R. Micó, C. Rihuete, and R. Risch. eds. 1999a. *La cova de Es Càrritx y La cova de Es Mussol. Ideología y sociedad en la prehistoria de Menorca*. Barcelona: Consell Insular de Menorca.
- Lull, V., R. Micó, C. Rihuete, and R. Risch. eds. 1999b. *Rituales de vida y muerte en la prehistoria de Menorca. La Cova des Càrritx*. Barcelona: Consell Insular de Menorca.
- Lull, V., R. Micó, C. Rihuete, and R. Risch. eds. 1999c. *La Cova des Mussol, un lugar de culto en la Menorca prehistórica*. Barcelona: Consell Insular de Menorca.
- Lull, V., S. Gili, R. Micó, C. Rihuete, and R. Risch. 2002. "Social and ideological changes in the Balearic Islands during the Later Prehistory." In *World Islands in Prehistory. International Insular Investigations Series* edited by W. Waldren and J. A. Ensenyat, 117-26. Oxford: Archaeopress.
- Lull, V., R. Micó, C. Rihuete, and R. Risch. 2005. "Los cambios sociales en las islas baleares a lo largo del II milenio." *Cypsela* 15: 123-148.
- Lull, V., R. Micó, B. Palomar, C. Rihuete, and R. Risch. eds. 2008. *Cerámica talayótica. La producción alfarera mallorquina entre ca. 900 y 550 a.n.e.* Barcelona: Edicions Bellaterra.
- Menorca. 2012. "Fotos de Menorca." Last modified: May 2012. <http://www.menorca.ws/>.
- Micó, R. 2005a. "Towards a definition of politico-ideological practices in the Prehistory of Menorca (the Balearics Islands): The wooden carvings from Mussol cave." *Journal of Social Archaeology*, 5(2): 276-299.
- Micó, R. 2005b. *Cronología Absoluta y periodization de la Prehistoria de las Islas Baleares*. Oxford: Archaeopress.
- Micó, R. 2006. "Radiocarbon dating and Balearic prehistory: Reviewing the Periodization of the Prehistoric Sequence." *Radiocarbon*, 48 (3): 421-434.
- Plantalamor, L. 1991. *L'Arquitectura prehistòrica i protohistòrica de Menorca i el seu marc cultural*. Maó: Museu de Menorca.
- Plantalamor, L. and M.C. Rita. 1979. "Arqueologia prehistòrica i protohistòrica." In *Enciclopèdia de Menorca* 8: 1-193. Maó: Obra Cultural de Menorca.
- Ramis, J. 1818. *Antigüedades célticas de la Isla de Menorca desde los tiempos más remotos hasta el siglo V de la Era Cristiana*. Maó: Imprenta de Pedro Antonio Serra.
- Rihuete, C. 2003. *Bioarqueología de las prácticas funerarias: Análisis de la comunidad enterrada en el cementerio prehistórico de la Cova des Càrritx (Ciutadella, Menorca), ca. 1450-800 cal ANE*. Oxford: Archaeopress.
- Van Strydonck, M., M. Boudin, and A. Eryvynck. 2002. "Stable Isotopes (¹³C and ¹⁵N) and Diet: Animal and Human Bone Collagen from Prehistoric Sites on Mallorca and Menorca (Balearic Islands, Spain)." In *World Islands in Prehistory. International Insular Investigations Series* edited by W. Waldren and J. A. Ensenyat, 189-97. Oxford: Archaeopress.
- Verhagen, P., S. Gili, R. Micó, and R. Risch. 2007. "Modelling Prehistoric Land Use Distribution in the Rio Aguas Vally (SE Spain)." In *Case Studies in Archaeological Predictive Modelling*, edited by C. Bakels and H. Kamermans, 171-191. Leiden: Leiden University Press.
- Vidal B.T. 1969. *Evolución de la agricultura y de la propiedad rural en la isla de Menorca*. Revista de Menorca 21 (2): 5-46.
- Wheatley, D., and M. Gillings. 2002. *Spatial Technology and Archaeology: the archaeological applications of GIS*. London: Taylor and Francis.

A New Method of Spatial Analysis Based on the Extraction of Proximity Graphs

Diego Jiménez-Badillo

INAH, Museo de Templo Mayor, Mexico

Abstract:

This paper introduces the Relative Neighbourhood (RN) method of spatial analysis, a series of procedures oriented to extract relational networks from point sets, which can be used to discover meaningful association patterns among archaeological artefacts or sites. The underpinning elements of such method are the relative neighbourhood concept, the retrieval of proximity graphs, the measurement of graph-theoretic relational properties both at global and local levels, and the visualisation of spatially-based connectivity patterns. The method is particularly useful to analyse a special type of archaeological deposits, which we refer to as “spatially symbolic contexts”. However, it is important to highlight that the method is generic and can be applied to many different kinds of archaeological contexts or even as part of studies focused on exploring relations between archaeological sites.

Keywords:

Proximity Graphs, Relative Neighbourhood, Spatial Analysis

1. Introduction

This paper introduces the Relative Neighbourhood (RN) method of spatial analysis, a series of procedures oriented to extract relational networks from point sets, which can be used to discover meaningful association patterns among archaeological artefacts or sites.

We illustrate the application of the RN method using a group of Aztec offerings, an interesting case of spatially symbolic contexts. These comprise groups of objects whose arrangement and meaning are determined by symbolic ideas expressed through the spatial ordering of different categories of artefacts. In such deposits, the specific location of each object -as well as its spatial association with the rest of the assemblage- creates characteristic observable ‘patterns’. In studying contexts like these, analysts seek to identify meaningful arrangements from the spatial links of the artefacts. The following section describes these deposits, while the remaining sections explain the concept of relative neighbourhood and the procedures to analyse global and local spatial features.

2. The Aztec Offerings as a Study Case for Spatial Analysis

For more than a hundred years Mexican
Corresponding author: diego.jimenez61@gmail.com

archaeologists have discovered numerous caches of artefacts among the remains of the Sacred Precinct of Tenochtitlan, the main ceremonial centre of the Aztec people (Matos Moctezuma 1979, 1982, 1986, 1988). Nearly 200 deposits have been found so far (2012) within the boundaries of the archaeological site and some more in the surroundings.

The caches provide valuable evidence of offering rituals performed to request favours or to pay tribute to some of the many deities of the Mexica pantheon (López Luján 1994: 48, 52). Therefore, they are the key to understand important aspects of this religion.

Although some caches are very simple most of them are rather complex, both in terms of the variety and the quantity of artefacts deposited. Indeed, some contexts include hundreds or even thousands of items. Among the most noteworthy contents are figures of deities, as well as sceptres, pectorals, masks, ear-plugs, small-scale models of musical instruments, sacrificial tools, etc. Some offerings also include remains of beheaded people, as well as flora and fauna remains. Among the animals it is worth mentioning sawfish snouts, crocodiles, jaguars, lynxes, serpents and a great diversity of birds, fishes and marine shells (Figs 1, 2, and 3).

The selection of objects for a particular oblation responded to specific goals. For instance, if

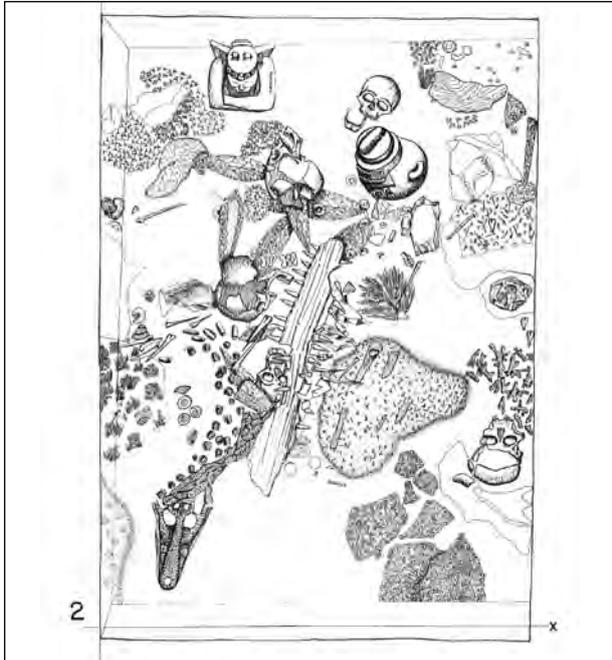


Figure 1. One of the offerings recovered from the remains of the Sacred Precinct of Tenochtitlan (Offering 60).



Figure 3. Some objects recovered in the offerings of the Sacred Precinct of Tenochtitlan.



Figure 2. Offering H, recovered from the remains of the Sacred Precinct of Tenochtitlan.

the purpose of the offering was to propitiate timely rains, then the priests selected objects that according to Aztec religion were suitable for receiving forces from the gods of fertility. Chosen artefacts may include sculptures resembling the gods themselves, their insignia, or items symbolizing the region of the cosmos where such deities were believed to inhabit. Such an offering may also include items intended to keep other forces, like drought and death, under control (López Luján 1994).

2.1 The spatial dimension of the offerings

Aztec religion revolved around the concept of complementary opposition of divine energies and the offerings are examples of a system of relations governed by combinatorial rules that were expressed onto space. In Aztec world-view a rule dictated that those spaces with an excess of “cold” influences must be balanced with objects symbolically associated with “hot” energies. Another rule indicated that “feminine” artefacts ought to be balanced with “masculine” items; sculptures of the god of rain and fertility usually were placed opposite to figures of the god of fire, etc.

Those rules guided not only the kind of artefacts chosen for each particular offering, but more importantly they prescribe the spatial arrangement of the items. This explains why specific categories of artefacts are found facing cardinal directions; below or above certain boundaries; next to or in opposition to other objects or features, and so on (López Luján 1994).

The need to revealing the structure of these archaeological contexts led us to developing a method that explores the network of spatial relationships between the offering items in order to discover whether specific classes of artefacts are recurrently located in proximity or in opposition to each other. If some patterns are discovered, then it would be easier to determine their meaning.

When looking for a solution for that challenge we came across with an interesting concept of adjacency known as relative neighbourhood. This allows exploring different degrees of association between artefacts and provides the foundation for the method described below (Jiménez-Badillo 2004).

3. The Concept of Relative Neighbourhood

Relative neighbourhood is an interesting notion that formalises mathematically the idea of two objects being “more or less” adjacent and therefore associated up to a certain measurable degree (Lankford 1969; Toussaint 1980a, 1980b, 1980c, 1988; Gabriel and Sokal 1969; Matula and Sokal 1980; Kirkpatrick and Radke 1985; Urquhart 1980, 1982, 1983, Adamatzky 2010).

In earlier publications we detailed its use in archaeology for discovering contextual relationships or for finding different levels of connection between sites or artefacts (Jiménez and Chapman 2002; Jiménez-Badillo 2004, 2006, 2009a, 2009b, 2012).

Analysing relative neighbourhoods is better for exploratory analysis than using absolute concepts such as nearest neighbour.

Imagine a point set $P = \{p_1, p_2, p_3 \dots p_n\}$ and take any two elements, say p_i and p_j . We say that p_i and p_j are *relative neighbours* if, and only if, both are at least as near to each other as they are to the rest of the points. This is expressed mathematically as follows:

$$d(p_i, p_j) \leq \min[\max\{d(p_i, p_k), d(p_j, p_k)\}] \text{ for } k = 1, \dots, n, k \neq i, j \text{ and } p_k \in P$$

The equation implies delimiting a region of influence for a pair of points and then checking if such region is empty (i.e. if no other point is located inside its boundaries). If that happens we declared those two points as relative neighbours drawing a link

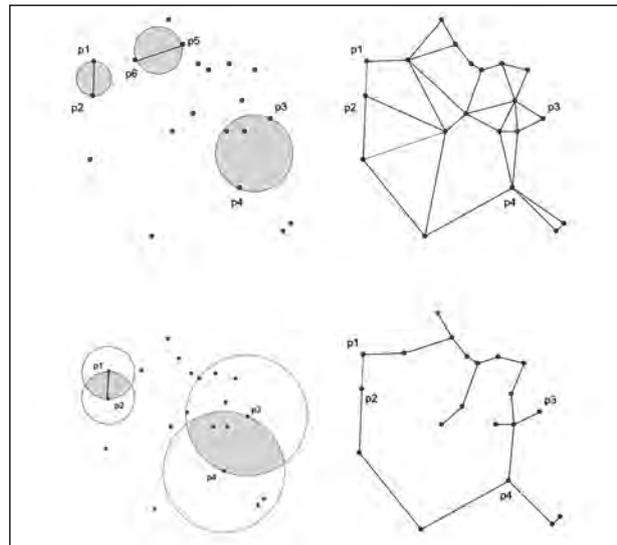


Figure 4. Test of “region emptiness” based on drawing circles (above) and lune-shaped regions (below). In this example, the neighbourhood of the points (p_1, p_2) is empty. Therefore, they are declared relative neighbours and a link is drawn between them. On the contrary, the region corresponding to pair (p_3, p_4) is not empty and the points are declared no-neighbours.

between them. On the contrary, if the region contains any other point of the dataset, the relationship is declared null and no link is drawn.

Applying such procedure to every pairwise permutation produces a relational network. Interestingly, the shape and connectivity of the network depend on the spatial configuration (i.e. topology) of the whole point set, and therefore corresponds to its morphological structure.

The region of influence can be defined as a circle whose perimeter passes through p_i and p_j or as the intersection of two circles of radius $d(p_i, p_j)$ centred at p_i and p_j respectively (i.e. *versica piscis*). In the first case the resulting network is called Gabriel Graph (GG) while the second alternative produces the Relative Neighbourhood Graph (RNG) (Fig. 4).

For exploratory reasons we can enlarge or reduce voluntarily the region of influence using a parameter beta (β) (Kirkpatrick and Radke 1985). High values of beta produce coarser networks while low values of beta produce networks with higher edge-density. Testing several values of beta produces

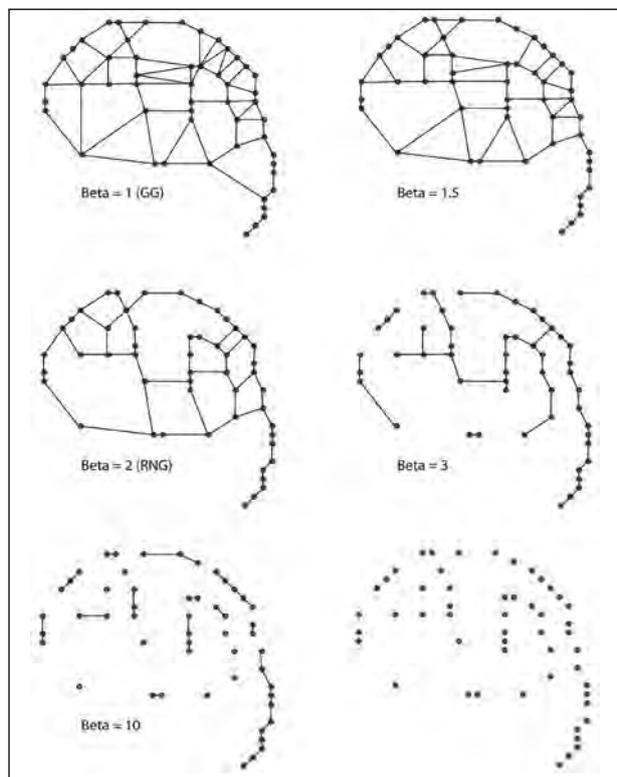


Figure 5. A family of Beta-skeletons resulting when the region of influence changes due to the introduction of a parameter β . Notice the variation in the number of connections detected with different beta values.

a parameterized family of networks known as Beta-skeletons (Fig. 5) (see more details in Jiménez-Badillo 2004: 96-145, 2012).

Extracting networks using consecutive values of beta constitutes a great exploratory procedure to analyse the spatial structure of archaeological contexts, mainly because interesting details are often exposed through the connectivity changes between successive Beta-skeletons. Algorithms to compute these graphs have been developed by Huang (1990), Hurtado et al. (2001), Jaromczyk and Toussaint (1992), Jaromczyk and Kowaluk (1987, 1991), Rao (1998), Su and Chang (1990, 1991a, 1991b), Toussaint and Menard (1980).

4. Analysing the Global Structure of Aztec Offerings

One way to analyse spatially symbolic contexts like the offerings consists in comparing their spatial structure at a global level. All artefacts involved

are recorded with a point, which corresponds to its center of gravity. We find that differences in areal extend of the objects do not affect the analysis considerably. This is due to the fact that the neighbourhood relations depend on the relative position of all the artefacts in the set and not on the absolute distance between two objects. In contrast, when linear items are present (such as the sawfish snouts found in offerings 22 and 58), the convention is to record the objects with several points marked at fixed intervals.

We start extracting a family of Beta-skeletons for each offering under study using a value of $\beta = 1$. Then, we continue incrementing beta by one-tenth steps, that is 1.1, 1.2, 1.3, etc. It is worth noting that the skeleton obtained with $\beta = 1$ corresponds to the Gabriel Graph, while $\beta = 2$ correspond to the Relative Neighbourhood Graph.

We propose using the Gabriel Graph as a benchmark for comparing spatial structure of different point sets because it has some important geometric properties, namely:

The Gabriel Graph is always connected and contains the largest number of edges in a family of Beta-skeletons. Therefore, it can be considered as an upper connectivity threshold (UCT) for any spatial distribution of nodes.

The Gabriel Graph is unique, which means that for any specific spatial distribution of nodes there is one and only one Gabriel Graph. In practice this means that similar node spatial distributions tend to have similar GGs, and vice versa, very different spatial distributions of nodes tend to produce very different graphs. Therefore, we may think of GG as the signature of a point set.

When the nodes under study are located in 2D space, the Gabriel Graph is always a connected planar graph. As we explain in the following section this gives us the possibility of quantifying connectivity properties by using standard measures of Graph Theory. In a previous paper we explain the extraction of two connectivity measures called control and integration (Jiménez-Badillo 2012), while in this paper we extend the method with additional measures.

Another interesting value of beta is what we called the Lower Connectivity Threshold (LCT). This is defined as the highest value of beta producing a connected Beta-skeleton with the lowest number of edges; in other words the highest beta value before the graph gets split into two or more sub-graphs. In any spatial distribution of nodes such value is greater than 2 because a graph extracted with $\beta = 2$ corresponds to the RNG and this is always a connected graph (see more details in Jiménez-Badillo 2004: 166-172).

4.1 Calculating some measures of global spatial structure

The next step is to measure global spatial structure in the Beta-skeletons extracted previously. The most basic measure is the edge to vertex ratio (EVR), which yields edge-density; then we can observe other characteristics such as the number of planar faces (if the nodes are in 2D space) or the number of induced cycles in 3D space.

Another, more interesting measure is the so-called planar network structure (Gabriel and Sokal 1969; Matula and Sokal 1980). This is applicable to connected Beta-skeletons (i.e. those that are not split into two or more subgraphs), which include Gabriel and Relative Neighbourhood graphs. Connected graphs are those that contain a path from each vertex to any other vertex. The measure exploits the additional fact that connected Beta-skeletons derived from a 2D point set are always planar graphs. Planar graphs are those that can be drawn in such a way that no edges cross each other. They are important because they have very well known connectivity properties given by Euler's theorem, expressed in the following formula (Haggett and Chorley 1969):

$$\text{Number of vertices} - \text{number of edges} + \text{number of faces} = 2$$

Euler's formula predicts how many connections can be found in a fully connected planar graph and therefore can be used as a benchmark to compare the connectivity of a real adjacency graph with its fully-connected counterpart.

Planar Network Structure is given by:

$$\alpha(G) = \frac{\text{number of edges} - \text{number of nodes} + 1}{2(\text{number of nodes}) - 5}$$

As Matula and Sokal (1980) rightly notes, the above measure corresponds to the actual number of interior faces in graph G divided by the maximum number of interior faces in an ideal planar connected graph on as many vertices. Thus, they call it planar face density and developed the following formula more appropriate for GGs and RNGs (i.e skeletons for $\beta = 1$ and $\beta = 2$, respectively):

$$\alpha(G) = \frac{\text{number of interior faces in a planar connected proximity graph on } n \text{ nodes}}{\text{maximum number of interior faces in any planar graph on } n \text{ nodes}}$$

4.2 Producing profiles of spatial structure for Inter-offering comparisons

Either one of the above measures can be used to produce what we call graph connectivity profiles. These are used to compare the morphology of the offerings through changes in the value of beta. Those that have similar spatial layout will have similar profiles. When this happens we could assume that they belong to the same class of caches and therefore we may analyse them together when looking for local connectivity patterns.

The shape of the curve in each profile is important. In some cases, tiny increases of beta produce significant drops in the number of edges. In other situations the numbers of connections may not change dramatically even with strong changes of beta. All this depends on the morphology and relational structure of the point sets under study.

The meaning of these variations depends on each application. In the case of the offerings, it helps us identify classes of offerings with similar spatial layout. Indeed, through this procedure we have identified several offerings that contain similar categories of artefacts distributed almost in the same way, which suggest a common meaning.

Comparing several offering profiles in the same chart would help establish similarities and differences between several archaeological contexts. As an example, we illustrate the edge-to-vertex (EVR) profiles of three caches known as offerings 22, 58 and U (Figs 6, 7, and 8).

Figure 9 exhibits major differences in the spatial structure of offering U and offering 22. Both curves follow dissimilar paths and present a very clear intersection caused by a strong contrast of

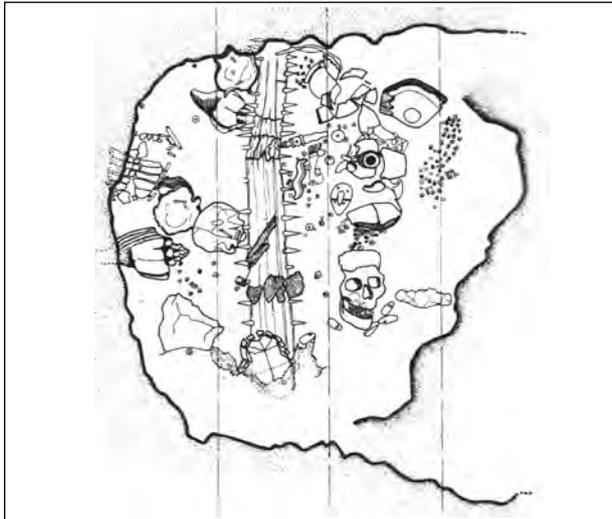


Figure 7. Excavation drawing of Offering 58 (level 4). It features (a) sawfish; (b) two human skulls; (c) a broken pot with the effigy of a fertility goddess; (d) three scepters (i.e. xicahuaztli, serpentiform, and deer-head); (e) Xancus shell; (f) turtle shells; and (g) mother-of-pearl shells.

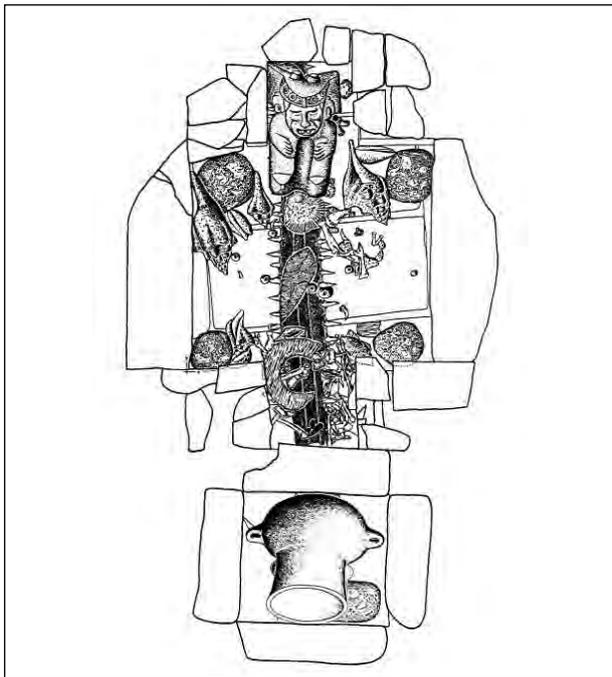


Figure 8. Excavation drawing of offering U. It features a sawfish, four copal balls, several marine shells, eagle and lynx remains, a pot with portrait of Tlaloc (god of fertility), a sculpture of Xiuhtecuhtli (god of fire), a sacrificial knife, some greenstone beads, etc. (Drawing by Fernando Carrisoza, Museo del Templo Mayor, INAH, Mexico).

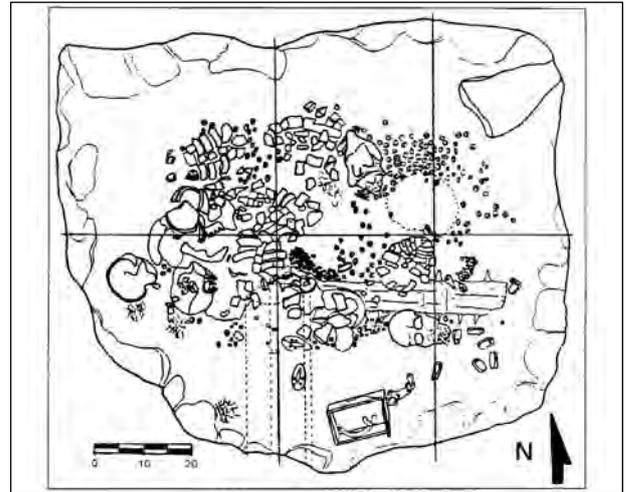


Figure 6. Excavation drawing of Offering 22.

slope and differences in length. In order to highlight such difference we overlaid tendency lines (the bold lines) using a standard polynomial formulae. This simplifies the general shape of the curves facilitating comparison procedures. Notice, for example, that despite having almost the same number of objects, offering U starts with an EVR higher than 2.0, while its counterpart hardly reaches 1.7. On the other side, offering 22 is longer but finishes with almost as fewer edges than offering U.

In contrast, the comparison of offering 22 and 58 yield some important similarities (Fig. 10). The most remarkable is that they appear closer to each other in both shape and length. Both curves in fact follow mostly a parallel trajectory. There is only a minor difference between the two. This is revealed by the intersection of both curves at $B = 1.5$. However, if we displace one of the profiles to the right it is notable that both curves coincide almost entirely. What is the significance of these observations? Most of all, they provide a gross indication of global structural similarities and dissimilarities among the three point sets under consideration. Concordances are stronger between offerings 22 and 58, whereas offering U represents a departure from both of them. This latter cache, for instance, shows a higher degree of linearity in the middle of the assemblage as well as certain concentration of items in its corners. In contrast, the other two caches seem to show only one aggregate in the lower portion, while the alignment is not as strong as in offering U. This is not surprising given the fact that offerings 22 and 58 share almost the same type of objects (see

Table 1). Whether or not those variations actually reflect different meanings of the caches is a question that the RN method would help to answer with additional measures oriented to discover patterns at a local level.

5. Analysing the Local Structure of Aztec Offerings

Another, equally important way to analyse the Aztec offerings consists in identifying spatial arrangements of artefacts that are significant at a local level. To this end, we apply a visual clustering technique. The goal is to find clusters constituted by the same combination of artefacts and to investigate whether the same arrangements are found across several caches. If certain combinations of artefacts appear repeatedly in many offerings we would be in a position to infer that such arrangements were intentionally created and therefore we can treat them as “symbolic themes”. The meaning of these patterns could finally be deciphered with the help of iconographic as well as historic and ethnographic information.

5.1 Visual clustering through the extraction of Limited Neighbourhood graphs

Local analysis is based on the extraction of a lesser known graph called Limited Neighbourhood graph. This was defined by Urquhart (1982), but unfortunately has not attracted much attention among specialists.

The LNG also attaches regions of influence to pairs of points –just like the Beta-skeletons- but adds an extra parameter sigma (σ) to transform the shape of such region. This produces two new shapes for the extraction of proximity graphs which are composed by a combination of circles and lunes, as follows:

$$R_1(p_i, p_j, \sigma) = \text{Lune}(p_i, p_j) \cup \{x: \sigma \min[d(x, p_i), d(x, p_j)] < d(p_i, p_j) \forall i \neq j\}$$

$$R_2(p_i, p_j, \sigma) = \text{G-Circle}(p_i, p_j) \cup \{x: \sigma \min[d(x, p_i), d(x, p_j)] < d(p_i, p_j) \forall i \neq j\}$$

Where Lune(p_i, p_j) is a region defined on the relative neighbourhood principle; G-Circle(p_i, p_j) is a region defined on the least square adjacency criterion or Gabriel neighbourhood; and x is the variable representing those locations in space that defined the R shape.

R_1 corresponds to the shape illustrated in Figure 11a, while R_2 corresponds to the region in Figure 11b.

The main effect of any of these two shapes is cutting the resulting graph into several components. They isolate the most obvious aggregations of nodes within a set. Therefore, LNG is a useful tool to detect local morphological traits in a heuristic way, complementing any knowledge given by the global ones viewed with graphs that are always connected, such as RNG and GG.

As sigma controls the local fragmentation of the graph, it is obvious that applying different values of sigma produces a sequence of nested clusterings (Urquhart 1982, 177).

Urquhart presents evidence that his hierarchic strategy is useful as a visual clustering approach. As he characterises it: “A visual model of the clustering problem is one in which clusters are defined (...) in a way that relates to human visual perception” (Urquhart 1982, 173).

Within this framework sigma represents a factor of *relative edge consistency*, from which the index of dissimilarity for each clustering can be easily calculated: $d^* = 1/s$.

Graphs extracted with different values of sigma would immediately fulfil three important criteria of clustering (Urquhart 1982, 175):

- **Connectivity:** all points within a cluster c_i element of $C(P)$ should be connected. This means that points belonging to the same cluster must be highly connected, while those belonging to different clusters must be separated.
- **Consistency:** for adding a new point t with a set of neighbours $N(p_i)$, if c_i is element of $C(P)$ and $N(p_i) \cap c_i$ then there exists a cluster c'_i element of $C(P \cup p_j)$ such that c'_i is identical with $(c_i \cup p_j)$ for some partition of P . That is, if a point is recognised as member of a group of neighbours at some level, that membership should remain along the whole hierarchy containing the same group of neighbours.
- **Local stability.** Inserting a point p in cluster c_i

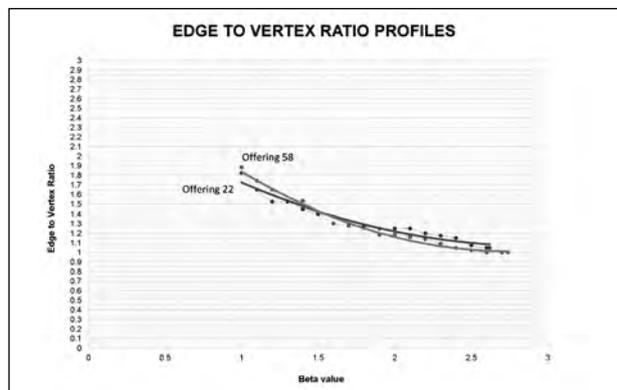


Figure 9. Comparison of edge-to-vertex ratio (EVR) profiles from offerings 22 and U.

should not affect cluster c_j , when $c_i \neq c_j$.

Consequently, the LNG is capable of detecting a much wider range of aggregation patterns than other visual approaches (Urquhart 1982, 174). In Figure 12 we illustrate the range of cluster types detected by this particular method. These include not only well-separated groups, but also aggregations that exhibit local changes in point density, those with a 'bridge' connecting two subclusters, and points having a Gaussian distribution.

The application of this clustering technique to the Aztec offerings seeks to determine whether the caches under study contain clusters constituted by the same types of artefacts. If repetitive combinations of artefacts were found, then it would be reasonable to assume that such regularities were intentional. Therefore, they could be regarded as 'symbolic themes'. The significance of those patterns could finally be interpreted with evidence provided by other sources of Aztec symbolism such as iconography, as well as historic and ethnographic information. Suppose, for example, that we identify a cluster formed by four greenstone beads located at the corners of an imaginary square, plus one more at the centre complemented by a sacrificial knife. We assume this to be significant and therefore investigate whether other offerings contain arrangements with the same combination of artefacts. If the search is positive, we refer to the pattern as a symbolic theme.

Such a pattern actually exists. It has been found in the integration core of offering U and it is repeated in two offerings numbered as 16 and 16a. Some experts -based on iconographic analysis

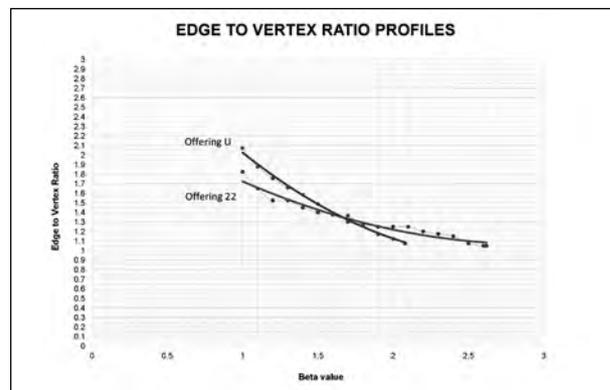


Figure 10. Comparison of edge-to-vertex ratio (EVR) profiles from offerings 22 and 58.

of codices, sculptures, ceramics, and documentary evidence-, interpret the pattern as a *quincunx*, the most schematic reference to the five major points of the Aztec universe, which includes the four cardinal points plus the centre. Furthermore, López Luján (1994, 172-192) has found recurrent evidence for the association of this arrangement with the god Xiuhtecuhtli. It is worth mentioning that sculptures of this deity have been found in several caches, including offering U itself (Fig. 8).

The procedure is perhaps the most simple to follow and it is also one of the more useful of the RN-Method:

To start, we may choose a beta value of 2. This, in addition to some sigma value, would yield clusters based on the R_1 shape defined by Urquhart (1982), illustrated in Figure 11a. Alternatively, we may start with $\beta = 1$ and any value of sigma, so the clusters would be based on the R_2 shape (illustrated in Figure 11b). A third option, which we found particularly useful in offering analysis, is to begin with a beta value that matches the lower connectivity threshold (LCT). As we defined it earlier, LCT is the highest beta value before the graph gets split into two or more sub-graphs. In the case of offering U, this corresponds to $\beta = 2.08$, while in offering 22 $\beta = 2.62$, and in offering 58 $\beta = 2.74$. Such a threshold is then combined with any sigma value.

At the same time we have to choose a value of sigma. We recommend to apply a whole series of values within the range $0 < \sigma \leq 1$. For example: $\sigma = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$. Whatever the final combination of beta and sigma values, the

	Object type	OFFERING 22 Quantity	OFFERING 58 Quantity
1	Conch shell	6	9
2	Copper bells	2	2
3	Cradle	1	1
4	Fertility goddess on pot	1	1
5	Knife for sacrifice	1	1
6	Necklace of 7 olive shells	1	1
7	Nose-plug Xipe-Totec	1	1
8	Pendant oyohualli	2	2
9	Perforated conch shell	0	2
10	Resine of copal	1	1
11	Sawfish (Pristis pectinatus)	1	1
12	Sceptre chicahuaztli	1	1
13	Sceptre deer-head	1	1
14	Sceptre serpentiform	1	2
15	Sea urchin	5	0
16	Skull	1	1
17	Skull-mask	1	1
18	Turtle shell (Pseudemys scripta)	6	7
19	Xancus shell	1	2
	Total	34	37

Table 1. Categories of artefacts found in Offering 22 and Offering 58. Notice the similarities in the quantity and type of objects, as well as the similar spatial layout (see Figures 8 and 9).

purpose is to produce a family of nested graphs, each one representing a specific level within the hierarchic clustering process. Within this framework, an index of dissimilarity can be easily measured for each edge, by taking into account how long it remains connected to a cluster. This is calculated as follows:

$$d^* = 1/s.$$

An edge that disappears from the graph at $\sigma = 0.1$, for example, would have a very high dissimilarity value; whereas an edge that remains present until σ reaches a value 0.9 would have low dissimilarity.

In fact, sigma represents a factor of *relative edge consistency*. The level of consistency refers to how essential are certain edges for the structure of the pattern. Inconsistent edges are those that appear joining two clusters that should be separated. 'Inconsistency' also applied to redundant connections, which generally are not very meaningful. It is worth mentioning that lower sigma values tend to yield more 'inconsistent' edges than higher sigma values. This means that very few clusters appear in the graph, making it difficult to visualise any significant arrangement of artefacts. In contrast, high sigma values yield only the most 'consistent' edges. These represent the strongest, more meaningful connections, but would perhaps

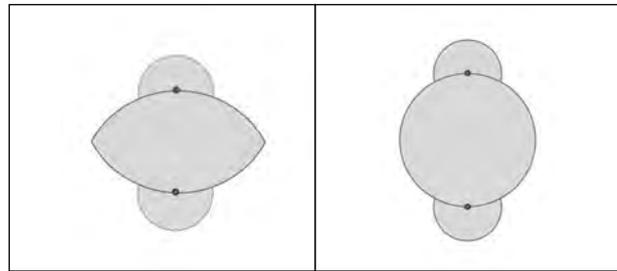


Figure 11. The regions of influence that produce Limited Neighbourhood Graphs: (a) Shape R1, a limited neighbourhood derived from the RN-lune; (b) shape R2, a limited neighbourhood derived from the Gabriel circle.

eliminate some other essential edges. Therefore, very high sigma values might be insufficient to reveal interesting patterns. Therefore, we suggest exploring each offering with different beta/sigma values until some pattern emerges. More attention should be given to arrangements perceived with sigma values equal to 0.4, 0.5 or 0.6.

For the analysis of offerings 22 and 58 we choose the lower connectivity threshold as the value of beta and keep it constant. Then, we try different values of sigma. At the end, the combination $\sigma = 0.6$ and $\beta = 2.63$ for offering 22; and $\sigma = 0.6$ and $\beta = 2.74$, for offering 58 led to the identification of the following artefact arrangements organized along the sawfish snout, all of which are present in both offerings (Figs 13 and 14):

Arrangement one (Fig. 15):

- a) Sceptre deer-head
- b) sceptre serpentiform
- c) sceptre *chicahuaztli*
- d) obsidian nose-plug.

Arrangement two (Fig. 16):

- a) Human skull
- b) Necklace of 7 *Oliva* shells
- c) Copper bells
- d) Cradle (only in offering 22)

Arrangement three (Fig. 17):

- a) Human skull
- b) Sacrificial knife

c) Two *oyohualli* pendants

Arrangement four:

- a) Several turtles
- b) Several mother-of-pearl shells
- c) *Xancus* shells
- d) Sea urchins (only in offering 22)

Arrangement five (Fig. 18):

- a) Pot with human effigy
- b) Either turtle shell or mother-of-pearl shell

Once the patterns have been found, the analysis proceeds with an examination of iconographic and documentary sources (codices, historic texts, etc.) in order to decipher their meaning. In previous works, we offered a detailed interpretation of the above arrangements highlighting the importance of the sawfish-snout as a metaphor of the earth and explaining the symbolism of each combination of artefacts (Jimenez-Badillo 2004, 223-247; 2011).

Due to limitations of space, we exemplify this step only with regard to Arrangement 2. The human skull belongs to a child who was subject to a rite of beheading. This practice was very common in sowing and harvesting rites that emulated different stages of maize growth. Those known as Ochpaniztli, Tititl, and Huey Tecuilhuitl - each lasting twenty days, were performed every year to mark different phases of the agricultural cycle and included the decapitation of teenage girls that personified earth, water and maize deities (Graulich 1999, 233-252, 379-401). Huey Tecuilhuitl was especially important because coincided with the arrival of corn in cultivated fields. At that time, a young girl was dressed with the attire of the goddess Xilonen, which included a necklace of seven gold pendants with corncob shape. We believe that the seven shells found in offerings 22 and 58 were alternative symbols of corncobs or seeds. Xilonen personified a 'tender corncob' and by dressing like the goddess the girl was supposed to acquire her germination powers, in other words she became a sacred "seed". At a certain point of the rite, the girl's heart was extracted and her throat cut with the snout of a sawfish. Only after the girl was killed the common people were allowed to harvest

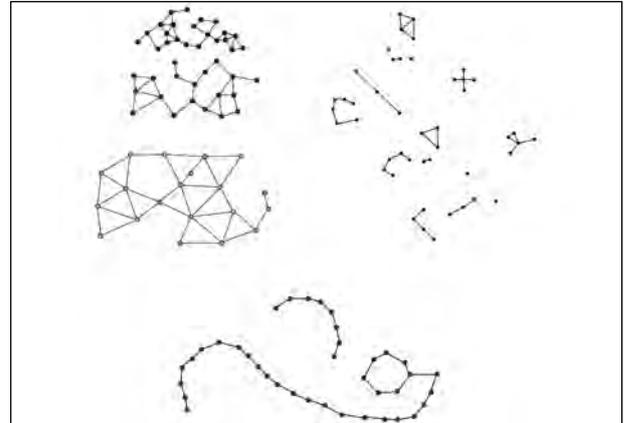


Figure 12. Different kinds of clusters that can be properly identified by applying the limited neighbourhood concept: (a) variable density; (b) Gaussian; (c) linear.

and consume earth's new produce (Durán 1967, 125-141; Sahagún 1956, 173).

The ceremony fulfilled the goddess request for devolving germination powers to earth in order to renew the agricultural cycle. This request is described in the myth of Cipactli, the earth monster that explicitly demanded sacrifices and blood offerings from mankind as restitution for the life that she would provide on earth. Interestingly, the snout of the sawfish was called *acipactli*, which translates as 'tool for sacrifice'.

It is precisely in the context of sacrifice to mother-earth goddesses where we can situate the meaning of the cradle models found in offerings 22 and 58. Durán (1967, 125-133) gives interesting clues while describing the ceremonies held during Huey Tecuilhuitl. He mentions a curious ritual performed every time that Cihuacoatl-Quilaztli have spent more than eight days without receiving human sacrifices. In those occasions, her priests put a sacrificial knife within a cradle and had a woman carried it to the market. There, she requested a saleswoman to look after the cradle promising to return soon. After several hours of waiting, the saleswoman discovered that the cradle contained only a knife and from this she deduced that the goddess had visited the market to show her hunger and to reprimand human beings for the lack of sacrifice. The concept of sacrifice with the purpose to obtain new life was so embedded in Aztec mentality, that the whole existence was conceived as a cyclic process of life-death-life (i.e. as opposed to a linear one). In this mindset, the idea

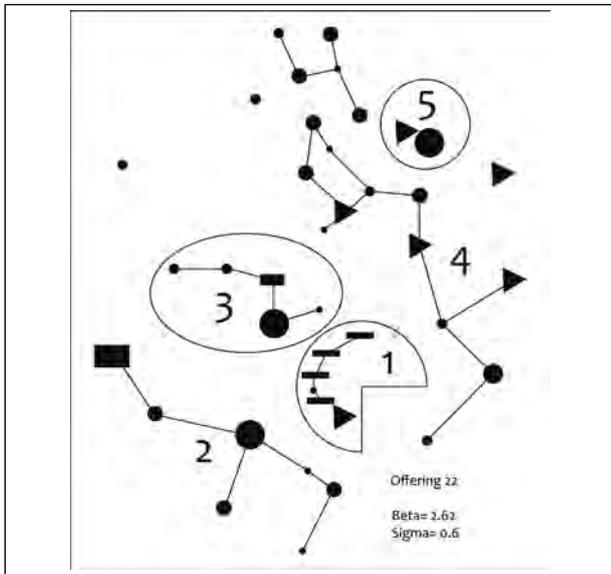


Figure 13. The five artifact arrangements found in Offering 22 through the application of the limited neighbourhood concept. These arrangements coincide with those found in Offering 58. The numbers identified each pattern. (1) Deer-head sceptre, serpentiform sceptre, chicahuaztli sceptre, and nose-plug; (2) human skull, necklace of seven shells (*Oliva* sp), and copper bells; (3) Another human skull, sacrificial knife, and two oyohualli pendants; (4) Xancus shells, conch-shells, turtle shells and sea urchins; (5) pot portraying an unidentified deity and turtle (offering 22), or conch-shell (offering 58). The cradle appears linked to this pattern in offering 22, while in offering 58 it is associated to shells.

of devolving vital forces to earth was accepted as a necessary act of preservation for the entire cosmos.

6. Conclusions

We conclude this paper by highlighting the criteria that justify the adoption of proximity graphs in the analysis of spatial symbolic contexts:

1. The representation of artefact relations in the form of graphs provides a good foundation to analyze archaeological contexts both quantitatively and in a visual way.

2. The proximity graphs described in this paper make possible the discovery of connectivity patterns without making any *a priori* assumption about the morphological features that would eventually emerge from the point set under study.

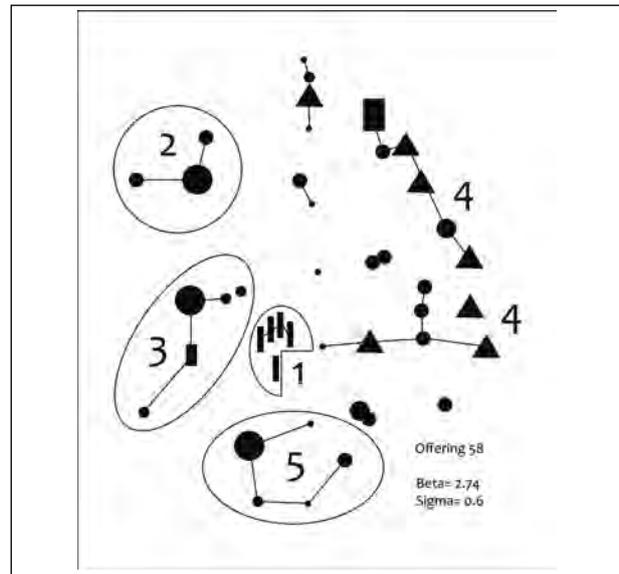


Figure 14. The five artefact arrangements found in Offering 58 through the application of the limited neighbourhood concept. These arrangements coincide with those found in Offering 22: (1) Deer-head sceptre, serpentiform sceptre, chicahuaztli sceptre, and nose-plug; (2) human skull, necklace of seven shells (*Oliva* sp), and copper bells; (3) Another human skull, sacrificial knife, and two oyohualli pendants; (4) Xancus shells, conch-shells, turtle shells and sea urchins; (5) pot portraying an unidentified deity and turtle (offering 22), or conch-shell (offering 58). The cradle appears linked to shells.

3. Both RNG and GG satisfy the condition of being unique. That means that two or more different patterns can be compared in order to determine how similar they are with respect to each other. As Kirkpatrick and Radke (1985, 218) say: “On one hand, we might like to know if two or more input sets have the same or similar forms. Alternately, we might inquire whether the form of a single input set is regular or homogeneous, and if not, isolate the irregularities or homogeneous components.”

4. The adoption of Beta-skeletons and Limited Neighbourhood Graphs brings additional benefits to the method:

(a) Beta-skeletons allow identification of the most strong and consistent morphological features in the point set by highlighting those groups of nodes whose connections remain constant throughout changes in the region of influence.

Figure 15. Arrangement 1 found through the application of the limited neighbourhood concept. It comprises four elements: a) Sceptre deer-head; b) sceptre serpentiform; c) sceptre chicahuaztli; and d) obsidian nose-plug.

Figure 16. Arrangement 2 found through the application of the limited neighbourhood concept: a) Human skull; b) Necklace of 7 Oliva shells; c) Copper bells; d) and e) Models of cradles.

Figure 17. Arrangement 3 found through the application of the limited neighbourhood concept: a) Human skull; b) Sacrificial knife; c) Two oyohualli pendants.

Figure 18. Clay pot portraying a fertility deity from Offering 22. The same type of vessel was found in Offering 58.

(b) Limited Neighbourhood Graphs on the other hand, reveal different possibilities of how the objects might be symbolically linked, attaching dissimilarity measure to the point connections.

(c) Together, Beta-skeletons and Limited Neighbourhood Graphs make the procedure suitable for exploratory spatial analysis.

(d) Finally, they provide a formal way to compare systems across different levels of relative proximity.

References

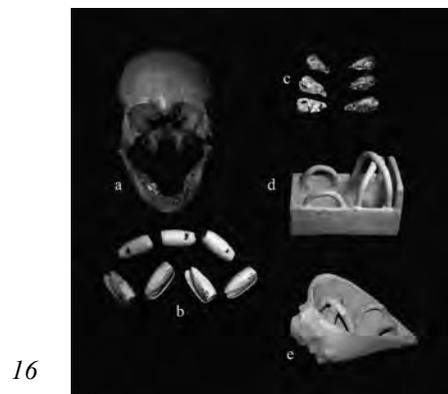
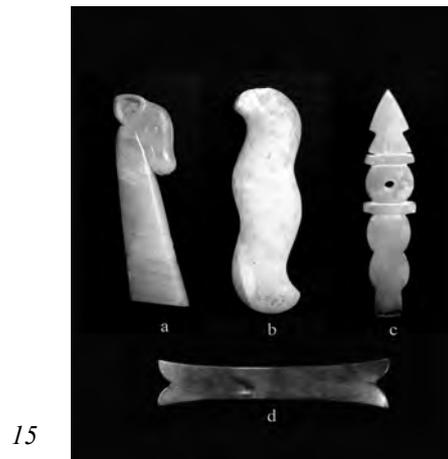
Adamatzky, A. 2010. "On Excitable β -Skeletons." *Journal of Computational Science* 1(3): 175-186.

Durán, D. 1967. *Historia de las Indias de Nueva España e Islas de Tierra Firme*, edited by A. M. Garibay. México: Editorial Porrúa.

Gabriel, K. R., and R. R. Sokal. 1969. "A New Statistical Approach to Geographic Variation Analysis." *Systematic Zoology* 18: 259-278.

Graulich, M. 1999. *Fiestas de los pueblos indígenas. Ritos Aztecas. Las fiestas de las veintenas*. México: Instituto Nacional Indigenista.

Haggett, P., and R. J. Chorley. 1969. *Network analysis in*



geography. London: Edward Arnold Publishers.

Huang, N.-F. 1990. "A Divide-and-Conquer Algorithm for Constructing Relative Neighbourhood Graph." *BIT* 30: 196-206.

Hurtado, F., G. Liotta, and H. Meijer. 2001. "Optimal, Suboptimal, and Robust Algorithms for Proximity Graphs." *Lecture Notes in Computer Science* 2115: 2-13.

Jaromczyk, J. W., and M. Kowaluk. 1987. "A Note on Relative Neighbourhood Graphs." In *Proceedings of the Third Annual Symposium on Computational Geometry*, edited by D. Soule, 233-241. New York: ACM.

Jaromczyk, J. W., and M. Kowaluk. 1991. "Constructing the Relative Neighbourhood Graph in 3-Dimensional Euclidean Space." *Discrete Applied Mathematics* 31: 181-191.

Jaromczyk, J. W., and G. T. Toussaint. 1992. "Relative Neighbourhood Graphs and their Relatives." *Proceedings of the IEEE* 80(9): 1502-1517.

Jiménez-Badillo, D. 2004. *A Method for Interactive Recognition of Three-dimensional Adjacency Patterns in Point Sets, Based on Relative Neighbourhood Graphs: An Archaeological Application*. PhD diss., University of London.

Jiménez-Badillo, D. 2006. "El Análisis Espacial de las Ofrendas Mexicas por Medio de Gráficos de Proximidad." In *Arqueología e Historia del Centro de México: Homenaje a Eduardo Matos Moctezuma*, edited by L. López-Lujan, D. Carrasco, and L. Cué, 455-482. México: Instituto Nacional de Antropología e Historia.

Jiménez-Badillo, D. 2009a. "Una Aplicación de la Teoría de Gráficas en Arqueología (Primera parte)." *Carta Informativa: Revista de la Sociedad Matemática Mexicana* 60: 1-8.

Jiménez-Badillo, D. 2009b. "Una Aplicación de la Teoría de Gráficas en Arqueología (Segunda parte)." *Carta Informativa: Revista de la Sociedad Matemática Mexicana* 61: 1-8.

Jiménez-Badillo, D. 2011. "Interpretación de dos ofrendas del Templo Mayor de Tenochtitlan con el apoyo de fuentes pictográficas." In *Las Imágenes Precolombinas. Reflejo de Saberes*, edited by M.C. Valverde Valdés, and V. Solanilla,

361-391. México, Universidad Nacional Autónoma de México.

Jiménez-Badillo, D. 2012. *Relative Neighbourhood Networks for Archaeological Analysis*. In *Revive the Past, Proceedings of the 39th Conference in Computer Applications and Quantitative Methods in Archaeology*, Beijing, China. 12-16 April 2011, edited by M.Q. Zhou, I. Romanowska, Z.K. Wu, P. F. Xu, and P. Verhagen, 370-380. Amsterdam: Pallas Publications.

Jiménez-Badillo, D., and D. Chapman. 2002. "An Application of Proximity Graphs in Archaeological Spatial Analysis." In *Contemporary Themes in Archaeological Computing*, edited by D. Wheatley, G. Earl, and S. Poppy, 90-99. Oxford: University of Southampton, Department of Archaeology, Monograph 3.

Kirkpatrick, D. G., and J. D. Radke. 1985. "A Framework for Computational Morphology." In *Computational Geometry*, edited by G. Toussaint, 217-248. North Holland: Elsevier Science Publishers.

Lankford, P. M. 1969. "Regionalization: Theory and Alternative Algorithms." *Geographical Analysis* 1: 196-212.

López Luján, L. 1994. *The offerings of the Templo Mayor of Tenochtitlan*. Colorado: University Press of Colorado.

Matos Moctezuma, E. 1979. El proyecto Templo Mayor: Objetivos y programas. In *Trabajos arqueológicos en el centro de la Ciudad de México (antología)*, edited by E. Matos Moctezuma, 13-26. Mexico: Instituto Nacional de Antropología e Historia.

Matos Moctezuma, E. 1982. *El Templo Mayor: Excavaciones y estudios*. Mexico: Instituto Nacional de Antropología e Historia.

Matos Moctezuma, E. 1986. *Vida y muerte en el Templo Mayor*. Mexico: Ediciones Océano.

Matos Moctezuma, E. 1988. *The Great Temple of the Aztecs: Treasures of Tenochtitlan*. London: Thames and Hudson.

Matula, D. W., and R. R. Sokal. 1980. "Properties of Gabriel Graphs Relevant to Geographic Variation Research and the Clustering of Points in the Plane." *Geographical Analysis* 12: 205-222.

- Nooy, W., A. Mrvar, and V. Batagelj. 2005. *Exploratory Social Network Analysis with Pajek*. Cambridge: Cambridge University Press.
- Rao, S. V. 1998. *Some Studies on Beta-skeletons*. PhD Diss. Indian Institute of Technology.
- Sahagún, B. 1956. *Historia general de las cosas de la Nueva España*, edited by A. M. Garibay. México: Editorial Porrúa.
- Su, T.-H., and R.-C. Chang. 1990. "The k-Gabriel Graphs and their Applications." In *Lecture Notes on Computer Science 450: Proceedings of the International Symposium SIGAL '90*, edited by T. Asanom T. Ibaraki, H. Imai, and T. Nishizeki, 66-75. New York: Springer-Verlag.
- Su, T.-H., and R.-C. Chang. 1991a. "Computing the k-Relative Neighbourhood Graphs in Euclidean Plane." *Pattern Recognition* 24(3): 231-239.
- Su, T.-H., and R.-C. Chang. 1991b. "On Constructing Relative Neighbourhood Graphs in Euclidean k-Dimensional Spaces." *Computing* 46: 121-130.
- Toussaint, G.T. 1980a. "The Relative Neighbourhood Graph of a Finite Planar Set." *Pattern Recognition* 12: 261-268.
- Toussaint, G.T. 1980b. "Comment to the article: 'Algorithms for Computation of Relative Neighbourhood Graph' by R. B. Urquhart." *Electronic Letters* 16(22): 860-861.
- Toussaint, G.T. 1980c. "Pattern Recognition and Geometric Complexity." In *Proceedings of the Fifth International Conference on Pattern Recognition*, 1324-1347. Miami Beach: International Association for Pattern Recognition and IEEE Computer Society.
- Toussaint, G.T. 1988. "A Graph-theoretical Primal Sketch." In *Computational Morphology*, edited by G. T. Toussaint, 229-261. Amsterdam, North-Holland: Elsevier Science Publishers.
- Toussaint, G.T., and R. Menard. 1980. "Fast Algorithms for Computing the Planar Relative Neighbourhood Graph." In *Proceedings of the Fifth Symposium of Operations Research*, 425-428. Köln: University of Köln.
- Urquhart, R. 1980. "Algorithms for Computation of Relative Neighbourhood Graph." *Electronic Letters* 16(14): 556-557 (includes a comment by G. T. Toussaint, 860-861).
- Urquhart, R. 1982. "Graph Theoretical Clustering Based on Limited Neighbourhood Sets." *Pattern Recognition* 15(3): 173-187.
- Urquhart, R. 1983. "Some Properties of the Planar Euclidean Relative Neighbourhood Graph." *Pattern Recognition Letters* 16: 317-322.

Bayesian Spatial Modelling of Radiocarbon Dated Archaeological Artefacts Using R-INLA

Juhana Kammonen, Tarja Sundell
University of Helsinki, Finland

Petro Pesonen
University of Helsinki and National Board of Antiquities, Finland

Markku Oinonen and Päivi Onkamo
University of Helsinki, Finland

Abstract:

Radiocarbon dated archaeological artefacts can be used as spatiotemporal proxies for prehistoric population size and distribution. In this paper, a radiocarbon dataset from eastern Fennoscandia is predisposed for tests with statistical software suitable for spatial and spatiotemporal analysis. We present a Bayesian computing approach using the recently developed R-INLA (Integrated Nested Laplace Approximation in R-environment) software. We adapt our previous spatial model to R-INLA comparing the results with our earlier MCMC (Markov chain Monte Carlo) approach. We also take the first steps of employing a spatiotemporal Poisson point process model into the analysis. A series of test runs show that results are roughly congruent with MCMC approach. Moreover, the INLA approach proves exponentially faster and memory-efficient. Model families are defined through compact R statements, which allows for systematic exchange between different statistical models. This paper is part of a multidisciplinary effort in spatiotemporal analysis of archaeological data from eastern Fennoscandia.

Keywords:

Spatiotemporal Analysis, Eastern Fennoscandia, R-INLA

1. Introduction

In this paper, we present spatial analysis with the R-INLA software using a set of radiocarbon dates and a dataset from Typical Comb Ware (TCW) from the Finnish National Archaeological Database. The data is the same as used previously by Kammonen et al. (2012) in spatial analysis with MCMC methods of the WinBUGS software. The results are compared with those of the previous approach. In addition, performance of the R-INLA software is assessed with a simple benchmark test in laptop computer environment.

1.1 Archaeological background

During recent years, the archaeological record of eastern Fennoscandia, especially the territory of Finland and ceded Karelia (Fig. 1), has faced increased research activity on both sides of the border. Ceded Karelia refers to the region southeast of the current borders of Finland, which was part

Corresponding author: juhana.kammonen@helsinki.fi

of the country before Moscow Armistice in 1944. Due to post-World War II standstill, most of the Karelian Isthmus remained only fractionally studied until the collapse of Soviet Union. Due to several Fenno-Russian joint research projects the Stone Age research in northern part of Karelian Isthmus has now been intensified (see e.g. Lavento 2008; Nordqvist et al. 2009; Seitsonen et al. 2012).

The archaeological records from all of the excavations conducted in Finland have been documented in the Registry of Ancient Monuments, a national database of the National Board of Antiquities (http://kulttuuriymparisto.nba.fi/netsovellus/rekisteriportaali/mjreki/read/asp/r_default.aspx). This database is currently being updated and digitized to enable extensive use of data analysis. The augmentation is done in our multidisciplinary research project Argeopop (<http://www.helsinki.fi/bioscience/arageopop>). Radiocarbon dating of the archaeological finds in the database is an ongoing process. At its present state the database already allows for spatial, and in

the future, also spatiotemporal analysis of the data.

Based on archaeological studies, eastern Fennoscandia is believed to have received its first inhabitants from the south, southeast and east in prehistory. The human occupation in the area begins at approximately 9,000 years cal BCE, immediately after the retreat of glacial ice of the Vistula Ice Age. When the first settlers arrived, vast areas that are located on dry land today were underwater, especially on the western coast of the region. From a small founder population and very likely through a series of founder effects (Sundell et al. in press), there was a period of population growth and a temporary population maximum around 4000-3500 cal BCE (Tallavaara et al. 2010). The population peak period coincides with a local maximum in mean temperatures, which may have made living conditions more pleasant. The population peak period is followed by a distinct weakening in the archaeological signal (Tallavaara et al. 2010). This is indicative that after the peak period, the population in the area has gone through a severe decline but has eventually recovered – without being replaced. This kind of demographic event is called a population bottleneck and is assumed to explain some peculiar components of the genetic profile of present-day Finns (Sajantila et al. 1996; Kittles et al. 2009). Population bottlenecks and their effect on genetic diversity have been extensively studied with population genetic analyses and computer simulation methods (e.g. Sundell et al. 2010; Sundell et al. in press; Valtonen et al. 2012).

A spatially located archaeological find is a distinct sign of previous human activity in that location, a self-evident fact that from the viewpoint of this paper validates the concept of archaeological signal in our analysis. Consequently, studying the distribution of finds on a larger geographical area for different time periods is likely to provide clues of differential human activity in the area. As the radiocarbon dated artefacts of the National Archaeological Database have a record of the spatial location of the find, the dataset is most suitable for spatial and spatiotemporal analysis. In addition to radiocarbon dating, it is possible to date archaeological artefacts typologically, which is the case with the Typical Comb Ware (TCW) dataset of 676 finds from the study area. This ceramic style is associated with the hunter-gatherers that are known



Figure 1. Reference map showing the location of Finland (black) and ceded Karelia (dark grey) in Northern Europe.

to have been living in the area around 4000-3500 cal BCE, the temporal span for the radiocarbon dataset used in this study.

1.2. Computational and statistical issues

Latent Gaussian models are a popular and flexible class of models. However, Bayesian inference for these models has proved very time consuming. This is usually due to overwhelmingly multidimensional posterior probability distributions of the response variables. Recently, Rue, Martino and Chopin (2009) have introduced the Integrated Nested Laplace Approximation (INLA) method for deterministic fast approximate inference. INLA provides a way to approximate posterior marginal distributions of non-Gaussian response variables frequently encountered in Bayesian computing.

The INLA method was designed to provide fast inference for a large class of practical Bayesian problems (Simpson et al. 2011). The R-INLA package works through an R (<http://www.r-project.com>) interface to the INLA program. The INLA program itself is written in C. The syntax for the R-INLA package is based on the inbuilt Generalized Linear Model (glm) function in R. The R-INLA package is available at <http://www.r-inla.org>

Finally, R-INLA has been extended to include models with data from various sources. Individual

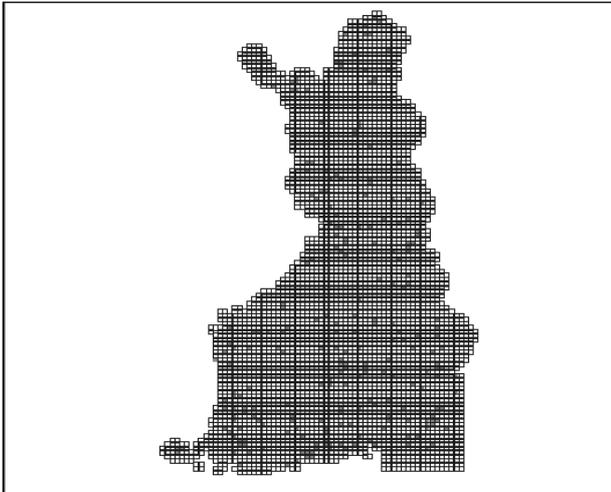


Figure 2. Finland and ceded Karelia divided into a grid of 3997 squares. Each square covers an area of 10 km to 10 km. Red dots mark the locations (on 10 km to 10 km scale) of radiocarbon dated finds (9000-1500 cal BCE) of the National Archaeological Database. For full colour pictures please see the online versions of the proceedings.

likelihood functions required by the different sets of data can be programmed in R-INLA by reformatting the data into a matrix where the number of columns is equal to the number of likelihoods.

2. Materials and Methods

Following the initial settings of our previous approach (Kammonen et al. 2012) Finland and ceded Karelia were first divided into a grid of 10km to 10km square cells. The division results in a total of 3997 cells. Initially, each cell were given an integer value 0 if there were no archaeological finds in this cell, and 1 if there were one or more archaeological finds (Fig. 2). The grid is contiguous, as to say that every cell has at least one neighbour and the grid doesn't contain any isolated components. The grid was not configured to account for geographical factors such as lakes, rivers, hills and sea shore. Thus, the only boundaries of the grid are the actual spatial limits of the grid which correspond roughly to the land area of Finland and ceded Karelia (Fig. 2).

We employed the Besag-York-Mollie (BYM) model inbuilt in R-INLA. The BYM-model is based on image analysis (Besag 1986) and assumes similarity of neighbouring areas in geospatial applications. The value of response variable p_i , which corresponds to the posterior probability of

making at least one archaeological find in cell i , was achieved via a logit link:

$$\text{logit}(p_i) = \alpha_0 + \lambda_i \quad (1)$$

in which α_0 is the general probability of making an archaeological find anywhere in the study area and λ_i is the local spatial effect in cell i . λ_i was given an improper conditional autoregressive (CAR) prior distribution. Posterior distributions were then produced according to Bayesian framework and based on archaeological data.

To compare the results of the INLA approach with those of Kammonen et al. 2012, we used the exactly same radiocarbon dataset (187 dates from 4000-3500 cal BCE) and the Typical Comb Ware dataset of 676 finds. Also the study area is exactly the same. We also ran a series of test runs in a simple laptop computer environment (1.6 GHz AMD processor with 2 gigabytes of Random Access Memory). The aim of this was to assess the performance of the software.

The results were visualized with the graphics methods of the R-interface as well as Pitney Bowes' Business Insights (<http://www.pbinsight.com>) MapInfo 10.0. Geographical details such as rivers, lakes, country borders and an estimate of the Baltic Sea shoreline were added for illustration purposes and are identical to those used in Kammonen et al. 2012.

3. Results

The built-in BYM-model of R-INLA presents the user with a completely different interface than the previous MCMC approach. Instead of the R/Spplus-like syntax of the WinBUGS language (Lunn et al. 2000), model descriptions in R-INLA are implemented by defining formulae in the R programming environment. These formulae are then passed to the `inla()`-function of the software that performs the actual approximations. Assigning prior probability distributions and initial values to hyperparameters is largely built-in, but allows for very smooth exchange between priors and latent model families.

The comparison between the MCMC approach of Kammonen et al. 2012 and the INLA approach is

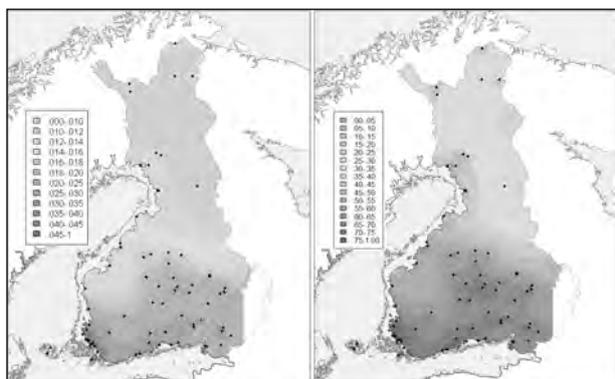


Figure 3. Results (posterior mean values) of the spatial analysis of radiocarbon dates from period 4000-3500 cal BCE from Kammonen et al. 2012 (on the left) and the corresponding results with the INLA approach (on the right). For full colour pictures please see the online versions of the proceedings.

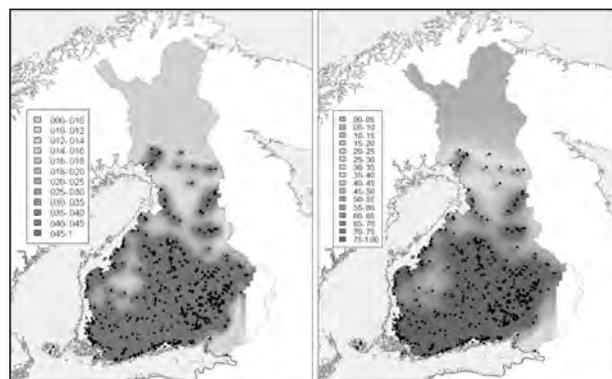


Figure 4. Results (posterior mean values) of the spatial analysis of Typical Comb Ware (TCW) finds from Kammonen et al. 2012 (on the left) and the corresponding results with the INLA approach (on the right). For full colour pictures please see the online versions of the proceedings.

presented below (Figs. 3 and 4) The results for the test runs to assess the performance of the R-INLA software are in Table 1.

4. Discussion and Conclusions

The posterior mean maps (Figs. 3 and 4) outline an apparent congruence between the results of the MCMC approach (Fig. 3 left; Fig. 4 left) and results of the INLA approach (Fig. 3 right; Fig. 4 right). The probability (1) achieved through the response variable p_i with the INLA approach highlights the same geospatial areas as the response variable of the MCMC approach. This is especially evident in the analysis using TCW ceramic finds as data, where areas with higher posterior probability and the range of the spatial effects are all but identical (green-yellow-red colour scale in Fig. 4).

In addition, the INLA approach proved computationally faster (Table 1) and less memory consuming. We could even run the model with 1km to 1km grid cell size in a reasonable time, something that could not be done even in supercomputer environment with the MCMC approach (Table 1, bottom row). Evidently, R-INLA is a very

efficient piece of software for spatial analysis using differential datasets with irregular spatial lattices.

As expected, the inbuilt BYM-model of R-INLA works similarly than that presented in the earlier MCMC approach (Kammonen et al. 2012), meaning that a spatially located archaeological find results in an archaeological signal in the location. The similarity assumption of neighbouring areas in the BYM-model then accordingly propagates this signal to adjacent cells, which is disclosed by the elevated posterior probability in those areas. The main setback with the BYM-model is that the response variable p_i has an improper CAR-prior distribution, meaning that a closed form for the probability density function does not exist. This is exactly where the computational problems of the approach arise – especially with models that are more complicated than presented in this case study. An obvious mathematical solution to this kind of problem would be to approximate the probability density function and the INLA approach offers a fast and reliable way to do this.

The results of the INLA approach are relatively compatible with the results produced with the

Grid cell size	T_{MCMC}	T_{INLA}
20 km to 20 km	1200 sec	26 sec
10 km to 10 km	3600 sec	270 sec
1 km to 1 km	-	8800 sec

Table 1. Comparison of average computing times (T) between the MCMC approach of Kammonen et al. 2012 (T_{MCMC}) and computing times with the corresponding model in R-INLA (T_{INLA}).

methods in our previous approach (Kammonen et al. 2012) despite the apparent differences in the scale of the posterior means (green-yellow-red colouring in Figs. 3 and 4). With the help of R-INLA we have refined our methodology of spatial analysis of archaeological data to the point where extensions into the spatiotemporal domain become feasible. In future, a set of guidelines (“a toolbox”) for fitting various spatiotemporal models (Illian et al. submitted) should be used to enable lightweight and systematic exchange between different models.

To enable spatiotemporal extensions of the methodology, our aim is to first phase in a Poisson point process model. Poisson point processes are a type of a random process and are well studied objects in probability theory. A thorough discussion of the basic theory is given in Daley and Vere-Jones (1988). On one hand, Poisson process is a stochastic process which calculates the number of objects and the time that these objects occur in a given time interval. The number of events in disjoint intervals is independent and has a Poisson distribution. On the other hand, Poisson point process means the application of a Poisson process into a space with 2 (or more) dimensions. As spatiotemporal models, they are most feasible in exploring archaeological research questions often concerning both space and time.

Spatiotemporal modelling is a hot topic in computer-aided archaeology. Having a systematic method to produce and refine spatiotemporal models is attractive as new data becomes available for free use in the form of location information and enormous GIS databases. Eventually phasing in a Poisson point process model and later an even more suitable Cox process model is the first step in actual spatiotemporal analysis of archaeological data from eastern Fennoscandia.

References

Argeopop. 2012. “Argeopop - University of Helsinki.” Last modified 27 June 2012. <http://www.helsinki.fi/bioscience/argeopop>

Besag, J. 1986. “On the Statistical Analysis of Dirty Pictures.” *Journal of the Royal Statistical Society. Series B*, 48: 259-302.

Daley, D.J., and Vere-Jones, D. 1988. *An Introduction to the Theory of Point Processes*. New York: Springer.

Illian, J., Sørbye, S.H., and Rue, H. *In press*. “A toolbox for fitting complex spatial point process models using integrated Laplace transformation (INLA).” *Annals of Applied Statistics*.

Kammonen, J., T. Sundell, E. Moltchanova, P. Pesonen, M. Oinonen, M. Heger, M. Haimila, and P. Onkamo. 2012. “Bayesian spatial analysis of archaeological finds and radiocarbon dates: An example from Finland 4000-3500 cal BC.” In *Revive the Past, Proceedings of the 39th Conference in Computer Applications and Quantitative Methods in Archaeology*, Beijing, China. 12-16 April 2011, edited by M.Q. Zhou, I. Romanowska, Z.K. Wu, P. F. Xu, and P. Verhagen, 318-3250. Amsterdam: Pallas Publications.

Kittles, R. A., A. W. Bergen, M. Urbanek, M. Virkkunen, M., Linnoila, D. Goldman, and J. C. Long. 1999. “Autosomal, Mitochondrial and Y Chromosome DNA Variation in Finland: Evidence for a Male-Specific Bottleneck.” *American Journal of Physical Anthropology* 108: 381-99.

Lavento, M. ed. 2008. *Karelian Isthmus. Stone Age studies in 1998-2003*. *Iskos* 16. Helsinki: Suomen Muinaismuistoyhdistys.

Lunn, D. J, A. Thomas, N. Best, and S. Spiegelhalter. 2000. “WinBUGS -- a Bayesian modelling framework: concepts, structure, and extensibility.” *Statistics and Computing* 10: 325-337.

Nordqvist, K., O. Seitsonen, and M. Lavento. 2009. “Waterways and the Stone Age and Early Metal Period studies on the Karelian Isthmus – The Pre-World War II studies and the research carried out by the University of Helsinki in 1998-2006.” *Quaternary International* 203(1-2): 25-32.

Registry of Ancient Monuments. 2012. “Muinaisjäänösrekisteri.” Accessed 27 June 2012. http://kulttuuriymparisto.nba.fi/netsovellus/rekisteriportaali/mjreki/read/asp/r_default.aspx.

R-INLA. 2012. “The R-INLA project.” Accessed 27 June 2012. <http://www.r-inla.org>.

Rue, H., S. Martino, and N. Chopin. 2009. “Approximate Bayesian inference for latent Gaussian models by using

- integrated nested Laplace approximations.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 71 (2): 319-392.
- Sajantila, A., A.-H. Salem, P. Savolainen, K. Bauer, C. Gierig, and S. Pääbo. 1996. “Paternal and maternal DNA lineages reveal a bottleneck in the founding of the Finnish population.” *Proceedings of the National Academy of Sciences USA* 93: 12035-9.
- Seitsonen, O., K. Nordqvist, D. V. Gerasimov, and S. N. Lisitsyn. 2012. “The Good, the Bad, the Weird: Stone Age and Early Metal Period Radiocarbon Dates and Chronology from the Karelian Isthmus, North-West Russia.” *Geochronometria* 39(2): 101-121.
- Simpson, D., F. Lindgren, and H. Rue. 2011. “Fast approximate inference with INLA: the past, the present and the future.” *arXiv: 1105.2982v1*.
- Sundell, T., M. Heger, J. Kammonen, and P. Onkamo. 2010. “Modelling a Neolithic Population Bottleneck in Finland: A Genetic Simulation.” *Fennoscandia Archaeologica* XXVII: 3-19.
- Sundell T., J. Kammonen, M. Heger, J. Palo and P. Onkamo. *In press*. Retracing Prehistoric Population Events in Finland Using Simulation (this volume).
- Tallavaara, M., P. Pesonen, and M. Oinonen. 2010. “Prehistoric population history in eastern Fennoscandia.” *Journal of Archaeological Science* 37: 251-60.
- Valtonen, M., J.U. Palo, M. Ruokonen, M. Kunnasranta and T. Nyman. 2012. “Spatial and temporal variation in genetic diversity of an endangered freshwater seal.” *Conservation Genetics* 5(13): 1231-1245.

From Space to Graphs to Understand Spatial Changes Using Medieval and Modern Fiscal Sources

Xavier Rodier

University François-Rabelais/CNRS, France and University of the Witwatersrand, South Africa

Mélanie Le Couédic

University of Pau/CNRS, France

Florent Hautefeuille

University of Toulouse II Le Mirail/CNRS, France

Samuel Leturcq

University François-Rabelais/CNRS, France

Bertrand Jouve

University of Lyon 2 Lumière, France

Etienne Fieux

University of Toulouse I Paul Sabatier/CNRS, France

Abstract:

Medieval and modern fiscal documents provide valuable information about the spatial organization of the countryside. They usually describe each plot, indicating the owner's name, the surface area, the land use, the neighbours, and the location. These lists, which contain hundreds of plots with their relative location, have great potential for analyzing the spatial pattern of village territories and the evolution of the landscape, especially when there is a series of successive documents for the same territory, as is often the case. Nevertheless, they have been dramatically under-exploited because it is difficult to reconstruct the landscape from fieldbooks without maps – and field-maps in France only appeared in the 17th-18th centuries. To overcome this difficulty, we propose a model of plots described in fiscal documents based on the use of graph theory. The main idea is to take spatial information from registers and former plot plans and transform it into graphs. The translation of spatial data into graphs should allow a bridge to be created between mapped and unmapped historical documents. On these graphs, each plot corresponds to a vertex and each neighbourhood relationship (adjacency) - visible on a plan or described in the documents as a neighbour - to an edge. The work is carried out using two complementary approaches: the first focuses on establishing a database dealing with records without maps. The second extracts the graph from former plot plans in a Geographic Information System. The goal is to match the graphs produced by these two approaches in order to analyze spatial changes by comparing the graphs, which is not possible from plans or registers alone. First, the protocol to transform plot plans into graphs in the GIS will be developed. Secondly, we will consider the possibilities of comparing two graphs corresponding to two successive states in the GIS.

Keywords:

Fieldbook, Cadaster, Geographic Information Systems, Graph Theory, Medieval Ages, Modern Period

Corresponding author: xavier.rodier@univ-tours.fr

Among all the documents available to historians and archaeologists to reconstruct the dynamics of land use during mediaeval and modern times (13th-18th centuries), there is a category of sources that is particularly rich in spatial organization data, namely land registers including *terriers* (fieldbooks) and cadastres, commonly called “*compoix*” in the south of France. For France alone, there are several tens of thousands of fiscal registers prior to the 19th century, the oldest going back to the 13th century. These documents, drawn up for the purpose of levying taxes on land ownership (private for the *terriers*, public for the cadastres), are generally in the form of registers, the most recent documents sometimes including plans, and they can be considered as the forerunners of present-day cadastres. This type of source, found in France and throughout Europe, provides a wealth of high quality information, its main interest lying in the form of the register, namely the list of names and lands, together with a large number of details, notably the amount of dues to be paid (“*allivrement*” in the cadastre or “*compoix*”). The standardized registration of these long series (often more than several thousand items) makes the use of analysis tools, particularly statistics, particularly apt. Above all, the documents give not only the name and profession of the landowner, but also sometimes the name of the farmer and often the surface area, land use (arable land, vineyard, meadow, game preserve, woodland, house, etc.), location through microtoponyms, the names of the owners of adjacent plots (known as “*confronts*”), etc., making it possible to situate the plots in relation to each other within a territory, with the help of a *terrier* if there is one, or of the Napoleonic cadastre. Various historical issues regarding land organization (landscape structures, distribution of land use, property and farm organization, etc.) can thus be addressed.

Although these mediaeval and modern fiscal documents provide a wealth of spatial information, with whole areas described in detail, plot by plot, and at successive periods, historians are never able to correlate this spatial information over long periods of time, which would enable them to identify the dynamics of reorganization. This difficulty is increased by the absence of plot plans (with a few exceptions) prior to the 17th-18th centuries. The researcher thus has to put together the pieces of a puzzle using topographic and microtoponymic

details, and particularly references to neighbouring plots (*confronts*). This work, which is long and fastidious, is difficult due to gaps in the information (e.g. neighbouring plots that are not described), ambiguities (e.g. the names of neighbours), and errors. It is however conceivable to reconstruct the land divisions of a restricted area, but in no case for the whole of a rural area. Historians de facto use essentially recent documents from the time that plans were available. The lack of any cartographic referent complicates the use and processing of the whole information, and particularly the possibility of observing changes in the plot plan from one state to another in order to gain a better understanding of spatial dynamics. Thus, processing the spatial information in the *terriers* and *compoix* generally remains limited to one-off initiatives, using “home-made” solutions with little emphasis on the dynamic character (Claveirole and Pélaquier 2001; Brunel et al. 2002). The 1985 conference in Rome (Biget et al. 1989) provided a ground-breaking and essential way forward for research in this area. It raised issues that still exist today, highlighting the need to create a tool capable of managing the mass of spatial data in mediaeval urban cadastres. An article written at that time even considered the possibility of using a mathematical object, graphs, to speed up the process of reconstituting mediaeval urban plot divisions from “*compoix*” (Montpied and Rouault 1989). This project was first discussed in a short article in the journal *Le médiéviste et l'ordinateur* (Montpied and Rouault 1982). The proposed experiment was essentially limited to constructing the graph (the plots were reduced to a single dot, and two neighbouring plots were indicated by two adjacent peaks on the graph), and then to visually identifying “topological accidents” on the graph. As far as we know, no further progress has been made in terms of developing this type of method to “reconstruct” ancient plot divisions since this work in the mid-1980s, although the most recent studies have highlighted this technical shortfall (Leturcq 2007; Truel 2007; Conesa 2007; Poirier 2010).

Picking up this idea of using the concept of mathematical graphs (Berge 1958), a research programme was launched in 2010, supported for three years by the French National Research Agency (ANR), called MODELESPACE¹. The aims

¹ ANR-09-BLAN-0322-02 «Modélisation des espaces préindustriels». This 2010-2012 project is led by Florent

of this programme are to go beyond the manual reconstruction of plot plans by generating a tool which would eventually make it possible to carry out ambitious projects to analyse the spatial dynamics of urban or rural areas for which several *terriers*, *compoix* and cadastres are available, constituting successive records of land organization. This tool for modelling spatial dynamics thus proposes a change of scale, as it should make it possible to incorporate the successive states of a single space into the same data base, to provide a graphic model of each of these states, and to measure spatial changes by comparing these different representations, overcoming the obstacle of using heterogeneous documents (notably plan vs. register). The aim is thus not to create an automatic tool to reconstruct a plot structure as close as possible to past reality; that would not only be unrealistic, but also based on a static concept of areas and landscapes.

There has been extensive work on graph theory by mathematicians in the last 25 years, notably with regard to its application to social sciences. The use of graph theory to analyse large interaction networks (social networks, internet networks, protein interactions, transport, etc.) developed rapidly in the 2000s with the possibility of direct on-line access to vast quantities of data and the development of mathematical models of graphs (Watts and Strogatz 1998; Barabási and Albert 1999). We propose combining these now classical network analysis tools with topology tools² to study graphs in which each node represents a plot and two nodes are linked by an edge when the two plots are geographically adjacent to each other. Since the ground-breaking study by Atkin (Atkin 1977), recent studies have led to advances in the interactions between graph and topology (Babson et al. 2006; Chen et al. 2001; Dochterman 2009; Boulet et al.

Hautefeuille (FRAMESPA), Bertrand Jouve (ERIC) and Samuel Leturcq (Citeres-LAT). The internet site can be consulted at the following address: <http://modele-espace.univ-tlse2.fr/>

² Topology, the science of spatial relations, is a branch of mathematics created by the French mathematician Poincaré at the end of the 19th century in order to formalize reasonings of a qualitative nature. Topology is concerned with the spatial relationships between objects, whether or not they are located geographically. It allows an approach, a qualitative description, to be formalized mathematically. The neighbourhood relationships between objects are expressed in terms of continuity, contiguity, proximity, adjacency and connectivity. Modifying the spatial relationships brings about a change in topology (Pirot2004, 426).

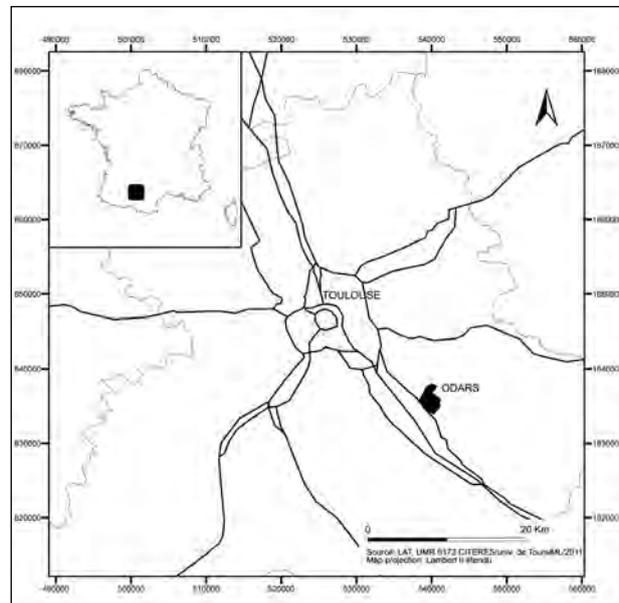


Figure 1. Map locating the district of Odars (31, France).

2010). The present study applies this work to an actual data base.

To test this method, we chose the parish of Odars on the outskirts of Toulouse (Fig. 1). The records of this parish contain a particularly large collection of *compoix* and *terriers*, with and without maps, making it possible to develop methods to measure spatial dynamics. There are three successive *compoix* for 1479, 1551 and 1598, a *terrier* with a map drawn up in 1759, and a “Napoleonic” cadastre dated 1811.

Based on this corpus, we look first at the general principles of the work to create the series of graphs. Secondly, we describe the difficulties encountered in comparing these graphs, and the possible courses of action to measure changes in the organization of land in Odars between the 15th and 19th centuries.

1. The Foundations of Spatial Analysis: Spatial Construction and Adjustment of the Graphs³

1.1 Graph construction

The procedure involves transforming spatial

³ For further details, see Le Couédic et al. 2012 (forthcoming).

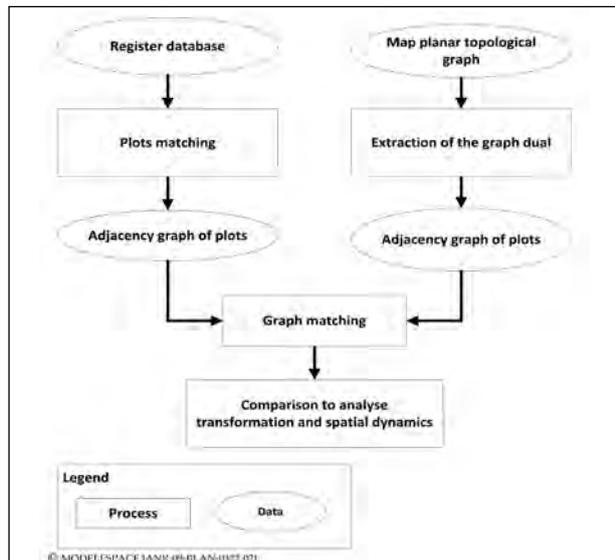


Figure 2. Stages of processing registers and plans.

information from registers and ancient plot plans into graphs (called “dual graphs”) in which each plot corresponds to a peak and each neighbourhood relationship (visible adjacency on a map or described in documents by mention of “confronts”) corresponds to an edge. Depending on the type of source, with or without maps, two different procedures are followed (Fig. 2).

The first (Fig. 3) involves extracting the dual graph (Berge 1958, 206-217) of the ancient plot plans in a Geographic Information System (GIS). The first step is to vectorize the original plan using a topological structure in order to obtain a planar graph of the plot system⁴. The peaks of this graph are the intersection points of the plot boundaries and the curves are the plot boundaries. From this graph, the dual graph is then extracted, in which each plot is represented by a peak, and two neighbouring plots are linked by an edge. In this dual graph, the attributes of the plots are assigned to the peaks that represent them in order to obtain what we call the plot adjacency graph.

The second process involves designing a tool to enter and analyze the spatial data of the fiscal

4 A plot plan is a topological planar graph. Thus a graph G is planar if it can be represented on a plan in such a way that the peaks are at distinct points and simple edges only meet at their extreme points. A topological planar graph is a representation of a planar graph G on a plane” (Pirrot 2004, 422-423; Berge 1970).

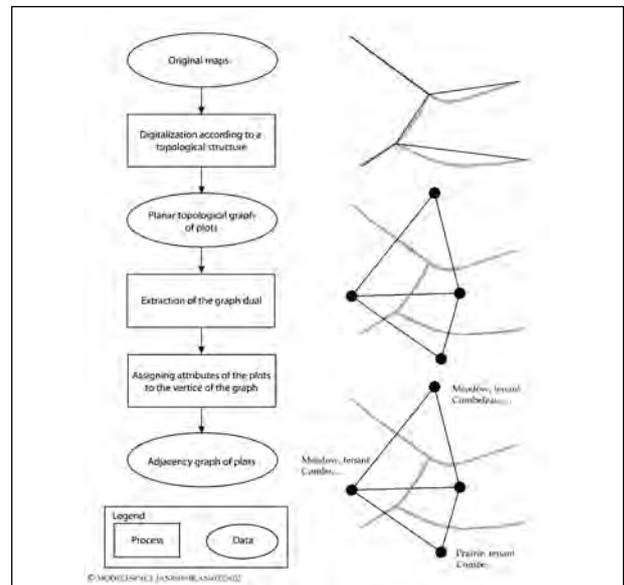


Figure 3. Left: Transition from original plans to adjacency graph of plots of land. Right: construction of the different types of graph from a plot plan (boundaries in grey); top, topological planar graph; middle, dual graph; bottom, adjacency graph.

registers in order to draw the graphs. The database is created in such a way as to incorporate the data of several registers and matrices into a single set in two steps⁵. In the first step, all the information contained in the registers is entered, not only taking into account the variety of editorial forms, but also using tools that can link words designating the same person or the same place with different spellings (disambiguation tools). The second step provides algorithms for pairing plots with neighbourhood links (Fig. 4). These algorithms allow pairs of plots with a common boundary to be extracted automatically. These features are recognized directly or are based on different workaround methods to remove ambiguities (place names, neighbour’s neighbour, etc.). This method makes it possible to generate a graphml file automatically, which can be read directly by a graph display programme such as YED or GEPHI (Fig. 4). A clear distinction is made between the *confronts* based on ownership (which may be linked to another plot described elsewhere in the register), and those which are “invariant”, corresponding to roads, rivers or demesnes not described elsewhere in the compoix. The latter category has the advantage of being relatively

5 The Tercomp base can be consulted on line: <http://modele-space.univ-tlse2.fr/>

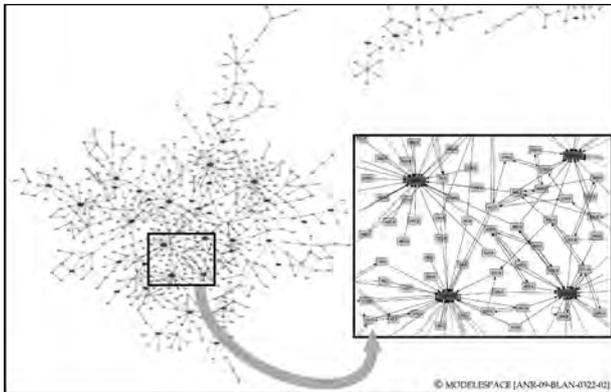


Figure 4. Dual graph aspatial from the cadastre (without plan) of Odars (1598), issued from the database Tercomp.

stable over time and thus can be used as a basis for constructing the graph. It should be stressed that while the Tercomp tools enable most of the plots to be located in spite of errors, gaps or ambiguities in the registers, some of the information remains disconnected from the graph (Fig. 4). When the registers do not include maps, the graphs are incomplete, with a few exceptions. The reproduction quality of the graph obviously depends on the accuracy of the data in the register. In its current version, Tercomp can be used to model fiscal registers whether or not they show the orientation of the *confronts*. However, the results are less good when they do not show the orientation. Currently, the database allows the reconstitution of 50 to 70% of *confronts*, enabling more than 90% of plots to be linked to the related graph, based on the invariants.

1.2 The spatial adjustment of the graphs

Each state of the documentation is translated into a graph. However, the graphs need to be spatially adjusted so that they can be compared in order to measure the spatial dynamics. The graphs based on the Napoleonic cadastral map (1811) and the *terrier* map of 1759 pose no adjustment problem, in that the cartographic information base allows very precise geolocation of each peak of the two graphs. This is not the case for the graphs based on information taken from registers which do not have maps (Fig. 4). In order to make these graphs spatially comparable to the former, we attempted to georeference them using invariant geolocation elements as landmarks for all or part of the period covered by the document. These landmarks include:

Point features (e.g. church, chapel, cemetery, castle, crossroad or remarkable house). This is a very mixed category and can only be used in conjunction with the two following categories.

Linear features (e.g. permanent stream or river, or significant road). This category, comprising essentially linear elements, has the advantage of covering the whole of the studied space evenly. However, it has two disadvantages. First, the linear features are modelled using a single point which generates concentrations of *confronts* in a spider map. Secondly, the same linear element can be designated in different ways in one register. This leads to doubles, requiring thorough sorting of these data.

Surface features: toponymic entity⁶ (e.g. *communes*, parishes, seigneuries, etc.); neighbourhood entity (e.g. communal, demesnes - in the case of *compoix*). This category is also homogeneous but has the disadvantage of the impermanence of place names. For example, in Odars, only 4% of the place names mentioned in 1476 are found in the Napoleonic cadastre. Furthermore, in the large communes, it is not unusual to find several identical microtoponyms, or ones which have been further subdivided.

The spatial adjustment of graphs based on registers without maps was simulated in an experiment on graphs based on the cadastre of 1811 and the *terrier* with map of 1759 (Fig. 5). The adjacency matrices of these two graphs, which are perfectly fixed in space, were exported into a graph-editing programme (YED) to remove all spatial information, in other words a similar situation to the graphs drawn from registers with no maps. These “aspatial” graphs were then spatially adjusted using the available landmarks. This resulted in distorted graphs, as can be seen in Figure 5. The distortion can be measured by comparing the aspatial graph of the 1759 *terrier*, readjusted using landmarks. This distortion shows the extent of the difficulty of this

⁶ “Toponymic entity” refers to a diffuse area within which a microtoponym is concentrated. For each step, it is possible to draw up a list of these toponymic entities, showing features that appeared or disappeared, changes in size or position, etc. (Vue 1997). These toponymic entities play an essential role during the construction of graphs from registers without maps when resolving ambiguities.

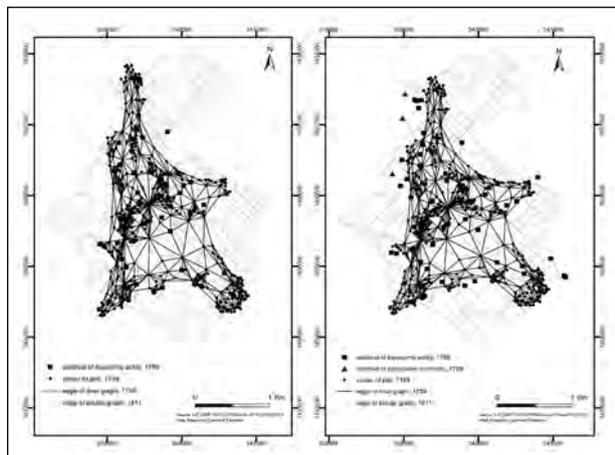


Figure 5. Two spatial adjustment tests of the dual graphs taken from the adjacency matrix of the plot plan of Odars (31, France) dated 1759 with a spring algorithm (on the left, adjustment based on place-name entities; on the right, based on place-name entities and neighbouring villages).

spatial adjustment phase. The aspatial graphs have topological distortions which make comparison tricky. A comparable protocol was set up for the registers without maps. The matrix of neighbouring plots is taken directly from the Tercomp database. It is also possible to take into consideration the surface area of the plots when this is known, and to show the orientation of the links joining two plots using colour coding. Once these points have been fixed, the data can be adjusted by setting the Force-directed algorithm (Brandes 2001; Di Battista et al. 1999) which controls the graph display. While technical improvements are still possible, particularly by improving the inclusion of the *confront* orientation in the graph display, they will not lead to a real spatialization of data; but this is not the aim. Our aim is first to match the dual graphs of plots generated from sources as different as registers and maps, and secondly to make them comparable, or at least to identify the patterns revealing characteristic spatial organizations.

2. Comparing Graphs: Some Possible Lines of Action

The Modelespace team is currently working on methods for comparing the graphs. Our aim here is thus not to present a finished result but the various lines of action considered to achieve a valid method of measuring the dynamics of land use. It is

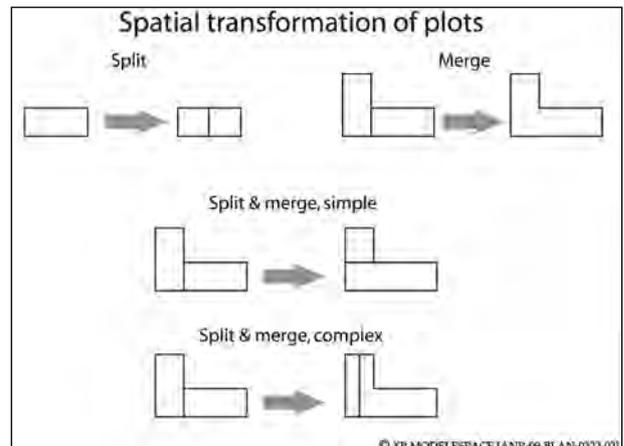


Figure 6. Spatial transformation of plots.

not possible to compare graphs at the plot scale from one period to another, even with the most complete and precise restitution from a register. Coherent sub-graphs between the invariant points need to be drawn. The aim is to characterize the spatial organizations of these sectors by finding patterns that represent the plot structure in the graphs.

2.1 Creating a catalogue of plot patterns

In order to analyze land-use dynamics by comparing graphs of successive states of the same area, it is essential first to create a catalogue of spatial transformations and their graphic representation. This lexicon of transformations provides the first interpretation tool.

In brief, the following cases can be found (Fig. 6):

- Stability: no perceptible change between t_1 and t_2
- Splitting: a plot at t_1 is split into n fragments at t_2
- Merging: several plots at t_1 have been combined to form a single plot at t_2
- A combination of splitting and merging, leading to a relative stability of the plot system: plots at t_1 are reorganized, leading to a shift in the boundaries at t_2 without any major change of the plot structure,
- A complex combination in which plots simultaneously split and merge: at t_1 , plots are

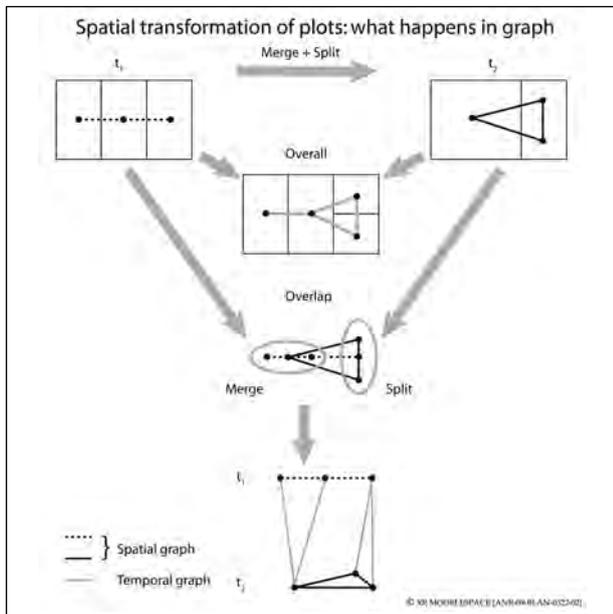


Figure 7. Cumulative graph: general principle.

reorganized, leading to a shift of boundaries at t_2 whereby the plot system is completely transformed.

2.2 The cumulative graph: general principle (Fig. 7)

One comparison method involves cumulating the data of the two graphs t_1 and t_2 to create a third graph that can help interpret the transformations between t_1 and t_2 . This is based on the principle of breaking the space down into the number of entities that are produced when several states of plans are combined in a overall graph. This principle gives good results when analyzing spatial transformations from maps. By contrast, tests carried out from the dual graphs of plot plans indicate that identifying changes in the overall graph is too complex, and the method has thus been abandoned.

2.3 The spatio and temporal graph

This approach, taken from work carried out by GIS specialists (Del Mondot et al. 2010), is based on the principle of analyzing plot transformations over time. The relationships between the peaks of graphs t_1 and t_2 are used to create a temporal graph in which each edge is interpreted according to the transformation lexicon (stability, splitting, merging). It should be stressed that this technique is

only applicable when the relationship between plots from one state to another can be established, which is rarely possible for documents prior to the 18th century. The experiment is currently being carried out from graphs drawn from the maps dated 1759 and 1811. In this case, the relationships have been established spatially by overlapping the two states.

3. Conclusions

The results achieved so far enable the creation of graphs, either from maps or from matrices in the registers. For maps, this involves simply a protocol of extraction from the dual graph. By contrast, for registers, reconstruction of the graph required developing the Tercomp database comprising disambiguation tools. For Odars, this enabled us to produce graphs that are more than 90% complete.

This initial phase having been completed, work now involves comparing graphs using several approaches that are currently being explored. It has already been shown that the relevant scale of comparison cannot be the plots, and that sectors delimited by invariant points need to be identified.

In the short term, analyses based on the graph structure will be completed by statistical exploration of the attributes attached to the peaks: surface area, type of crop, owner's name, place name, etc. The aim is to identify significant indicators, alone or in combination, of spatial organizations for each state, in order to measure the dynamics of landscape change.

References

Atkin, R.H. 1977. *Combinatorial connectivities in social systems: an application of simplicial complex structures to the study of large organizations*. Basel: Birkhäuser.

Babson, E., H. Barcelo, M. de Longueville, R. Laubenbacher. 2006. "Homotopy theory of graphs." *Journal of Algebraic Combinatorics* 24: 31-44.

Barabási, A.-L., and R. Albert. 1999. "Emergence of scaling in random networks." *Science* 286: 509-512.

Berge, C. 1958. *Théorie des graphes et ces applications*. Paris: Dunod.

- Berge, C. 1970. *Graphes et Hypergraphes*. Paris: Dunod.
- Biget, J.-L., Hervé, J.-C., Thébert, Y. 1989. *Les cadastres anciens des villes et leur traitement par l'informatique*. Rome: École française de Rome.
- Boulet, R., Fieux, E., Jouve, B. 2010. "Simplicial simple-homotopy of flag complexes in terms of graphs." *European Journal of Combinatorics* 31: 161-176.
- Brandes, U. 2001. "Drawing on physical analogies." In *Drawing Graphs*, edited by M. Kaufmann and D. Wagner, *Lecture Notes in Computer Science*, 2025: 71–86.
- Brunel, G., O. Guyotjeannin, O., and J.-M. Moriceau. 2002. *Terriers et plans-terriers du XIIIe au XVIIIe siècle. Actes du Colloque de Paris (23-25 septembre 1998)*. Caen-Paris: Association d'Histoire des Sociétés Rurales/École Nationale des Chartes.
- Chen, B., S.-T. Yau, and Y.-N. Yeh. 2001. "Graph homotopy and Graham homotopy." *Discrete Mathematics* 241 (1-3): 153-170.
- Claveirole, A., and E. Pélaquier. 2001. *Le compoix et ses usages. Actes du Colloque de Nîmes - 26 et 27 novembre 1999*. Montpellier: Publications de l'Université Paul Valéry-Montpellier 3.
- Conesa, M. 2007. "Capbreu et paysage. Remarques sur l'utilisation d'une source notariale dans l'étude des paysages des Pyrénées de l'est. (Cerdagne, XVIe-XVIIIe siècle)." *LIAME, Bulletin du centre d'histoire moderne et contemporaine de la Méditerranée et de ses périphéries* 14: 97-124.
- Del Mondo, G., J. G. Stell, C. Claramunt, and R. Thibaud. 2010. "A graph model for spatio-temporal evolution." *Journal of universal computer science* 16 (11): 1452-1477.
- Di Battista, G., P. Eades, R. Tamassia, and I. G. Tollis, I. G. 1999. *Graph Drawing: Algorithms for the Visualization of Graphs*. Englewood Cliffs, NJ: Prentice Hall.
- Dochterman, A. 2009. "Homotopy groups of hom complexes of graphs." *Journal of Combinatorial Theory* 116(1): 180-194.
- Le Couédic, M., S. Leturcq, X. Rodier, F. Hautefeuille, B. Jouve, and E. Fieux. *In press*. "Du cadastre ancien au graphe. Les dynamiques spatiales dans les sources fiscales médiévales et moderne." *Archéosciences* 36.
- Leturcq, S. 2007. *Un village, la terre et ses hommes. Toury en Beauce (XIIIe-XVIIe siècle)*. Paris: CTHS.
- Montpied, G., and J. Rouault. 1982. "Reconstitution et cartographie d'un parcellaire à partir de cadastres textuels du XVe siècle." *Le médiéviste et l'ordinateur* 8: 11-13.
- Montpied, G., and J. Rouault. 1989. "Du texte au graphe: établissement d'une carte du parcellaire à partir des données de deux cadastres de la fin du Moyen Âge." In *Les cadastres anciens des villes et leur traitement par l'informatique*, edited by J.-L. Biget, J.-C. Hervé, and Y. Thébert, 359-380. Rome: École française de Rome.
- Pirot, F. 2004. "Glossaire." *Histoire et mesure. Dossier Système d'information géographique, archéologie et histoire* 3-4 (XIX): 422-427.
- Poirier, N. 2010. *Un espace rural à la loupe. Paysage, peuplement et territoires en Berry de la préhistoire à nos jours*. Tours: Presses universitaires François-Rabelais.
- Truel, Y. 2007. "Représentation fiscale au XVIIIe siècle: le Haut-Ségala quercynois. Le compoix de Saint-Cyrgues (1636)." Master Dissertation, Université de Paris I Panthéon-Sorbonne.
- Vue, B. 1997. "Microtoponymie et archéologie des paysages à Neuilly-L'évêque (52) du 13e au 20e siècle : comportement des microtoponymes au fil des siècles: le nom, l'espace et l'homme qui le nomme." PhD diss. Université de Nancy 2.
- Watts, D. J., and S.H. Strogatz. 1998. "Collective dynamics of "small-world" networks." *Nature* 393: 440-442.

Exploring the Effects of Curvature and Refraction on GIS-based Visibility Studies

Mariza Kormann and Gary Lock
University of Oxford, United Kingdom

Abstract:

The curvature of the earth affects visibility studies when using a DEM (Digital Elevation Model) and needs to be taken into consideration as pointed out by Conolly and Lake in 2006, although little work has been done in archaeology to explore the implications of this warning. This paper addresses related methodological issues in visibility studies namely the issue of reality, of edge effects, and of significance. We use the long barrows of the Danebury region in England as a case study. New methods to calculate and assess the combined effects of curvature and atmospheric refraction on visibility studies are presented. It is shown that within a visibility range of 10km there is a decrease in visibility of approximately 20% for the Danebury landscape; the effect is larger at areas closer to the horizon.

Keywords:

GIS, DEM, Visibility, Earth Curvature, Atmospheric Refraction

1. Introduction

The justification and attractiveness of performing visibility studies using GIS is that it is an attempt to model human visual perception and understand situated visual experiences and their significance to past landscape settings (Wheatley 2004). Much work on visibility has focused on the placement of monuments in the landscape. Archaeological features may be several kilometres away from the observer and the drop in visibility may severely affect how monuments are perceived at distance, especially on approach. The motivation for visibility research is that the analysis of visibility patterns can shed light on human occupation and landscape use and exploitation, and the relative importance of selected landscape regions and features. Examples of archaeologically-oriented visibility research include a survey of GIS and non-GIS visibility studies with particular application to stone circles in the Grampian region of Scotland (Lake and Woodman 2003), archaeology and GIS-based spatial technologies (Lock and Harris 2000), studies of visibility prominence of monuments (Llobera 2001), and an investigation of whether or not the siting of the Bronze Age cairns of North Mull Island in Scotland have a causal influence of their visible area (Fisher *et al.* 1997).

Despite widespread acceptance of GIS as a tool for visibility analysis, there has been justified

Corresponding author: mariza.kormann@kelllogg.oxon.org

criticism of early attempts at GIS-based analysis as it encourages objectivism and ignores subjective experience (Wheatley 1993; Gaffney and van Leusen 1995; Brück 2005:54). GIS-based methods for visibility and movement analysis have since developed to include for instance, cumulative viewsheds, Higuichi and directional (Wheatley and Gillings 2000), surface approaches (Lock *et al.* In press) and Fuzzy Cumulative Visibility Analysis (FCVA – a form of restricted total viewshed) (Kormann and Lock In press).

However, a number of theoretical and methodological issues concerning visibility studies still remain and have been pointed out by many authors over the last decade. In particular, van Leusen (1998) has discussed the problems of reality, of edge effects, and of significance. Van Leusen defines the issue of reality as being concerned with how accurately the data reflect the reality in the landscape, or how close the calculated viewshed is to the real viewshed? The obvious factors influencing this definition of reality are related to the quality of the DEM and the level of sophistication in the algorithms used to calculate visibility.

Concerning the quality of the DEM the only reasonable approach is to select the best available and being aware of intrinsic inaccuracies. It is often accepted that a DEM with 10m cell would yield adequate resolution for visibility calculations (Lock, Kormann and Pouncett, in press). Concerning the quality of the algorithms used to calculate visibility

maps, the ideal would be that algorithms should take into consideration the curvature of the earth, the effects of atmospheric refraction, the effects of light scattering on airborne particles, light absorption and uneven illumination. These combined factors affect visibility of distant points in such a way as to deviate considerably from simple calculations that are only based on DEM heights, as shown below.

There are two overriding concerns related to building accurate models of atmospheric parameters: first, it is difficult to model atmospheric refraction precisely as particles' multiple scattering effects and absorption together with atmospheric particle composition are not precisely known and second, illumination changes continuously. Therefore, the reduction in atmospheric contrast affecting visibility of distant points is an elusive measure and can only be estimated. Attempts at modelling refraction and single scattering effects are described in French (1981) and Bohren and Fraser (1986). The algorithms discussed in this paper are based on the models of the earth's curvature and simplified refraction model described by French (1981) and Young (2012). It is stressed here that atmospheric light scattering and uneven illumination are not included in the models proposed in Section 2.

The issue of edge effects is here addressed in several ways. First, when calculating visibility for cells that lie close to the boundaries of the DEM, virtual cells are considered within the given radius such that valid visibility indices are calculated up to the boundary cells. This negates the need for a buffer region, which is the standard solution to edge effects, as adding virtual cells works as a buffer and the fall off effect is eliminated. Second, by counting the cells and calculating correct percentages being careful not to include cells in the calculations when visibility radii exceed the boundaries of the study area, that is, cells that lie in the virtual buffer do not carry statistical value. For example, if a particular viewshed partially falls outside the DEM then the percentage of visible cells must only refer to the total number of cells that fall inside the DEM within that viewshed. Third, it is proposed in this paper that the earth's curvature and refraction correction be applied in a way that caters for edge effects. This is achieved by calculating a correction matrix that is 4 times larger than the actual DEM (twice the width and twice the height of the DEM). The method is

described in Section 2 and ensures a consistent way to adjust the relative heights of each cell in relation to the observer point.

Finally the last of the three theoretical and methodological issues raised by van Leusen (1998) is the issue of significance, which is related to how significant the results are when compared to general properties of the landscape. It is clear that using visibility to infer the merits of a particular site location one has to directly compare with wider statistical properties of the landscape. The choice of visibility radius can lead to misinterpretations as a small radius favours a higher percentage of cells being located at lower elevations relative to the observer, while a larger radius favours cells located at higher elevations (the peaks and ridges effect). The 'preferential' visibility of peaks and ridges for larger radii are documented in the literature, e.g. (Wheatley 1995) and (van Leusen 1998). In order to address these issues, it is proposed here that whenever visibility studies are conducted several radii should be used; this can be described as generating visibility within bands where each band is defined by a given radius. Then for each band independent statistics can be collected providing extra levels of information per visibility band that would not be available if a single radius were to be used.

Addressing the methodological issues above, this paper proposes methods to calculate and assess the combined effects of curvature and atmospheric refraction on visibility studies. The models are described in Section 2. Section 3 describes a case-study using the long barrows of the Danebury region in England and Section 4 attempts to determine patterns in the data. Finally a discussion and conclusion is presented in Section 5.

2. Correcting for the Curvature of the Earth and Atmospheric Refraction

2.1 Earth curvature

The Earth's curvature and atmospheric properties significantly affect visibility range and must be taken into consideration as pointed out by Ruggles and Medycky-Smith (1996) and Conolly and Lake (2006). A detailed analysis has been carried out by French (1981) of the geometry of

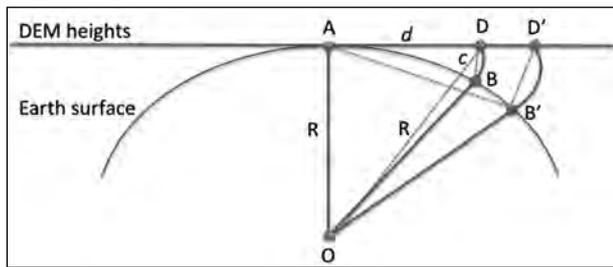


Figure 1. The curvature correction values are defined by the distance BD (not to scale, curvature is exaggerated).

visibility as a function of altitude and atmospheric refraction. The studies focus on how far can the horizon be seen, assuming that there is enough contrast to the naked eye between the horizon and the sky. An approximate correction for refraction is proposed in which it increases the distance to the visible horizon by about 9%. The analysis is valid up to a vertical distance of about 10km.

Bohren and Fraser (1986) extended the analysis to consider the effects of light scattering by particles in the atmosphere. Their analysis was focused on determining at what distance the horizon would cease to be visible, as it is clear that even on extraordinarily clear days it is not possible to see forever due to contrast reduction. Strong assumptions render their model a very simple one, as it assumes no multiple scattering effects, constant illumination, and ignores light absorption by particles in the atmosphere. However, the results are plausible and agree with the observation that for long distances and high altitudes there is not enough contrast between earth and sky to accurately determine where the horizon lies. The conclusions are that the exact location of the horizon is not possible to determine for altitudes greater than 10km.

While such analysis highlights the role of atmospheric contrast reduction for both long distances and high altitudes due to uneven illumination, light scattering and absorption, it also indicates that these factors can be lumped into a single contrast reduction correction factor. The analysis and the models of curvature and refraction presented below derive from the work described in French (1981) with contrast reduction not considered – the justification being that a contrast reduction factor as defined by Bohren and Fraser (1986) can be applied to the results from

curvature and refraction, if desired. That would be an acceptable replacement for an otherwise difficult to model effect. Similar analysis is also presented in Young (2012) in which curvature and refraction are lumped together.

DEM heights are normally defined on a rectangular grid and, if one extends the grid to represent the entire Earth's surface then clearly this would point to a flat Earth. Although incongruous, this would mean that on a visibility study using a GIS if the observer were on the highest point on Earth, then it would be possible that even the furthest cells on the Earth surface could be visible. The flat Earth surface model is depicted in Figure 1 by the horizontal line labelled "DEM heights".

In Figure 1, R is the radius of the Earth and the observer is at location A looking at cell D . However, the cell D is not at its apparent position on a flat earth, but on the surface of the earth at location B . Therefore, the distance BD is the amount the cell has to be lowered in relation to observer A and it is a function of the distance AD . It is clear that as D increases towards D' the location becomes occluded by the curvature of the earth and this is represented by point B' . The average radius of the earth (R) is known and the distance $d = AD$ is the distance from the viewpoint to a given cell. Given that the angle between AD and AB is very small and that the distance BD is very small compared to the radius of the earth, it is reasonable to consider the right triangle $OA-AD-DO$ in Figure 1 and call the correction factor $c = BD$ and express the relationship by Pythagoras' theorem:

$$R^2 + d^2 = (R + c)^2 \quad (1)$$

from which c can be calculated as:

$$c = \sqrt{(R^2 + d^2)} - R \quad (2)$$

As visibility is calculated from every cell location, it would be time-consuming and unpractical to calculate new values of c for all possible different viewpoints. To solve this problem a method is proposed to pre-calculate the corrections that must be applied to each height of the DEM placing these correction factors into a matrix. The cell at the centre has a correction value of 0 (zero) while all other cells have a correction factor that increases

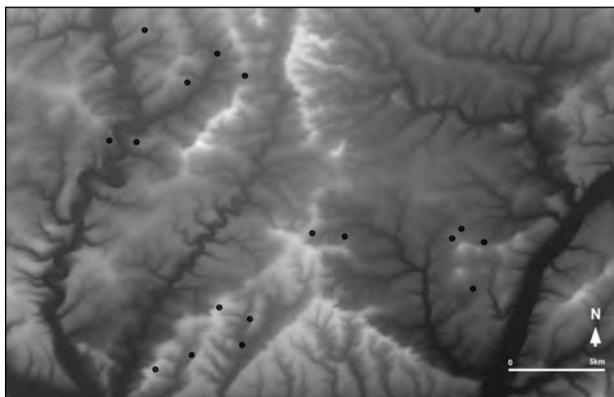


Figure 2. The Danebury region topography map showing the location of the 18 long barrows. Lighter areas are higher elevations.

with the distance from the viewpoint. These factors are constant and only need to be calculated once (from the Earth's radius at the required latitude).

2.2 The effects of atmospheric refraction

The Earth's atmosphere works as a lens with a bending effect – the result is that this enables us to see further from any viewpoint (Young 2012). From Figure 1, consider that an observer located at the position marked A would see B as the furthest point in the horizon if only the earth's curvature effects were considered. Because the atmospheric lens effect effectively 'raises' some points that lie beyond the horizon, now point B' becomes visible, although the geometry implies that it is not visible. The difference between B'–B is the effect of atmospheric refraction. This is normally very difficult to model as it depends on atmospheric composition and on a number of other parameters that vary with the time of the day, brightness, altitude, location, and so on. An approximate model has been proposed (French 1981; Young 2012) and it is adopted here. The model states that the increased visibility due to atmospheric refraction can be incorporated into the model of the Earth's curvature by simply adding to the Earth's radius $1/6$ of its value:

$$R' = R + R/6 \quad (3)$$

This enlarged radius R' is used to calculate a correction matrix C_M which contains the values of c to be applied to each cell in the DEM as a function of its distance d to the observer cell. The C_M matrix is calculated through Equation (2) where both

curvature of the earth and atmospheric refraction are calculated in a single step. Because both are corrected at the same time, the expression *curvature correction matrix* refers both to the curvature and refraction correction matrix.

The Neolithic long barrows in the Danebury region of central England are used here as a case study (Lock and Harris 1996). The DEM for the Danebury region is depicted in Figure 2 with 2000×3000 cells with cell size of 10m, so the distance d is defined at every 10m interval. The values of c are pre-calculated and saved to the correction matrix C_M . In order to cater for edge effects the size of the correction matrix is set to 4000×6000 cells, which is four times as large as the original DEM – this will ensure that when the correction matrix is centred on any feature of the DEM, all cells of the DEM will be covered by the matrix. Figure 3 shows a graphical representation of a pre-calculated matrix C_M where the point at its centre represents a minimum correction value, which increases towards the boundaries. The value at the centre of matrix C_M will be zero (because $d=0$ at this point) and will increase towards the borders of the matrix in all directions. For instance, for a cell 30km away from the centre the correction will be 70m for the Danebury landscape (this is calculated through equation 2 with $D=30,000$).

Integral to this approach is what is the maximum distance that can be seen from a viewpoint at 1.60m (the eye level of a person) and this can be calculated simply as the distance d when $c=1.60m$. The Earth is not a perfect sphere and the radius at the poles is smaller than the radius at the Equator (6,356km and 6,378km respectively). To get an accurate representation of the Earth's radius for the Danebury region at latitude 51.133281 degrees its value was estimated as 6,365km. If the terrain were a perfectly flat plane such as at sea, the maximum distance that could be seen from a viewpoint at 1.60m is then calculated as $d=4,513m$.

If the effects of refraction are now added by using $R'=7,426km$ and the previous calculation is repeated the result of $d=4,875m$ is obtained. With this simple example for a viewing point at a height of 1.60m, it is shown that the effects of atmospheric refraction increase the viewing distance by 8% when compared to curvature alone. This agrees with results reported in the literature (French 1981).

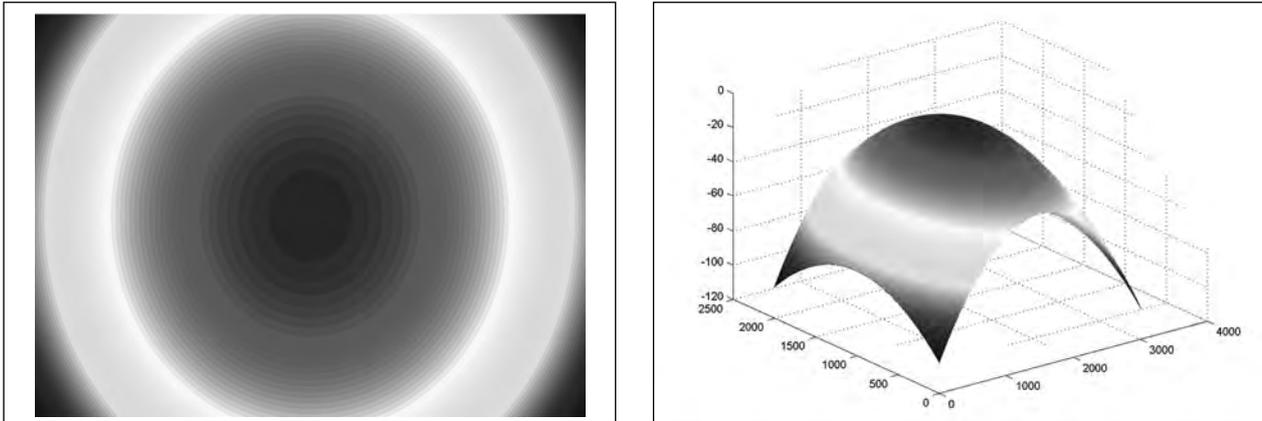


Figure 3. Representation of the curvature plus refraction correction matrix C_M . Regions near the centre show small corrections, away from the centre larger corrections (not to scale, curvature effects are exaggerated).

2.3 Applying the correction matrix to DEM heights

Once the correction matrix C_M is obtained as above, it is used to generate a new DEM for each location from where visibility is to be calculated. The C_M matrix must first be overlaid onto the DEM matrix such that its centre is exactly on the cell of interest. Since the DEM and C_M matrices have different dimensions (the C_M is four times larger than the DEM), C_M must be cropped to the same size as the DEM:

$$C_{M(\text{cropped})}[\text{Rows}, \text{Columns}] = C_M[(R_{\text{DEM}} - r : R_{\text{DEM}} - r + R_{\text{DEM}}), (C_{\text{DEM}} - c : C_{\text{DEM}} - c + C_{\text{DEM}})] \quad (4)$$

Where (r,c) are the rows and column indices of the location where visibility is to be calculated, (:) means the range of rows and columns (from:to), and $(R_{\text{DEM}}, C_{\text{DEM}})$ are the total number of rows and columns of the DEM. Then a new, corrected DEM can be generated for that location:

$$\text{DEM}_{\text{new}} = \text{DEM} - C_{M(\text{cropped})} \quad (5)$$

The corrected DEM_{new} is then used to generate viewsheds. Note that each cell of interest will have its own corrected DEM through Equations 4 and 5, but the generation of such DEMs only involves a simple cropping and a matrix subtraction operation.

The proposed method offers a number of advantages over existing methods available in ArcGIS for instance. Here, only a single correction matrix is calculated for any given latitude; then for each cell in the DEM (i.e. for each observer point) to generate a new, corrected DEM centred

at the observer point is just a matter of cropping (Equation 4) and a matrix subtraction (Equation 5). In contrast, in ArcGIS for each observer point all corrections are calculated anew for every single cell making it much more computationally intensive than the method proposed here. In addition, the proposed method estimates the Earth's radius at the required latitude, while in ArcGIS an average is used. This makes the proposed computations of curvature and refraction more accurate as they are performed using a more accurate Earth radius. For instance, for the Danebury region using the proposed method (implying using the Earth radius for the Danebury latitude) yields a more precise refraction correction index of 0.14 while in ArcGIS this is 0.13, a constant value that is independent of location on the Earth's surface. Finally, the proposed method is fully controlled by the user when creating the required corrected DEMs, while in ArcGIS it is a black box approach.

3. Assessing the Impact of Curvature and Refraction on Visibility

In order to assess the relative impact of the Earth's curvature and atmospheric refraction on visibility the following method was applied. First, define visibility bands for radii from 1 to 10km and calculate viewsheds using standard viewshed algorithms for each long barrow in the Danebury study region (18 barrows in total). Repeat the calculations using the algorithms for correcting for earth curvature and atmospheric refraction as defined in the previous section. Second, produce statistics by counting the number of visible cells

Number of visible cells: 18 Long barrows						
Visibility bands (km)	Standard viewshed		Curvature with refraction		Change in visible cells	Percentage change
	viewshed	Percentage	refraction	Percentage		
0.0 to 1.0	1449460	13%	1412902	13%	36558	3%
1.1 to 2.0	1191052	11%	1155786	10%	35266	3%
2.1 to 3.0	864387	8%	834940	8%	29447	3%
3.1 to 4.0	670912	7%	634964	6%	35948	5%
4.1 to 5.0	523145	5%	493708	5%	29437	6%
5.1 to 6.0	506543	5%	471655	5%	34888	7%
6.1 to 7.0	428990	5%	395647	4%	33343	8%
7.1 to 8.0	309588	4%	277797	3%	31791	10%
8.1 to 9.0	224704	3%	191199	3%	33505	15%
9.1 to 10.0	189460	3%	155154	2%	34306	18%

Table 1. Summary statistics for visible cell counts for 18 long barrows.

Number of visible cells: 18 random locations						
Visibility bands (km)	Standard viewshed		Curvature with refraction		Change in visible cells	Percentage change
	viewshed	Percentage	refraction	Percentage		
0.0 to 1.0	947518	8%	921287	8%	26231	3%
1.1 to 2.0	761838	7%	734491	6%	27347	4%
2.1 to 3.0	588188	5%	560087	5%	28101	5%
3.1 to 4.0	376884	3%	353348	3%	23536	6%
4.1 to 5.0	344982	3%	318540	3%	26442	8%
5.1 to 6.0	250260	3%	230884	3%	19376	8%
6.1 to 7.0	152549	2%	134406	2%	18143	12%
7.1 to 8.0	138369	2%	108932	1%	29437	21%
8.1 to 9.0	102884	1%	83945	1%	18939	18%
9.1 to 10.0	101047	1%	85012	1%	16035	16%

Table 2. Summary statistics for visible cell counts for 18 random locations.

at each visibility band for both standard viewshed and with curvature and refraction algorithms. Edge effects are taken into account by standardizing results to percentages and being careful to use the correct total number of cells in cases where any of the radii go over the borders of the study region.

The significance of the results needs to be compared to visibility properties of the general landscape. Since there are 18 long barrows in the Danebury region, visibility for further 18 random locations are also calculated to give a uniform sample size. Results are shown in Tables 1 and 2.

Tables 1 and 2 provide a summary of visible cell counts for both barrows and random locations. The columns under “Standard viewshed” show the number of visible cells from standard viewshed calculations with percentages for each band. The columns under the label “Curvature with refraction” show the effects of the earth’s curvature and refraction within each band; note that the number of

visible cell count drops in each and every band. The columns “Change in visible cells” and “Percentage change” are calculated by the number of change in visible cells in relation to the standard viewshed

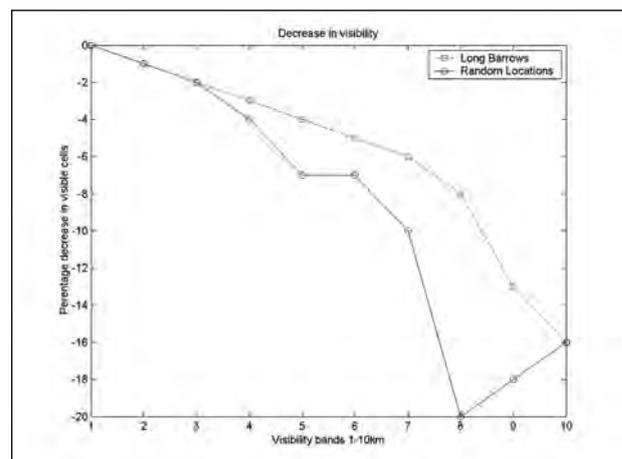


Figure 4. Average percentage decrease in visible cells per visibility band. Dashed line: long barrows; solid line: random locations.

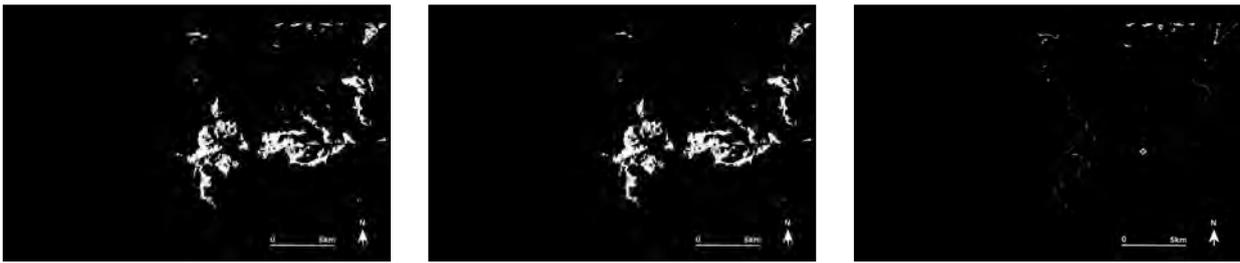


Figure 5. The effects of visibility reduction due to earth curvature and refraction for barrow 10. Left. visible cells calculated with standard viewshed. Middle. with curvature and refraction. Right, the difference or the actual cells that go out of view due to curvature and refraction.

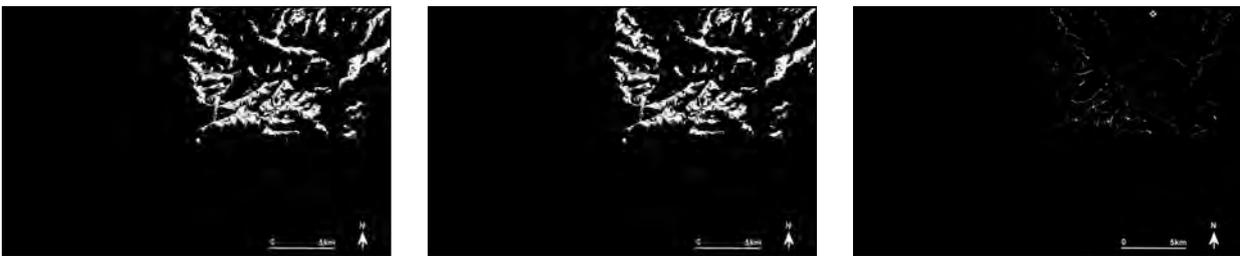


Figure 6. Barrow 18. Left. standard viewshed. Middle. with the effects of curvature and refraction. Right. the cells that go out of view due to curvature and refraction.

calculation. This is to assess the significance of the drop in visibility when the combined effects of Earth curvature and atmospheric refraction are taken into account.

It is apparent from comparing the two tables (the “Percentage” columns) that long barrows are located in areas of higher visibility than random locations (probably the expected result). Overall, the impact of the Earth’s curvature and atmospheric refraction on visibility reduction is significant when compared to standard visibility cell counts as shown by the last column “Percentage change”. The reduction in the number of visible cells becomes more significant for points further away from the viewing point; in other words, it has a large effect on cells closer to the horizon. This is shown for the larger radii where the reduction is up to 21% for random locations and 18% for barrow locations. This effect is illustrated in the summary statistics of Figure 4 (taken from Tables 1 and 2), where on the x-axis is shown the 10 visibility bands, and on the y-axis the average decrease in visibility for all barrows (dashed curve) and all random locations (solid curve).

A visual representation of visibility decay due to curvature and refraction is illustrated in Figures

5 and 6 for barrows 10 and 18 respectively. The images on the left show the viewsheds calculated using the standard method, and in the middle, using the method described in this paper with curvature and refraction effects. The right images show the actual cells that go out of view when the curvature and refraction effects are taken into account. It is noticeable that as the visibility bands increase (and thus cells are closer to the horizon) the effect is more prominent, but it also affects closer cells. These results clearly quantify the reduction in visibility as a function of the distance from the observer providing evidence for the effects of visibility decay and strongly indicate that these effects should not be ignored.

4. Patterns of Visibility

To explore the effects of how each visibility band changes across all barrows and random locations, the bands were plotted independently as shown in Figure 7 for standard viewshed (left column) and with the combined effects of curvature and refraction (right column). The lighter coloured round markers are for the 18 barrows and the dark square markers for 18 random locations. The plots on the top row show the 10 visibility bands on the x-axis and the percentage of visible cells on

Exploring the Effects of Curvature and Refraction on GIS-based Visibility Studies
 Mariza Kormann and Gary Lock

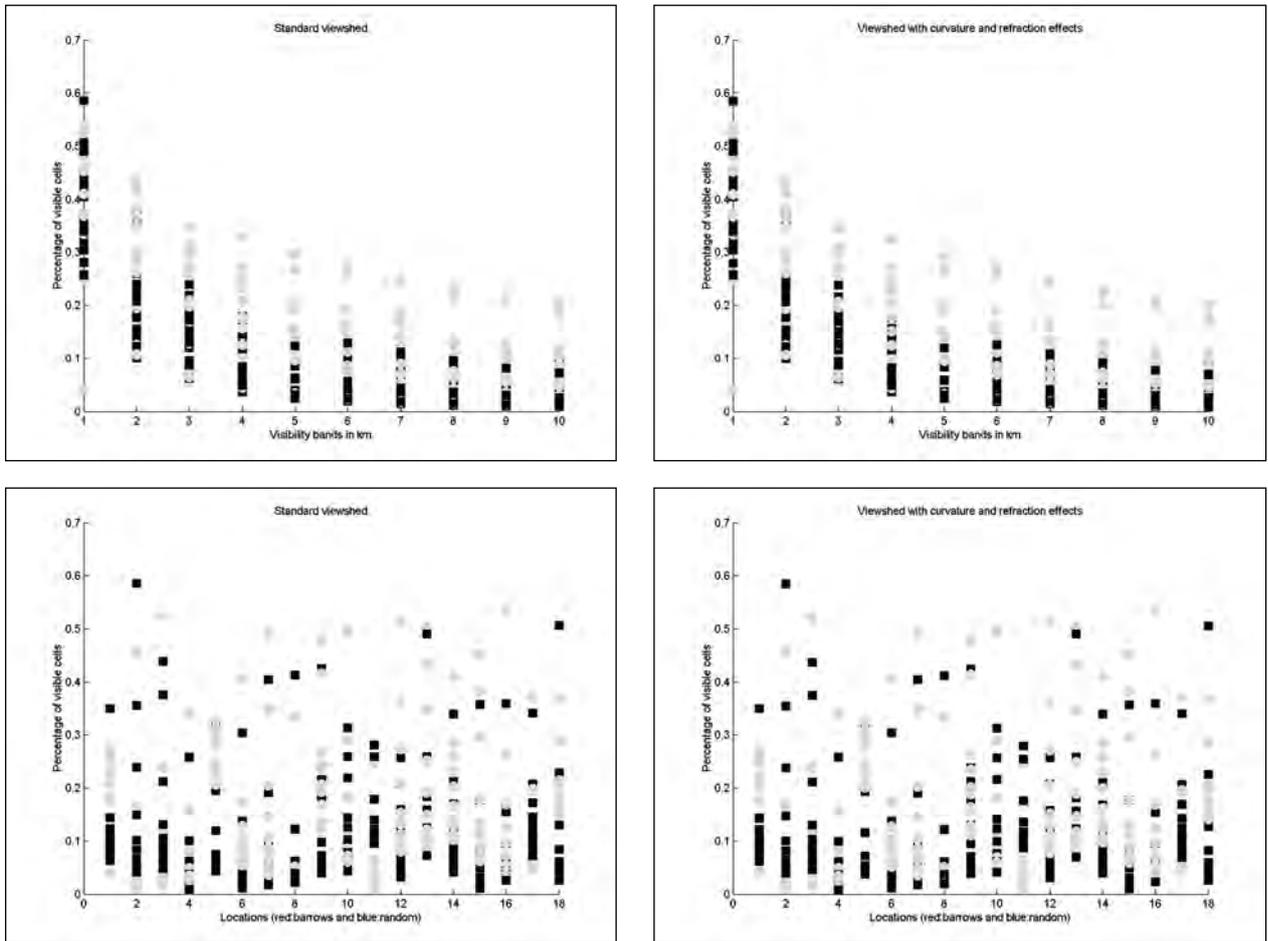


Figure 7. Top row. percentage of visible cells per visibility band for barrows (light round markers) and random locations (dark square markers). Bottom row. percentage per location. Left column. standard visibility, right column. with the effects of curvature and refraction.

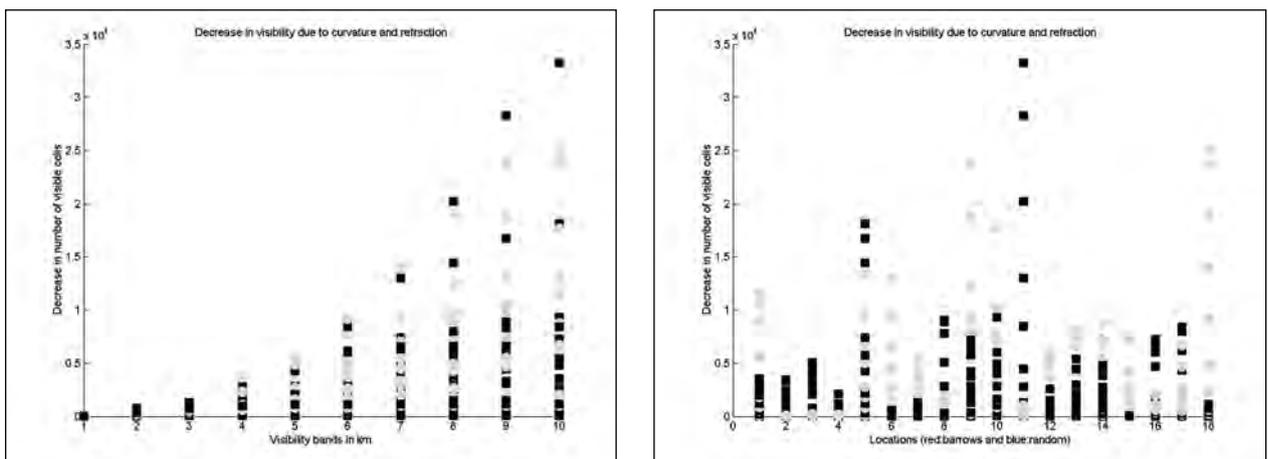


Figure 8. Decrease in the number of visible cells. Left, decrease per visibility band, right, decrease across all locations. Light round markers. long barrows, dark square markers. random locations.

the y-axis. The plots on the bottom row show 18 locations on the x-axis (for barrows and random locations) and the percentage of visible cells on the y-axis. Note that for all bands the long barrows have a significantly higher visibility than the random locations, the generally expected result based on the assumption that barrow locations were selected for their higher visibility.

When comparing the plots on the left column of Figure 7 against the ones on the right column, the decrease in overall visibility due to curvature and refraction is not clearly apparent. Although the decrease is significant in a number of cells as demonstrated in Tables 1 and 2 and in Figures 5 and 6, such decrease in relation to the overall number of visible cells is more clearly seen on the plot of Figure 8 left. The vertical axis of Figure 8 shows the number of cells that go out of view due to curvature and refraction for each visibility band and across all locations. The effects are more pronounced for larger bands and clearly cannot be ignored – a point that has been made earlier and is reinforced here.

5. Discussion and Conclusions

This paper has addressed methodological and theoretical issues in visibility studies by using the long barrows of the Danebury region in England as a case study. A new method is demonstrated in which corrections to the DEM heights to cater for the Earth's curvature in relation to the observer point are pre-calculated making it highly computationally efficient. These corrections are adjusted to a local landscape by using the earth's average radius at the given latitude.

The paper has addressed the issue of reality in visibility studies as defined by van Leusen (1998). The difficulties of accurately modelling the combined effects of atmospheric refraction, multiple light scattering and absorption by airborne particles together with uneven illumination have been pointed out. Both the Earth's curvature model together with an approximate model of refraction from the literature (French 1981; Young 2012) were adapted to support curvature and refraction corrected visibility calculations. Experimental results demonstrate that the effects are most noticeable for points towards the horizon; reductions of the order of 20% are observed for the case study for distances

around 10km. These results clearly show that the effects of the earth curvature and atmospheric refraction must be taken into account in visibility studies otherwise the calculated visibility may not correspond to reality. This is of importance within visibility studies generally but particularly where relationships between sites and visible horizons are investigated (for example archaeo-astronomy and stone circles as in Ruggles and Medyckyj-Scott 1996).

The issue of significance was also addressed. In order to compare visibility from selected feature points and draw conclusions in relation to the properties of the wider landscape, this paper has selected 36 locations (18 long barrows plus 18 random locations) and each viewshed was divided into 10 bands, from 1 to 10km. That means visibility maps were calculated 72 times (36 times using a standard visibility algorithm, and 36 times with the effects of curvature and refraction).

The effects of curvature and refraction were quantified and tabulated. Looking for patterns in the data, plotting a combined data over all locations and also for each visibility band proved useful to assess changes in visibility. The effects of reduction in number of visible cells due to curvature and refraction were quantified and visualized – it has been shown that the reduction can be up to 20% for the Danebury data for some visibility bands.

The main thrust of this paper has been concerned with methodological issues in visibility studies by analysing and quantifying the Earth curvature effects on visibility. The paper makes theoretical and methodological contributions in this regard and also addresses the issues of reality, edge effects and significance. The curvature models described here only consider atmospheric refraction, and it is clear that further work should address the modelling of complex atmospheric variables such as multiple particle interaction concerning light scattering and absorption and uneven illumination. Also, further work on quantification includes increasing the sample size for further testing the properties of barrows and landscape and the development of pattern analysis and data visualization methods.

References

- Allen, K. M. S., S. Green, and E. B. Zubrow. 1990. *Interpreting Space: GIS and Archaeology*. London: Taylor and Francis.
- Bohren, C. F., and A. B. Fraser. 1986. "At what altitude does the horizon cease to be visible?" *American Journal of Physics* 54(3): 222-7.
- Brück J. 2005. "Experiencing the past? The development of a phenomenological archaeology in British prehistory." *Archaeological Dialogues* 12(1): 45-72.
- Conolly, J., and M. Lake. 2006. *Geographical Information Systems in Archaeology*. Cambridge: Cambridge University Press.
- Fisher, P., C. Farrelly, A. Maddocks, and C. Ruggles. 1997. "Spatial Analysis of Visible Areas from the Bronze Age Cairns of Mull." *Journal of Archaeological Science* 24 (7):581-592.
- French, A. P. 1981. "How far away is the horizon?" *American Journal of Physics* 50(9): 795-9.
- Gaffney, V., and M. van Leusen. 1995. "GIS environmental determinism and archaeology." In *Archaeology and Geographic Information Systems: a European Perspective*, edited by G. Lock and Z. Stančič, 367-82. London: Taylor & Francis.
- Kormann, M., and G. Lock. *In Press*. *Exploring Differences: Implications for FCVA Visibility Indices and Scales of Analyses*. Oxford: Archaeopress.
- Lake, M.W., and P.E. Woodman. 2003. "Visibility studies in archaeology: a review and case study." *Environment and Planning B: Planning and Design* 30 (5): 689-707.
- Llobera, M. 2001. "Building Past Landscape Perception with GIS: Understanding Topographic Prominence." *Journal of Archaeological Science*, 28: 1005-14.
- Lock, G., and T.M. Harris. 1996. "Danebury revisited: an English Iron age hillfort in a digital landscape." In *Anthropology, Space and Geographic Information Systems*, edited by M. Aldenderfer and H.D.G. Maschner, 214-240. New York: Oxford University Press.
- Lock, G., and T. M. Harris. 2000. "Beyond the Map: Archaeology and Spatial Technologies. Introduction: Return to Ravello." In *Proceedings of the NATO Advanced Research Workshop*, edited by G. Lock, xiii-xxv. Ravello: IOS Press.
- Lock, G., M. Kormann, and J. Pouncett. *In press*. "Visibility and Movement: Towards a GIS-Based Integrated Approach." In *Berlin Studies of the Ancient World*, edited by S. Polla and P. Verhagen. Berlin: De Gruyter Excellence Cluster Topoi.
- Ruggles, C., and D. Medyckyj-Scott. 1996. "Site location, landscape visibility, and symbolic astronomy: a Scottish case-study." In *New Methods, Old Problems. Geographic Information Systems in modern archaeological research*, edited by H. Maschner, 127-146. Carbondale: Southern Illinois University.
- van Leusen, P.M. 1998. "Viewshed and Cost Surface Analysis Using GIS." Accessed June 2012. www.let.rug.nl/arge/Docs/caa98v2.rtf.
- Wheatley, D. 1993. "Going over old ground: GIS, archaeological theory and the act of perception." In: *Computing the Past: Computer Applications and Quantitative Methods in Archaeology*, edited by T. Madsen and I. Scollar, 133-38. Aarhus University Press.
- Wheatley, D. 1995. "Cumulative viewshed analysis: a GIS-based method for investigating intervisibility, and its archaeological application." In *Archaeology and Geographic Information Systems: a European Perspective*, edited by G. Lock and Z. Stančič, 171-186. London: Taylor & Francis.
- Wheatley, D. 2004. "Making space for an archaeology of place." *Internet Archaeology* 15. Accessed June, 2012. <http://eprints.soton.ac.uk/id/eprint/28800>.
- Wheatley, D., and M. Gillings. 2000. "Vision, Perception and GIS." In *Beyond the Map: Archaeology and Spatial Technologies*, edited by Gary Lock, 1-27. Amsterdam: Amsterdam IOS Press.
- Young, A. T. 2012. "Distance to the Horizon." Accessed June 2012. http://mintaka.sdsu.edu/GF/explain/atmos_refr/horizon.html
- Young, A. T. 2012.. "Calculating ray bending." Accessed June 2012. http://mintaka.sdsu.edu/GF/explain/atmos_refr/bending.html.

Process Formalization and Conceptual Modelling in the Study of Territorial Dynamics

Laure Saligny

CNRS, Université de Bourgogne, Dijon, France

Xavier Rodier

Université François-Rabelais, Tours, France

Estelle Gauthier

Université de Franche-Comté, Besançon, France

Nicolas Poirier

CNRS, Université Toulouse 2 - Le Mirail, Toulouse, France

Murielle Georges Leroy

Université de Franche-Comté, CNRS, Besançon and DRAC Lorraine, Metz, France

Frédérique Bertoncello

CNRS, Université de Nice- Sophia Antipolis, Nice, France

Olivier Weller

CNRS, Université Paris 1 Panthéon-Sorbonne, Paris, France

with the collaboration of Elise Fovet, Cristina Gandini, Jessica Giraud, Laure Nuninger, François Favory

Abstract:

The ArchaeDyn team has investigated territorial dynamics by comparing areas over long time spans between the Neolithic and Modern times. Datasets on various themes have been shared and indicators and analytical models produced. This paper presents both the formalization of scientific process used in the ArchaeDyn programme and a conceptual model of the systems and components so that synchronic and diachronic comparisons can be made. The aim is to clarify the transition from an archaeological feature (a site or an artefact) or a recording unit (survey area) as the input, to the characterization of spaces describing a system as the output. The approach is described by the successive steps corresponding to semantic, spatial or temporal analytical processes from the prospecting level to the level of complex objects such as the territories and spaces under study. These complex objects— consumption areas, agricultural areas and settlement patterns—are the subsystems in the dynamics of human territorial occupation. The main contribution of this formalization is that it describes synthetically the hypotheses tested and the approach implemented. In addition, it provides a collective validation of the research team's reasoning.

Keywords:

Models, Formalization, Analytical Process, Agricultural Areas, Settlement Patterns, Consumption Areas

Introduction

This paper is the outcome of collaboration as part of the ArchaeDyn programme funded by France's ANR (ANR-08-BLAN-0157) (Gandini et al. 2012). The ArchaeDyn programme was launched in 2004 to study territorial dynamics from Neolithic to Modern times. The programme is conducted by three workgroups focusing on separate themes: (i)

Corresponding author: laure.saligny@u-bourgogne.fr

the movement of raw materials and manufactured objects, (ii) agricultural areas; (iii) settlement patterns and territories.

All three workgroups exploit archaeological inventories collated as part of other scientific programmes. The ArchaeDyn members have implemented an analytical approach for sharing these datasets and producing indicators and analytical models with which to compare geographical spaces

over large time spans (Gandini et al. 2012; Poirier and Tolle 2008).

The sheer diversity of analysis produced in the eight years the programme has been running (for the same purpose of describing and comparing different 'pieces of space' occupied by humans) means that, although the analytical methods are properly applied, the process is not always clear at each stage.

This paper begins by presenting the theoretical models underpinning the programme's assumptions, defining the scientific goals in terms of thematic, spatial and temporal issues and hypotheses. The second part of the paper sets out the way in which each workgroup formalizes the scientific process. These process schemes explain both the transformation of study 'sites' into a characterization of 'pieces of space' and the different levels of studies, the various objects composing the systems under study and their interrelations. The purpose of this formalization is to develop a scheme by which to describe different objects and their transformations and to clarify the transition from an archaeological feature (a site or an artefact) or a recording unit (survey area) as an input, to the characterization of spaces as an output. The final part of the paper describes the conceptual models of each system in order to answer the questions: What are the objects composing the spaces under study? How do they evolve individually within these spaces? What is the basic spatial and temporal significant unit of the studies? With these conceptual models, the aim is to understand the rule for each dimension (semantic, time and space) at every step of the analytical process.

1. Theoretical Models and Hypotheses

The ArchaeDyn programme brings together many researchers working from common assumptions in order to compare spaces between Neolithic and Modern times. The programme is formed around three workgroups engaged in various areas of research at different spatial and chronological scales and using different archaeological data. One workgroup investigates the movement of raw materials and manufactured objects over geographical areas that may be as vast as Europe and that range in time from the

Neolithic to the Iron Age. Its data are mostly inventories of movables. A second group works on the micro-scale level on agricultural areas from the end of Protohistory to the Middle Ages based on fieldwalking data. A third group studies settlement patterns and territories from 800 BC to 800 AD on a regional scale by processing data from various sources (field surveys, textual sources, etc.).

The three workgroups have developed their own analytical models based on the specific theoretical models described below.

1.1 Product distribution model

In order to analyse the spatial phenomena related to the movement of Pre- and Protohistorical raw materials and manufactured objects in Western Europe, the workgroup began by identifying the processes involved in their diffusion (Gauthier et al. 2012a). Substantial modelling was carried out to list the elements composing the diffusion systems and their organization, their interactions, the factors influencing the location of products or their transfers. The general theoretical model suggested is intended for the products under study. The objective of this first model is to establish and define the relevant concepts and terminology.

Three possible 'domains' in the life cycle of a product are identified (Fig. 1): production, diffusion and consumption, with the latter including both the use and the hoarding of products. Each archaeological 'place' is associated with one of these three domains.

Unlike the first step of supply and the final step of voluntary or involuntary abandonment (or recycling), which define the beginning and the end of the life cycle, none of the intermediate stages is mandatory. The object may have followed a variable pathway, moving backwards and forwards among the three domains. The transition from one type of site to another is a transfer. A sequence of such transfers may entail the product being moved in space over time, which is termed diffusion.

1.2 Agricultural area model

The workgroup studying agricultural areas aims to evaluate human appropriation and infield

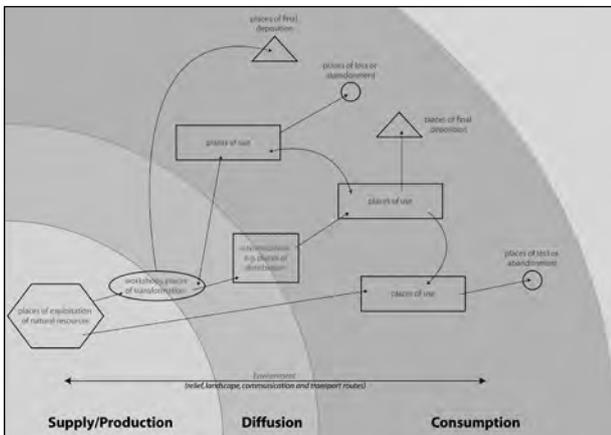


Figure 1. Modelling of transfers between functional places (Gauthier et al. in press, Fig. 6).

dynamics of exploited spaces in order to ensure domestic supply, from Protohistory to Modern times.

The possible transformations of these spaces can be summarized as follows:

- cultivated to uncultivated
- cultivated to forest,
- cultivated to pasture,
- cultivated to cultivated (change of crops),
- and vice versa.

These transformations may be repeated on a number of occasions over the course of time.

This agricultural area is studied on a local scale in many study areas from the identification of ardmarks (such as manurings identified from off-site material during fieldwalking surveys), or ancient field systems, or agrarian structures preserved under forest cover.

The aim is to ascertain the criteria that determined the location of intensely appropriated spaces, with regard to water, soils, relief and distance from settlements. Which settings were intensely and durably occupied? Which ones were ignored?

The idea is also to evaluate the validity of centrality models, like von Thünen's model (Thünen 1826) for agricultural activities (O'Kelly et al. 1996).

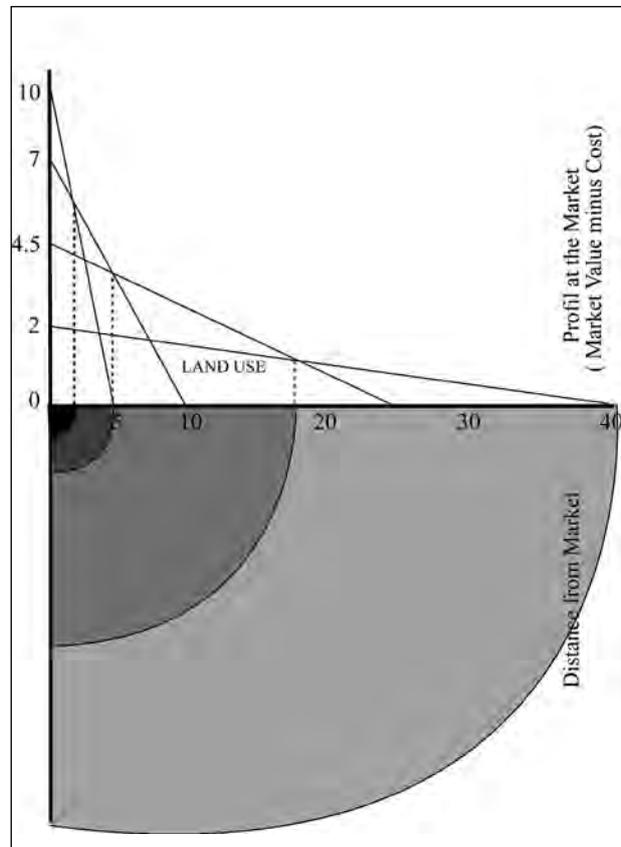


Figure 2. The Von Thünen model (<http://www.csiss.org/classics/uploads/vonthunen2.jpg>).

Is there a linear relationship between the degree of appropriation and the distance from settlements (Fig. 2)?

1.3 Settlement model

The workgroup aims to analyse the intensity, stability and organization of rural occupation in southern and central France from 800 BC to 800 AD using indices related to settlement. The goals are to characterize the regional settlement systems (density, spatial and hierarchical structure and evolution) by using common criteria and to understand the mechanisms or processes causing changes in the systems (Bertoncello *et al.* 2012). It is essential to compare study areas in order to evaluate the scale on which changes occur (regional or supra-regional), which may point to different causes.

The archaeological objects considered for the study of the settlement systems are the remains of dwellings or similar buildings (farm or craft

buildings). Religious sites and burial sites are excluded because they are not to be found in all the areas and they reflect specific spatial strategies and temporal resolutions that differ from those of habitat. Because networks of settlements are under study, the remains taken into account are considered in the broad sense of the term habitat, as defined in geography by Roger Brunet: 'Habitat is the grouping and layout of dwellings in a given space; it can include outbuildings for animals and stores, as well as workshops and other constructions for occupational use. [...] Rural habitat [is] anything built in the countryside' (Brunet et al., 1992: 229).

2. Process Formalization

To study these models and make synchronic and diachronic comparisons, it is necessary to describe and formalize the objects under study and the way they are processed and analysed.

For each schema (Figs. 3, 4 and 5), the components are the entities (or objects) and the processes.

The entities are grouped into six classes (the ovals in the diagrams):

- recording units which are arbitrary: they are used in field surveys to locate and quantify sherds.
- observed objects or primary data which are artefacts, remains, sherds, etc.
- constructed objects construed from the primary data which are interpreted and validated, for example a settlement or an agricultural area.

These observed and constructed objects are unambiguous objects.

- analytical objects or 'tools', devised for analysis and comparison,
- theoretical objects, used as references.

The analytical and theoretical objects are abstract objects, such as a cell in which qualitative and quantitative archaeological information can be aggregated.

- studied objects, the expected results, the 'pieces of space' or the spatial configuration which the ArchaeDyn programme aims to characterize.

Three kinds of processes are distinguished (the diamonds in the diagrams):

- process based on knowledge and on archaeological interpretation: they involve the classical archaeological approach,
- process involving statistical, spatial or geoprocessing methods,
- comparative process, which may include methods of analysis with archaeological descriptions and interpretations.

Each entity and process occurs at a level of observation and study of the phenomena.

The first level concerns the exploration phase (field survey or documentary inventories) with objects containing specific information or primary data. The second level is the functional level for characterizing 'pieces of space'. At this level, the abstract objects are developed so that data can be standardized, compared and analysed over time and space. It is the level of transition at which primary data are selected and converted into interpreted data. Moreover this level involves a transformation in the way space is perceived: it is the transition from isolated objects to a 'pieces of space'. This space is characterized by the presence or absence of certain objects and their relations. The final level is that of the characterization of 'pieces of space' describing the study area. It is at this level that dynamics between different periods or between different regions can be compared.

2.1 Process formalization for product diffusion systems

The team seeks to identify production, transformation or consumption areas and the circulation network connecting these areas, as described by information about flow and type of transfer (Fig. 3).

To this end, the investigators collate several heterogeneous inventories of archaeological

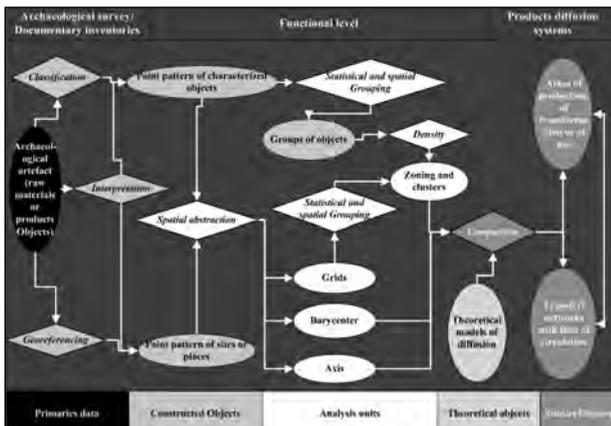


Figure 3. Process formalization for product diffusion systems (L. Saligny, X. Rodier – WG4, Archaeodyn ANR-08-BLAN-0157, 2012).

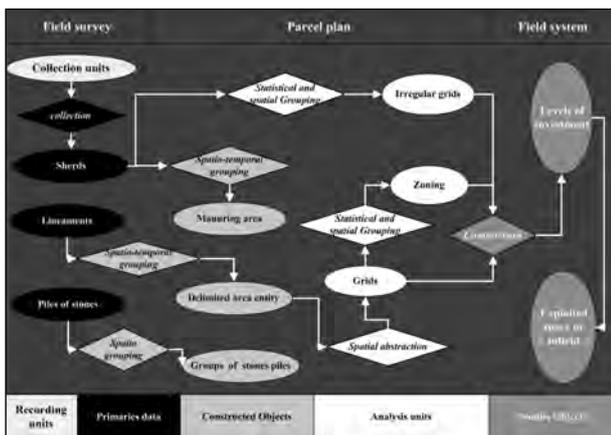


Figure 4. Process formalization for agricultural areas (L. Saligny, X. Rodier – WG4, Archaeodyn ANR-08-BLAN-0157, 2012).

artefacts. Then, a series of scientific processes drawing on archaeological knowledge is used to homogenize these data, such as a descriptive grid to classify objects as characterized objects. As the next stage, these objects are grouped into sets according to criteria of similarity and distance. A second method is the building of a georeferenced database of sites or locations where objects are found. The finds are usually located by reference to contemporary municipal boundaries. The last process at this stage is the interpretation made for sites and for objects at the same time. The presence of an object at a site could help in ascribing that place to a category of production, transformation or hoarding, etc. These constructed entities (sites and characterized objects) are used for building the analysis units such as grids and zoning by spatial abstraction or

statistical and spatial methods. The final process is to use comparative tools and methods and to make interpretations by reading maps or charts.

The end-results are proposals for items related to the description of the system: areas of production, transformation, use or hoarding and the transfer or circulation network as a mode of transfer and in terms of flow. These results are compared with a corpus of models used in archaeology (Renfrew 1975; Needham 1998; Pétrequin et al. 2012), ethnology (Godelier 1996) and geography (Reilly 1931).

2.2 Process formalization for agricultural areas

The approach implemented in this workgroup is formalized in Figure 4. The diagram obeys the same logic as above to depict the transition from one level to another through the transformation of objects and the processes employed for those transformations.

As in the previous case, three levels are identified. The first is the level of inventory of ardmarks recorded in many study areas, such as off-site sherds in field-survey collection units, agrarian structures and fragmented fossilized structures (lineaments or piles of stones) beneath forest cover .

For the second level, these archaeological objects are transformed into a spatial or time-space cluster based on knowledge or on concentration thresholds.

The parcel plan level is the level of description and interpretation of shapes and areas. It includes two categories of objects: archaeological objects, but by this stage they have been interpreted, and abstract objects, developed so the data can be standardized, compared and analysed.

A manuring or a delimited area entity is an interpreted object because it is the archaeologist who chooses (on the strength of empirical or mathematical criteria) whether or not to form a group. The zoning, for example, as an abstract and analytical object, is the outcome of aggregating neighbouring cells with the same attributes or indicators, this group being arrived at by statistical methods. This second level with the analysis units is also a transformation of space: we no longer study

isolated and interconnected objects but a space subdivided into areas.

This level is a transitional stage of abstraction at which these 'pieces of space' can be characterized by means of gradients or degrees of human appropriation (in terms of exploitation). This level is also one that makes regional comparisons possible and enables (or maximizes) the use of all archaeological sources. In case studies of manuring, recording by collection units prevents any studies at the level of parcels. The last level is that of exploited space around the habitats that allows inter-regional comparisons of dynamics (trends will be compared by comparing spectra or curves).

2.3 Process formalization for settlement systems

This approach can be summarized in four steps (Fig. 5):

1. Constructing the site database. This is the level at which archaeological sites are identified by field surveys. These sites are understood as localized and bounded concentrations of archaeological remains that are sufficiently characterized for them to be dated.

2. Constructing the settlement database. At this (functional) level, dating (pottery assemblage) and preliminary interpretation of the primary data from the sites makes it possible to identify settlements. The term 'settlement' refers here to a place where people settled (with no structural or functional connotation) at some time, in a more or less permanent and uninterrupted manner. A settlement may therefore correspond to a site or part of a site: several human occupations (= several settlements) could have occurred in succession at the same point of space (= on the same site). The functional level corresponds to a process of interpretation and decomposition of the primary information, based on knowledge. The site is an intermediate object for building interpreted settlements.

3. Hierarchical classification of settlements. In a systemic approach, it is assumed that the organization and evolution of the system depends on the interactions among its components, i.e. settlements in our case (Durand-Dastès et al. 1998). The nature and intensity of these interactions

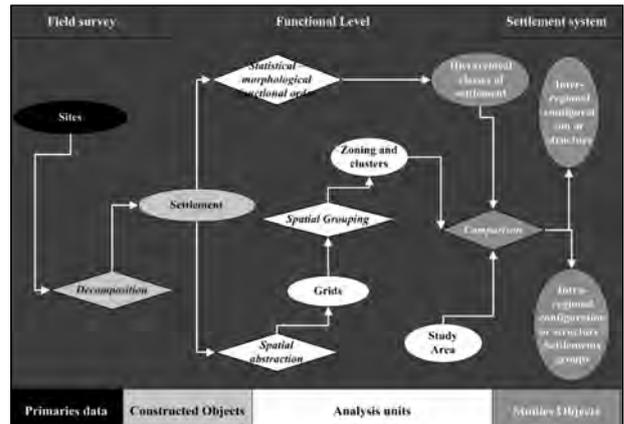


Figure 5. Process formalization for the settlement system. (L. Saligny, X. Rodier – WG4, Archaedyn ANR-08-BLAN-0157, 2012).

depend on the relative rank of the settlements within a functional hierarchy. In order to estimate this hierarchy, settlements are classified by a common set of criteria (size, building materials, length of occupation, function: Gandini et al. 2008) using Factor Analysis (AFC) and Agglomerative Hierarchical Clustering (CAH–Ward's method). The interpretation and validation of the classification combine the analysis of the profiles of the classes and archaeologists' knowledge of the settlements.

4. Spatial abstraction. Spatial tools and models are used to characterize the organization, intensity and stability of settlements in each study area. The study areas are constructed according to the scatter of settlements in each region (Nuninger et al. 2012). They are artificial objects designed to be the analysis units and exploration windows of the settlement systems and to perform inter-regional comparisons.

The functional level of abstraction is the transitional step in characterizing different settlement configurations in space and time.

These processes are intended to provide a standardized approach and common analytical procedures for comparing areas with heterogeneous archaeological databases.

3. The Conceptual Modelling

Once the scientific processes have been formalized in this way, each system is described

by a conceptual model in order to characterize each object in space and its evolution within the space. The purpose is also to explain the rules and importance of each dimension—space, chronology and theme/description/function—for building objects and their descriptors. So, the models are based on a deconstruction into three dimensions: space, time and semantics through which we are to understand the relationships between each dimension and each object in the system (Rodier and Saligny 2010; Rodier et al. 2010). The main aim of this conceptual modelling is to homogenize the archaeological inventories so as to be able to use them together and produce comparable results.

Each model has an upper and a lower part (Figs 6, 9, and 11) with the representation of nested boxes indicating a strict inclusion. At the top, the biggest box describes the system under study. The box contains a series of boxes representing each of the objects composing the system (described below, chap.2). The arrows specify the nature of the relationship between each object: the transfer of attributes, a spatial transformation (summary, account, etc.), and so on.

The lower part is composed of three dimensions used to describe objects: Description/Function, Space and Time. Each of these dimensions uses specific features for building and describing archaeological and analytical objects. Between each part, relationships are drawn representing the combinations of association, inclusion, inheritance or transfer.

3.1 The conceptual model for the product diffusion system

In this model (Fig. 6), the time dimension is composed of a time feature which consists of long phases or chrono-cultural periods like the Early,

Figure 7. Variation of the proportion of Neolithic Alpine axeheads with distance from the workshop at Alba (Italy) (data: Jade, P. Pétrequin (ed), JADE; graph: L. Nuninger, WG3, Archaedyn ANR-08 BLAN-0157, 2007).

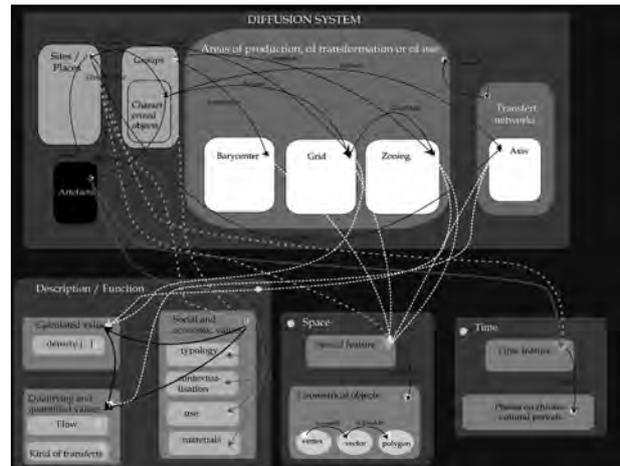
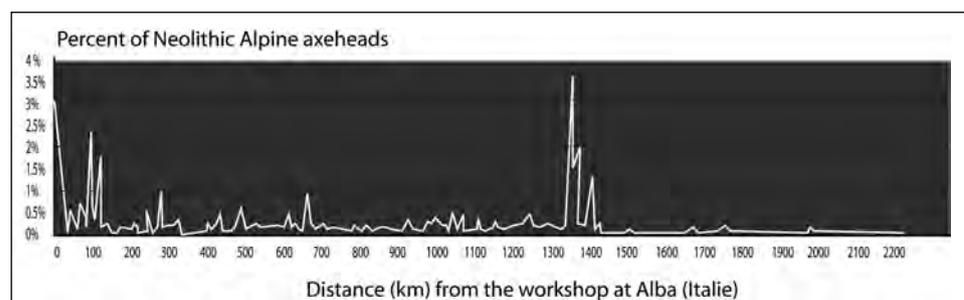


Figure 6. The conceptual model for the product diffusion system (L. Saligny, X. Rodier – WG4, Archaedyn ANR-08-BLAN-0157, 2012).

Middle, Late Neolithic, and so on. Time features are used only for building artefacts and places or sites. Through the interplay of relationships among objects, datings are allocated to the characterized objects, then to the other units of analysis (like grids), and finally to the objects studied (like networks or production areas).

The space dimension is composed of spatial features consisting of standard geometrical entities such as points, lines and polygons. Only sites and abstract units have their own geometry.

Through the interplay of relationships, the areas of production and transformation and the networks exist and are defined only by the existence of grids, zonings and axes.

The semantic dimension is composed of three interconnected boxes. The first, the social and economic values, describes artefacts according to different settings. The combination of these values is used to characterize and describe sites and groups.

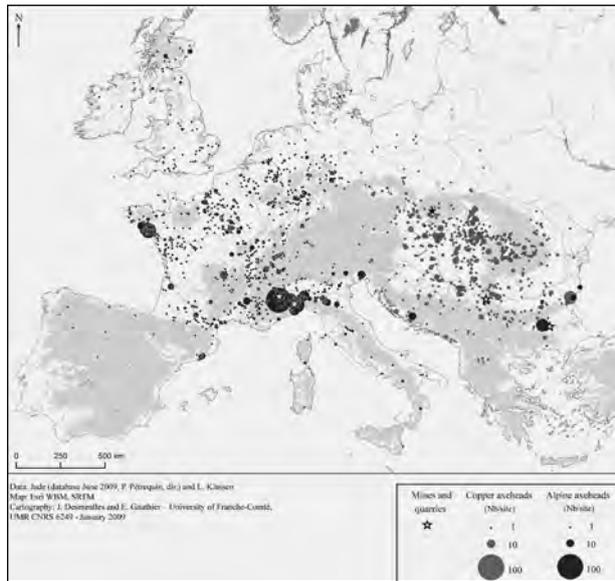


Figure 8. The repartition of alpine jade axeheads and copper axeheads during the 5th millenium in Europe (Pétrequin et al. 2011).

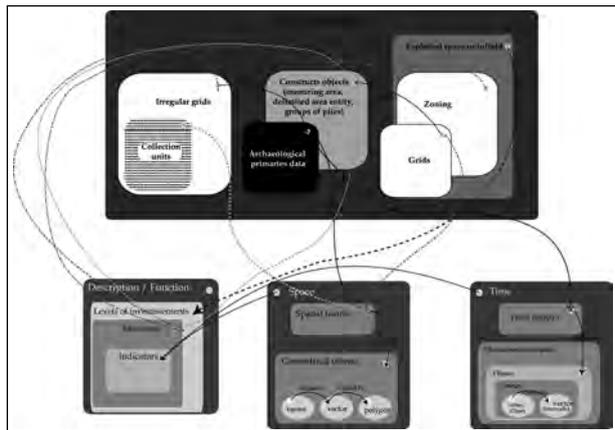


Figure 9. The conceptual model for agricultural areas (L. Saligny, X. Rodier – WG4, Archaeodyn ANR-08-BLAN-0157, 2012).

The other boxes, such as calculated values, qualified and quantified values, depend on selections and on the analysis of social and economic values. These new values are to describe grids, zonings and axes and they are based on the economic and social values of the areas or the sites connected by the axes.

The scientific process and the conceptual modelling of the diffusion system allow the workgroup to achieve interesting results. The following diagram (Fig. 7) and the map (Fig.8) presents the number of Neolithic Alpine axeheads

at different distances from the production area. Ignoring the concentration of the axeheads in the vicinity of the quarry sites, we can observe several peaks corresponding to the centres where the axes were distributed in the direction of north-western Europe. The graph highlights an exceptionally high peak showing the extraordinary concentration in Morbihan (Brittany) mainly on the site of Carnac, depicting the importance of the Carnacean phenomenon in Neolithic Europe. During the fifth millennium BC, Europe was divided between two systems, one with Varna and the other with the Morbihan as the epicentre, in which different socially valued objects were used by elites: copper in the East and jade in the West (Pétrequin et al. 2012).

3.2 The conceptual model for agricultural areas (Fig.9)

The space dimension in this study is not continuous and not fixed at the beginning because the external limits and the internal divisions change depending on whether or not they belong to the agricultural area: the areas that become uncultivated or built are omitted from the studies. So, the analytical units such as grids are not totally included in the exploited space box, because, during the time period under study, space can be transformed into uncultivated land or can be overbuilt.

In the time dimension, time is continuous, with an (almost regular) division into phases. By extension, we can consider that time is eurythmic: the rhythm is predetermined upstream, so it is regular and constrained. The time is a 'media time' or "temps-support" like a common framework for comparing study areas. Only the primary data are concerned by time and spatial features, but conversely, they do not have any descriptive or functional dimension. This third dimension is used with the functional level (see 2.2) for the constructed objects and the analysis units which have some description based on measurements. These measurements were obtained by analysing indicators which are created by analysis of space and possibly dating. In this study, the indicators of functional dimension are constructed from characteristics of spatial and temporal dimensions. In a second step, these indicators taken separately or combined become measurements for describing analytical units.

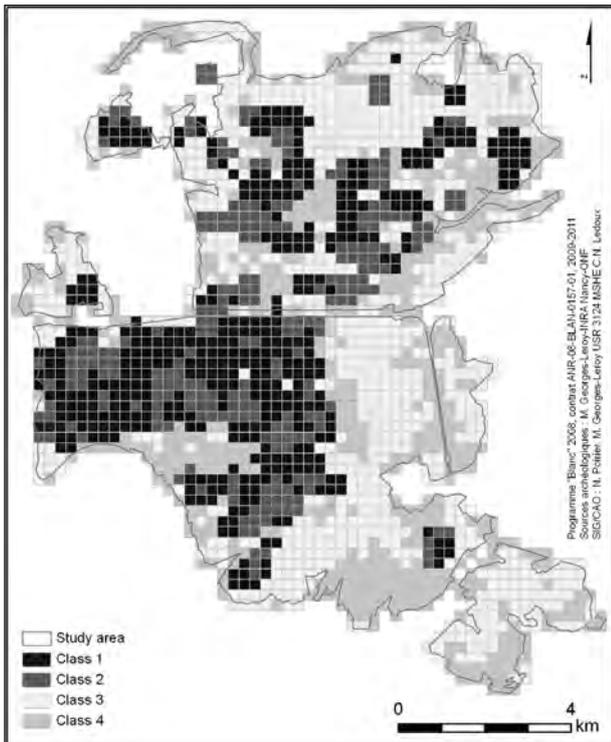


Figure 10. Intensity of human investment (Data: M. Georges-Leroy, INRA Nancy, ONF; GIS: N. Poirier, WG1-ANR ArchaeDyn- 2011).

The exploited space is described by the level of human investment in terms of stability, instability or durability. For example, Figure 10 shows the results with characterization of space based on the ancient field system preserved under the forest cover: the statistical combination of the lineament density, the sum of lineament lengths and the automated classification of plot shapes gives a measure of the intensity of human appropriation. We are thus able to distinguish areas with high densities of regular plots, areas with long lineaments, areas with high densities of disorganized isolated lineaments, or areas of low-density remains. These differences are directly related to the level of appropriation by ancient societies in exploiting these spaces for agricultural purposes, requiring or not requiring the building of a regular field system.

3.3 The conceptual model for the settlement system

In this model (Fig. 11), the settlement, interpreted entity, inherits the spatial dimension of the site. It is also the only object to be dated by the temporal feature.

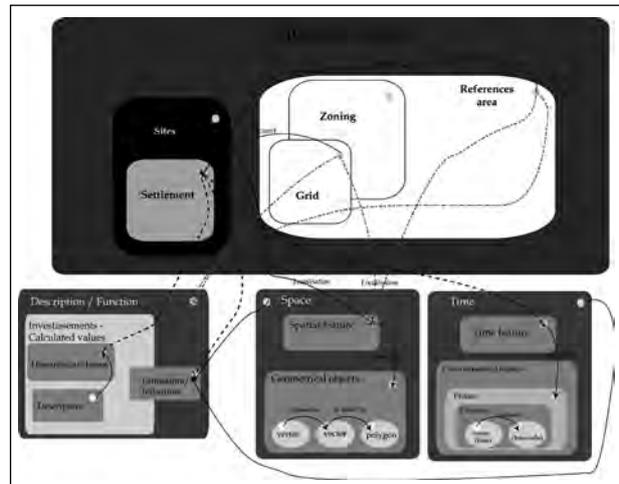


Figure 11. The conceptual model for the settlement system (L. Saligny, X. Rodier – WG4, ArchaeDyn ANR-08-BLAN-0157, 2012).

The grids and the study area are located by spatial features represented by different geometric objects. They are dated indirectly by the settlements they contain (with an inclusion relationship).

In this study, the functional dimension is built from objects of spatial and temporal dimensions, and their relationships, as in the conceptual model of agricultural area (see 3.2). It is the same process as above: the indicators taken alone or in association become measures with which to describe the grids and the study areas.

The main difference is the inclusive box, which is the settlement system in this case, as the aim is to compare different regional settlement configurations in southern and central Gaul.

Therefore the Gaulish settlement system is the global object, studied with various indicators of settlement density (Fig.12), hierarchy or spatial configuration (Bertoncello et al. 2012b).

4. Conclusions

In conclusion, the main contribution of this formalization is the synthetic description of the hypotheses tested and the approach implemented. This 'psychoanalytic work' means the ArchaeDyn team can discover the different obstacles and 'dead ends' and find solutions to them. In addition, it provides for the research team a collective validation

Archaeopress.

Gauthier, E., O. Weller, and J. Giraud, with the collaboration of M. Gabillot, and P. Pétrequin. *In press*. "The circulation of raw materials and manufactured objects. A challenge for spatial analysis." In *Spatial Analysis Applied to Archaeological Sites from Protohistory to the Roman Period*, edited by J. Bourgeois and E. Deweydt. Ghent: Ghent University Press.

Godelier, M. 1996. *L'énigme du don*. Paris: Fayard.

Needham, S. 1998. "Modelling the flow of metal in the Bronze Age." In *L'atelier du Bronziste en Europe du XX^e à s. au VIII^e s. avant notre ère, t.III, Production circulation et consommation du bronze*, edited by C. Mordant, M. Pernot and V. Rychner. Paris: CTHS et CRTGR, Université de Bourgogne.

Nuninger, L., L. Saligny, K. Oštir, N. Poirier, E. Fovet, C. Gandini, E. Gauthier, Ž. Kokalj, and F. Tolle. 2012. "Models and tools for territorial dynamic studies." In *Settlement Patterns, Production and Trades from the Neolithic to the Middle Ages. ArchaeDyn*, edited by C. Gandini, F. Favory and L. Nuninger, 23-37. Oxford: Archaeopress.

O'Kelly, M., D. Bryan. 1966. "Agricultural Location Theory: Von Thunen's Contribution to Economic Geography." *Progress in Human Geography* 20(4): 457 -75.

Pétrequin, P., A.-M. Pétrequin, M. Errera, S. Cassen, E. Gauthier, D. Hovorka, L. Klassen, and A. Sheridan. 2011. "From Mont Viso to Slovakia: the two axeheads of Alpine jade from Golianovo." *Acta Archaeologica Scientiarum Hungaricae* 62: 243-268.

Pétrequin, P., S. Cassen, M. Errera, L. Klassen, and A. Sheridan. 2012. *Jade. Grandes haches alpines du Néolithique européen. Ve et IV^e millénaires av. J.-C.* Ledoux, Besançon: Presses Universitaires de Franche-Comté et Centre de Recherche Archéologique de la Vallée de l'Ain.

Poirier, N., and F. Tolle. 2008. "Measurements of Diachronic Stability of Agrarian Exploitation." In

Layers of Perception. Proceedings of the 35th Computer Applications and Quantitative Methods in Archaeology Conference, Berlin, Germany, April 2-6, 2007, edited by A. Posluschny, L. Karsten and I. Herzog. CD-ROM. Bonn: Habelt.

Poirier, N., M. Georges-Leroy, F. Tolle, and E. Fovet. 2012. "The time-space dynamics of agricultural areas from Antiquity to Modern times." In *Settlement Patterns, Production and Trades from the Neolithic to the Middle Ages. ArchaeDyn*, edited by C. Gandini, F. Favory and L. Nuninger, 39-50. Oxford: Archaeopress.

Reilly, W.J. 1931. *The Law of Retail Gravitation*. New-York: Knickerbocker Press.

Renfrew, C. 1975. "Trade as Action at a Distance: Questions of Integration and Communication." In *Ancient Civilization and Trade*, edited by J.A. Sabloff and C.C. Lamberg-Karlovsky. Albuquerque: University of New-Mexico.

Rodier, X., and L. Saligny. 2010. "Modélisation des objets historiques selon la fonction, l'espace et le temps pour l'étude des dynamiques urbaines dans la longue durée." *Cybergeo: European Journal of Geography, Systèmes, Modélisation, Géostatistiques* 502. Last modified June 2010. <http://cybergeo.revues.org/index23175.html>.

Rodier, X., L. Saligny, B. Lefebvre, J. Pouliot. 2010. "ToToPI (Topographie de Tours Pré-Industriel), a GIS for understanding urban dynamics based on the OH_FET model (Social Use, Space and Time)." In *Making History Interactive. Proceedings of the 37th International Conference in Computer Application and Quantitative Methods in Archaeology, Williamsburg, Virginia, USA, March 22-26, 2009*, edited by B. Fricher, J. Crawford and D. Koler. Oxford: Archaeopress.

Thünen, J.H. 1826. *Der Isolierte Staat in Beziehung auf Landwirtschaft und nationalökonomie, oder Untersuchungen über den Einfluss, den die Getreidepreise, der Reichtum des Bodens und die Abgaben auf den Ackerbau ausüben*. Hamburg: Perthes.

Defining and Advocating Open Data in Archaeology

Stefano Costa

Working Group on Open Data in Archaeology, Open Knowledge Foundation

Anthony Beck

University of Leeds, United Kingdom

Andrew Bevan

University College London, United Kingdom

Jessica Ogden

L – P : Archaeology, United Kingdom

Abstract:

“A piece of content or data is open if anyone is free to use, reuse, and redistribute it - subject only, at most, to the requirement to attribute and share-alike” (<http://opendefinition.org/>). Driven by demands for greater transparency from government, general freedom of information and an increased awareness of the unanticipated re-use values of existing information, Open Data has seen dramatic growth in the past two years. Is archaeology part of this general trend? Our aim is to explore what it means to make archaeological data open and what processes are required to make it happen in a satisfactory way. There are three major goals: (a) individual and institutional advocacy, (b) ethical discussion and consensus-building, and (c) knowledge transfer (licenses guidance, wider academic context, repositories etc.). In this paper, we explore some of these issues in greater detail.

Keywords:

Open Access, Open Data, Open Archaeology, Open Knowledge Foundation, Linked Open Data, Knowledge Transfer

1. Introduction

‘Open’ is increasingly invoked as an attractive thing to be, across a host of human behavioural domains from scientific practice to corporate responsibility to governmental action (to name but a few). The concept readily suggests notions of public accountability, transparency of practice, plurality of opinion and scientific repeatability, thus providing a catchy banner for the advocacy of a range of perceived public goods. In contrast, while being ‘closed’ might occasionally carry positive connotations of increased security, it is less marketable and causes significant problems for a domain whose advances are predicated on a developing and accessible corpus. Some aspects of the open agenda are now several decades old and have had a tangible impact on many different sectors. For example, two well-established international initiatives are Open Access

(e.g. where publications such as journal articles are made publicly available without a fee) and Free/Libre and Open Source Software (FOSS, where source code is made available under liberal licenses that encourage onward use and modification), both arguably expressions of a wider change in the free circulation of information and knowledge. A small minority of archaeologists have also been embracing, discussing and/or contributing to such initiatives for some time, particularly those with an interest in customising digital tools or encouraging broader archaeological participation and dialogue beyond a few well-resourced Western institutions (e.g. in line with the participatory agenda espoused by World Archaeology). Needless to say, the Web has been a catalyst for these changes.

The latest chapter is Open Data, where priority is placed on full publication of vast tracts of undigested information that previously might only be referred to publicly in summary form, if

Corresponding author: stefano.costa@okfn.org

at all. As with Open Access and FOSS, Open Data is not an initiative that is specific to archaeology, but instead reflects much wider calls, for example, for increased access to government data (data.gov.uk, 2012) and data across academia (Royal Society, 2012). The underlying rationale of Open Data is that promoting untrammled access to large amounts of 'raw' information enables patterns of re-use and knowledge creation that were previously impossible and/or largely unanticipated. We seek to define what we mean by archaeological data in clearer terms below, but regardless of how this is done, most practitioners would certainly agree that archaeology produces a great deal of data and that a rapidly increasing portion of this is digital. Indeed, being digital is a necessary condition behind the free circulation of data over the Internet, and it is unsurprising that Open Data is particularly popular in the IT domain. In recent years, this popularity has generated an interest in Open Data in archaeology which is highlighted by the increasing number of sessions at the Computer Applications in Archaeology meetings dedicated to this topic. While there is undoubtedly an element of hype, this growing enthusiasm remains a very positive trend in our view. In this paper, we therefore explore how we might define the nature of archaeological data, and thereafter use this as a backdrop for discussing the strengths and weaknesses offered by existing models for data sharing. In particular, we are interested in advocating data-sharing models that emphasise sustainability, remain aware of different cultures of practice and that engage with broader social and ethical issues in archaeology.

2. What is Archaeological Data?

A traditional, often implicit view of archaeological 'data' places it in opposition to the act of 'interpretation'. From this perspective, data are (often standardised) information packets that are meant to capture the archaeological record or other kinds of archaeological evidence in raw form and which are typically obtained via a rigorous, scientific method (e.g. through controlled excavation or through instrumental analyses), whereas interpretation is seen as a subsequent, richer stage of the archaeological process. Wider debate in archaeology as a discipline (e.g. between so-called processual and post-processual perspectives) and value judgements about more and less important

kinds of archaeology (e.g. digging versus synthesis and theorising) have often found purchase around this assumed contrast. In recent years, however, this contrast has rightly been subject to revision, as various commentators have noted the widely acknowledged, if often forgotten, fact that all stages of archaeological practice involve theory-laden assumptions, and hence that data collection and interpretation are closely entwined. Current perspectives on this kind of issue tend to emphasise pragmatism in making such distinctions (e.g. Lucas 2012), and in line with this, here we retain an idea of data as being packets of information that are particularly amenable to easy reorganisation and/or automated processing. Even so, the kinds of information that can be thought of as data are much more diverse than some might initially think. Through interpretation and knowledge acquisition, data can be transformed. Sometimes the data transformation process consists of a series of steps, or a workflow. Each step can contain assumptions and generalisations that have an impact on the final interpretation. It is rare for this transformational process to be documented. Open archaeological data therefore should refer to all archaeological information that is shared (by whatever means) under licenses that allow re-use by everyone (e.g. according to the Open Knowledge Definition²). This definition is deliberately not a neutral description of the status quo in archaeology, but rather aims to provoke a wider shift in perspective about what constitutes archaeological data in the first place. We should also expand our sense of who the stakeholder might be for open data, avoiding the idea of an experts-only environment to one in which the re-users might be archaeologists, policy makers, planners and the public. Put simply, the data requires democratising. We hope the knock-on effect of such an effort will also raise digital awareness and literacy. This will affect the creation, management and use of data at all levels.

A broader definition of data enables a richer environment of reuse to develop. For example, through the removal of paywalls Open Access results in increased public availability of journals and publications in digital format. However, a fully open license also allows us to add value to Open Access texts by using powerful text-mining software

² Open Knowledge Definition <http://opendefinition.org/>

in order to perform automated categorisation and extraction of information. This is already popular in biomedical science (see Open Knowledge Working Group on Open Science 2012). By means of text-mining, information that was previously thought to be locked within the semantic complexity of natural language can be reborn and extracted as structured data. In line with this, the concept of a 'derived dataset' is key to the understanding of critical aspects of usage licenses and the determination of rights ownership, discussed below, but also for a long-term vision of how archaeological knowledge might circulate if made free to do so. Certain kinds of open archaeological data have the potential to be 'big data', not least because they can claim to have wide spatio-temporal coverage. However, it would be wrong to assume that only big data deserve to be open. As with anything else, scale and size should not be the sole parameter to assess quality. Models of dissemination may differ significantly depending on the size of the dataset, but small-scale datasets are both important in their own right, and prone to becoming large-scale via judicious aggregation with others at a later date. These initiatives allow us to rediscover a pathway to the kinds of synthetic perspectives on long term human culture last possible many decades ago (Bevan 2010).

3. Models of (Open) Data Sharing

It is clear that the sharing of archaeological data, especially via the Web and via licenses that encourage re-use, is nothing new or pioneering. Well-known and established services such as the UK Archaeology Data Service³, Open Context⁴ and the Digital Archaeological Record⁵ host a range of archaeological datasets. Private bodies such as Oxford Archaeology make their grey literature available as an open archive⁶. Research projects share through their websites parts or all of the data archives created during the project. Occasionally, researchers integrate their own published papers with digital data otherwise unavailable through traditional means. Some research teams have their

³ Archaeology Data Service <http://archaeologydataservice.ac.uk/>.

⁴ Open Context <http://opencontext.org/>.

⁵ the Digital Archaeological Record <http://www.tdar.org>.

⁶ Oxford Archaeology Library <http://library.thehumanjourney.net/>.

archives on the Web, but under restricted access. The variety of actors engaged in data sharing is largely mirrored by the variety of technical choices and tools that are adopted to fulfill the task. The spectrum ranges from interactive web applications (through WebGIS or query-able databases) to catalogues conforming to technical standards (such as the European Commission INSPIRE Directive⁷), and from complex datasets made interoperable via simplified formats and structures) to simple datasets available as web pages or documents. It is still rare to publish datasets alongside finished articles and rarer still to have any means to reproduce analysis and processing chains as part of the standard review process.

Given this wider context, we can make a distinction between two alternative models for sharing data: (a) 'living' datasets that either involve interactive applications or continue to experience incremental updates, and (b) static datasets released as a stable resource (i.e. similar to a paper that does not undergo any changes once published). These two models are not mutually exclusive, and technically it is possible to create snapshots from active datasets, or to develop interactive applications for the exploration of inactive datasets.

Some of these web applications (the first model) are easy to install and quite popular because of their ability to perform queries and explore map data via an intuitive user interface. However, this dynamic model is not without issues. First, the scalability of these systems is comparatively low, both in terms of long-term management and usability. Second, their duplicability is low as well, making long-term preservation problematic at best: such server driven systems tend to become one-offs. Moreover, in many cases, actual datasets are locked up in the web service and cannot be easily downloaded. The risk is that when the web service is eventually deprecated the data can be lost. Third, no matter how flexible the design of the interface and querying system, there will always be a limit to what users can do (i.e. the range of possible avenues for re-use is heavily constrained). Finally, web services in general are more difficult to develop and maintain as fully open systems (e.g. as discussed in detail in

⁷ European Commission INSPIRE Directive <http://inspire.jrc.ec.europa.eu/>.

documents like the Franklin Street Statement⁸ and the Open Software Service Definition⁹).

The second model, involving static datasets, is simpler from a technical point of view, and allows for the sharing of files and file archives by users who download them from, for example, a stable URL. The absence of built-in exploratory tools makes documentation particularly important in this case, even though catalogue software like CKAN are developing such tools. Documentation need not be limited to a formal description of the content, but should also be explicit about the who, when, where, what, why and how the dataset was created in first place. Data can be published in various formats, with some interoperable formats which have demonstrated widespread acceptance (.csv, .xml, .shp). However, the use of some of these interoperable formats can minimise the complexity of the underlying source data leading to the publication of incomplete datasets. In terms of data 'objects' this model is far more scalable, as from a technical point of view archiving is straightforward, and the cost of hosting even thousands of these datasets is linear to the quantity, not the quality or complexity. However, data aggregation for analysis beyond that envisaged by the original scale of data collection is complicated due to structural, syntactic and semantic heterogeneities (Bishr 1998).

In reality, the main data repositories cited above allow for some provision of both living and static data. Nevertheless, it is worth highlighting the strong analogies between the static model for data dissemination and traditional publication (either digital or printed). Hence, static approaches to data sharing not only are attractive for the technical reasons noted above, but also because they encourage a smoother transition from older (static, printed matter with restrictive licensing) to newer (static open data) knowledge sharing systems. Moreover, this approach can also work well within the current system of academic evaluation and reward, by making the dataset a typical type of publication such as a formal data paper. There are already some interesting examples in this direction,

such as the *Journal of Open Archaeology Data*¹⁰ which offers a concept of metadata that is not merely technical but also includes an explanation of the value of the dataset and the possible use-cases envisaged by the authors. The dataset as a finished and fixed item also represents a point of confidence for those who cite or use it, as happens with public domain radiocarbon calibration data.

A third emerging model is Linked Open Data (LOD), discussed only briefly here. A fundamental aspect of LOD is the use of Uniform Resource Identifiers (URIs), unique references that can be associated with both human- and machine-readable resources. LOD datasets are commonly distributed as downloadable, highly interoperable files (usually plain text). In one sense, LOD is quite similar to the static data model outlined above, but the use of URIs tends to encourage their treatment as a physical node in a wider network accessed via accompanying web service. Advanced querying of these kinds of LOD frameworks is possible via the SPARQL protocol, in a way which ushers in a highly interactive data environment, albeit one that at present can only be provided by well-resourced institutions. LOD requires datasets to be harmonised to some degree but this need not be very onerous for dataset creators. For example, tools such as Google Refine, are a helpful way to clean up data with an eye to LOD resources (such as those provided by Freebase). Even so, we would stress that these more involved tasks, associated with the construction of richer semantic web content, can and should be kept separate from the easier one of publishing open data in simple open formats. If we continue to place an emphasis on simple open data, we leave fewer excuses for not participating and yet also leave the door open if someone else is interested in enhancing the data via LOD at a later date.

4. Ethics

It is also worth discussing the ethical issues surrounding open data here, with regard both to the reasons why people should feel it incumbent on them to produce such data as part of their professional archaeological lives, and with regard to the potential misuse of open data in the archaeological domain. The socio-economic argument is particularly

8 "Franklin Street Statement", last modified July 14, 2008, <http://autonomo.us/2008/07/14/franklin-street-statement/>.

9 "Open Software Service Definition", last modified October 8, 2008, <http://opendefinition.org/software-service/>.

10 Journal of Open Archaeology Data <http://openarchaeologydata.metajnl.com/>.

compelling. The majority of archaeology is carried out in the interest of the public, often with public money and it is therefore only right that the public get access to the objects and outputs of activities that they have funded. This is particularly clear-cut for academics working at publicly funded institutions and/or on publicly funded projects. The issue initially appears more complicated in the case of developer-funded archaeology, where the funding comes potentially from private commercial sources and the client is notionally the developer. However, in the UK, for example, The National Planning Policy Framework (NPPF: Department for Communities and Local Government 2012) which describes the planning policies on the conservation of the historic environment in England and Wales stipulates that local planning authorities should 'make this evidence (and any archive generated) publicly accessible'. Hence, open data should be construed in this case as a crucial part of the mitigation-by-record process, and not as an afterthought.

A second argument in favour of open data touches on a problem of particular relevance in archaeology: unpublished research. There are various reasons for non-publication, but the sheer cost of bringing archaeological research (especially fieldwork) to formal publication is often a key issue. In such cases, we can nevertheless expect that some data exists. While some might argue that publishing data without offering a synthetic overview alongside it makes for a very limited resource, it is undoubtedly better to have access to data than to have nothing at all. For authors, there is at least some recognition of the work done, as is already the case with archaeological reports filed in grey literature archives. The opening up of grey data is, in our view, an economic way to meet the minimal obligations to publish research. This obligation has been proposed by the European Association of Archaeologists (EAA 2009) and by some national organisations (e.g. Research Councils UK; Associazione Nazionale Archeologi 2010), albeit sometimes together with a 5-10 year embargo period. Including raw data, publications as part of these regulations would represent a major step not only for open data *per se*, but for the wider aim of professional organisations. In contrast, the current view is that publication equals a short report, much smaller than the total amount of information resulting from archaeological activity. We need to transform our understanding of what constitutes

full and satisfactory publication. We would argue that greater favour be given to data rather than synthesis, and more generally to decouple these as two different kinds of publication.

Whilst retaining this sense of the importance and urgency of open data initiatives, it is still also worth noting some ethical arguments that might call for a more cautious approach to open data. For example, the open publication of archaeological data may conceivably lead to greater levels of archaeological looting, especially where this involves provision of exact spatial locations of sites. However, this argument is still largely theoretical and in dire need of some formal demonstration (that this does indeed lead to greater looting activity). In certain particularly vulnerable instances, it may be sensible to place restrictions over who has access to the finest-scale coordinates for archaeological finds. However, this has the unfortunate effect of making data available at multiple scales of granularity based on trust frameworks that are at present rather immature (e.g. probably just based on institutional affiliation), and there are further licensing implications that make it hard to consider this as open data in a strict sense. Furthermore, placing an embargo on data has significant and unquantified downstream implications for a whole range of legitimate activities that would make a positive contribution to knowledge.

A third ethical issue is the degree to which opening up data may conceivably perpetuate existing imbalances and injustices amongst archaeological institutions and practitioners. For example, at the institutional scale, the promotion of open data as a research output in its own right risks reaffirming the advantage enjoyed by bigger institutions who have the wherewithal to absorb the costs of making data open. At the individual scale, it may conceivably risk making it easier for more senior collaborators, company heads, etc. to swallow up some of the rights of individual data creators (especially those early in their careers) by taking the main credit for the overall products (e.g. with respect to the individual data outputs of excavators versus a project director), but ultimately we suspect the reverse danger that already exists is more worrying: currently, data access seems to scale with seniority (via who you know and how important they are; see Cella and Palombini 2012). However, as open

data increasingly becomes embedded in the data production process, it is likely that suitable metrics will be generated so that individual contributions to data impact and quality can be generated and retained.

5. Good Practice

The previous three sections have dealt with Open Data models and surrounding issues that are still a matter of significant debate, both within and beyond archaeology. There is a growing platform of good practices for (open) archaeological data, some already covered in existing documents (e.g. the Archaeology Data Service and Digital Antiquity's *Guides to Good Practice*¹¹), and most of these have garnered a sufficient level of acceptance to be considered as common culture. Issues surrounding technical interoperability are perhaps the best known and the key point to stress is a need for the use of open formats that are independent of any specific software platform. If we exclude the most common and simple problem cases (e.g. documents made with office suite software), the critical point is that currently there remain far too many *de facto* proprietary standards (e.g. .dwg) and far too many undocumented formats produced by survey, measurement and analysis tools. In contrast to technical format issues, metadata (structured background data describing a given dataset) is a method of documentation that is not very well known outside of expert domains. This is partly because of the lack of domain-specific standards, and partly because metadata are seen as mainly something to be done for large catalogues, while a single dataset is often not deemed worthy of this kind of treatment. Furthermore, some types of metadata are easier to add or more commonly produced (e.g. author and keywords of a document, bounding box of a spatial dataset), while the mere existence of a metadata structure does not necessarily mean it will be used (e.g. photographic images where this functionality is present but rarely used). Clearly, the specificity of each dataset asks for a detailed description of the methods and conditions with which data has been recorded and structured regardless of the nature of data. So, for example, while a photographic database may have all EXIF metadata about exposure time, even spatial location, it may or may

¹¹ "Guides to Good Practice," accessed September 23, 2012, <http://guides.archaeologydataservice.ac.uk/>.

not have information about the aim of the photographer, the criteria for choosing what to photograph, in what detail and so on.

The legal aspects of open data are arguably more complex than the technical ones, and also something that is presently under-addressed in the teaching of archaeology. Indeed, while choosing an open license for data is actually very simple, many people continue to prefer a restrictive closed license (perhaps because they are uncertain about the nature of potential re-use) or do not see the point of defining an explicit license (e.g. they have not realised that an open license is likely to be very beneficial for the circulation of their work). Finally, further problems can arise if open data need to be combined with proprietary data to create derived datasets. Some common cases of this issue occur with satellite imagery, but the same can happen with a photographic catalogue, or a dendrochronological calibration curve. The derived datasets will be necessarily restricted by the proprietary license, and it will not be possible to share them without breaking the license terms.

Bearing these issues of adoption in mind, there is nonetheless now a solid platform provided by two organisations that have been dealing with the problem of legal aspects of data for some time now, Creative Commons (CC) and Open Data Commons (ODC). There are not many licenses for open data, especially if compared to software licenses. We can make a distinction among three different types of open license: share-alike, attribution and public domain licenses. Share-alike licenses follow the idea of copyleft and of "cascaded" sharing of derived works under the same license (for further discussion, see Murray-Rust 2008; Cheliotis 2009; Stodden 2009). Wikipedia and all Wikimedia projects use a share-alike license, CC-BY-SA 3.0, that is appropriate for text and images. The Open Database License (ODbL) is better suited for factual data¹². Attribution licenses only require that copies and derived works maintain an indication of the original authors. In this case, too, the CC-BY 3.0 is best for creative content and ODC-BY for factual data. Public domain licenses are waivers of any right, including the right to be cited as

¹² "Open Database License FAQ," accessed September 23, 2012, <http://opendatacommons.org/faq/licenses/#why-not-use-a-creative-commons-or-freeopen-source-software-license-for-databases>

authors, and are appropriate for instrumental data without a clear creative dimension (e.g. chemical analysis), public domain data (bibliographic data), or big datasets that need to be aggregated in several iterations, making attribution tracking quite difficult to achieve. The two available licenses, CCO and the Public Domain Dedication License, are essentially identical, but the formulation of CCO is more universally interoperable with normative systems, having a fallback declaration in those cases when it is not allowed to waive all rights upon one's own work. Note also that three different licenses in the Creative Commons family are mentioned above, each one with its specific traits. Hence, there is no such thing as a "Creative Commons license" or "Creative Commons 3.0" and users should be guided in how they refer to their preferred license by the service available on the Creative Commons website.

Archaeological data collection is governed by a number of statutory, legislative and professional frameworks. These represent the policy environments under which contractual and academic archaeological works are enacted. This policy has a direct impact on how, and on what terms, archaeological data should be archived and disseminated. Unsurprisingly, many of these frameworks advocate the deposition of data and other resources in publically accessible repositories. For example, in the UK the Department for Communities and Local Government (2012, 32) state that evidence from archaeological works conducted as part of the planning process (and any archive generated) should be publicly accessible. Equally for the academic sector, Research Councils UK state that 'Publicly funded research data are a public good, produced in the public interest, which should be made openly available with as few restrictions as possible in a timely and responsible manner that does not harm intellectual property'. Unfortunately, most policy does not make a distinction between digital and analogue resources and, as digital archaeology is relative immature, many of the established repositories are designed for physical archives (plans, contexts, artefacts, ecofacts and synthetic reports). However, the majority of archaeological data in the future will be collected, analysed, interpreted and published *in-silico* (digitally). It is important that the policy environments, which in principle support publicly accessible archive deposition for analogue archives,

should do the same for digital archives and emphasise mechanisms that encourage re-use.

Finally, it is worth highlighting how important it is to develop a broader understanding and stronger awareness of copyright among all kinds of archaeological practitioners. We envisage the development and increasing importance of lightweight forms of copyright tracking and recording (e.g. by means of metadata and automatic tracking and versioning systems). Currently, authors' rights with regard to archaeological data are more often than not an obstacle to the circulation of knowledge rather than an efficient mechanism to protect one's work. As such, these rights need to be standardised and constrained in terms of their downstream effects, not only on active and future work, but also for existing (currently orphaned) works and grey literature.

6. Conclusions

For some archaeologists, Open Data requires a dramatic change of perspective, the acquisition of new skills and a major shift from curation of software to the curation of data. For others, Open Data may be as simple as doing what they are already doing. In all cases, technical development alone is not enough. Technologies must be accompanied by a proper social framework. Open Data has the potential to be trans-formative which means that it could significantly disrupt the sector. Education about the potential benefits of Open Data and the social implications of any transformation are required to take full advantage of the new possibilities of open archaeological data. Therefore, the link between funding, publication and Open Data is a key area, and we reaffirm the need to encourage national funding agencies, and international ones (private or public) to build in Open Data policies into the requirements of their grants and to check for such a track record in subsequent grants. Open Data needs to be a more relevant part of the archaeological publication, research, management, curation and policy process, and not merely an afterthought.

References

Archaeology Data Service and Digital Antiquity. 2012. "Guides to Good Practice." Accessed September 23. <http://guides.archaeologydataservice.ac.uk/>.

- Associazione Nazionale Archeologi. 2010. "Codice Deontologico." Last modified March 10. <http://www.archeologi.org/public/ANA.CODICE%20DEONTOLOGICO.MOD.pdf>.
- Autonomo.us. 2008. "Franklin Street Statement." Last modified July 14. <http://autonomo.us/2008/07/14/franklin-street-statement/>.
- Bevan, A. 2012. "Spatial methods for analysing large-scale artefact inventories." *Antiquity* 86 (332): 492-506.
- Bishr, Y. 1998. "Overcoming the semantic and other barriers to GIS interoperability." *International Journal of Geographical Information Science* 12, 299-314.
- Cella, E., and A. Palombini. 2012 "The future of the profession of archaeology: the protection of intellectual property rights and free access to data." In *Opening the Past. Archaeological Open Data. MapPapers* 3-II: 78-80.
- Cheliotis, G. 2009. "From open source to open content: Organization, licensing and decision processes in open cultural production." *Decision Support Systems* 47: 229-244.
- data.gov.uk. 2012. Open Data White Paper: Unleashing the potential. Accessed 3 February 2013. <http://data.gov.uk/library/open-data-white-pape>.
- Department for Communities and Local Government. 2012. "National Planning Policy Framework." Last modified March 27. <http://www.communities.gov.uk/publications/planningandbuilding/nppf>.
- European Association of Archaeologists. 2009. "The EAA Code of Practice." Last modified September 19. http://www.e-a-a.org/EAA_Code_of_Practice.pdf.
- Lucas, G. 2012. *Understanding the Archaeological Record*. Cambridge: Cambridge University Press.
- Murray-Rust, P. 2008. "Open data in science." *Serials Review* 34 (1): 52-64.
- Open Knowledge Foundation. 2008. "Open Software Service Definition." Last modified October 8. <http://opendefinition.org/software-service/>.
- Open Knowledge Foundation Working Group on Open Science. 2012. "Working Group Response to IPO Consultation on Text Mining Copyright Exception." Last modified March 21. <http://science.okfn.org/2012/03/21/response-to-ipo-consultation-on-text-mining-copyright-exception/>.
- Royal Society. 2012. *Science as an open enterprise*. London: The Royal Society.
- Stodden, V. 2009. "Enabling reproducible research: licensing for scientific innovation." *International Journal of Communications Law and Policy* 13: 1-25.

The Evolution of Territorial Occupation: Exploratory Spatial Data Analysis. Uncertainty and Heterogeneity of Data in Archaeology

Lucile Pillot and Laure Saligny

Université de Bourgogne, France

with the collaboration of Scott Madry, University of North Carolina at Chapel Hill, USA

Abstract:

The analysis of territorial changes over time raises many methodological questions. Despite a very frequent use of spatial analysis methods, few works underline their methodological shortcomings. This paper considers a comparison and discussion of exploratory statistical methods (notably Ripley's K function and its derivatives). It is worth noting three parameters in archaeological spatial analysis: scale, quality and quantity of input data. Archaeological data are intrinsically heterogeneous. Therefore, our approach proposes a multiscale process integrating heterogeneous data. We describe a range of methods for the description and analysis of heterogeneous spatial point patterns in archaeology. These methods are described through the spatial analysis of neolithic sites in the South-East of the Paris Basin (France). We conclude by discussing the problems of the logical link between spatial pattern recognition and the archaeological process.

Keywords:

Settlement Changes, ESDA, Point Pattern, Ripley's K Function

Introduction

The study of the evolution of human site-occupation is a common issue of current research programs in archaeology. Many studies seek to understand the settlement pattern over time using many inventories of occupied places (Poirier 2010 ; Gauthier et al. 2012 ; Nuninger et al. 2012). The main purposes of these studies are to observe spatial organizations, to identify any possible spatial patterns or particularities for each period of time and to put forth new theories. Archaeologists have long used methods for analyzing point patterns (Orton 2005) and produced indexes and maps of change between two chronological periods in order to compare them. The results contribute to raise the following questions: Is the space under scrutiny non-randomly organized? What are the visible spatial shapes? How do they evolve through time?

Corresponding author: lucile.pillot@u-bourgogne.fr

Abbreviations: CSR = Complete Spatial Randomness; ESDA = Exploratory Spatial Data Analysis; EN = Early Neolithic; KDE = Kernel Density Estimation; NDF = Neighborhood Density Function; MN = Middle Neolithic; PCF = Pair Correlation Function; PVC = Percent Volume Contour; SADIE = Spatial Analysis by Distance Indices.

With the development of the use of Geographic Information Systems (GIS) in archaeology, the use of analytical methods and spatial statistics has become both easier and automated. The studies that rely on these methods are based on measurements of distances between points. These became commonplace in archaeology. The method most used is the analysis of the nearest neighbor analysis (Clark and Evans 1954, Hodder and Orton 1976, Fletcher 2008). Other methods are applied and are becoming more widespread, such as for example *Quadrats* (Niknami, Amirkhiz 2008, Morgan 2009), Ripley's K function (Bevan and Conolly 2006, Winter Livneh et al. 2010), the K-mean method (Kintigh and Ammerman 1982), the recognition of *Hot-Spots* (Schwartz 2008, Pillot et al. accepted) or the density calculation with the method of *Kernels* (Baxter and Beardah 1997). All these methods of spatial and statistical analysis are borrowed from geography, epidemiology and ecology, and were developed for uses other than archaeology. These methods can involve certain conditions and implicit assumptions, such as the homogeneity of the studied phenomenon, or the value of the sample estimate for a complete population. However, these assumptions often remain underestimated or completely ignored (Crema et al. 2010). These studies often encounter

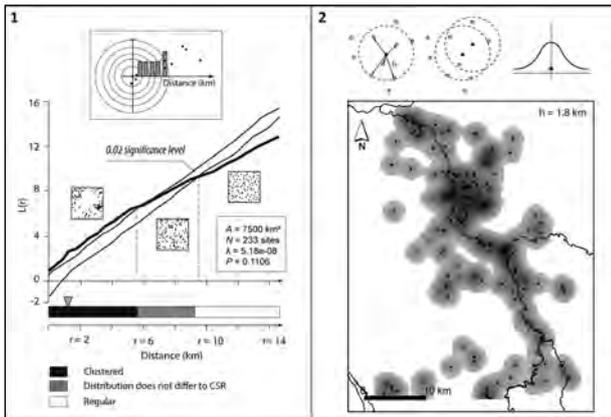


Figure 1. Methods: 1: Ripley's K function; 2: Kernel Density Estimation.

difficulties in using and comparing large sets of archaeological data exhibiting uncertainty or heterogeneity.

This paper seeks to discuss one commonly-used method (Ripley's K function and its derivatives) and to compare different results. It allows us to suggest a new approach of the use of these methods and to suggest how to render them better-suited to archaeological data.

1. Exploratory Spatial Data Analysis (ESDA) for Archaeological Point Pattern

1.1 Exploratory Spatial Data Analysis (ESDA)

In archaeology, we use a range of methods to describe and analyze patterns in sets of *events* (points). These methods pertain to the Exploratory Spatial Data Analysis methods (ESDA). These can be used to qualify, quantify and visualize territorial transformations and to test the hypothesis that human occupation patterns have changed over time. ESDA was thus carried out to identify any possible spatial particularities, including local *clusters* or *hot spots* (Diggle 1983, Cressie 1993, 1994, Anselin 1999, Banos 2001, Zaninetti 2005). These methods rely on the concept of "point process", observing the tendency of the observed points to form spatial clusters³. These analyze point patterns⁴.

3 A spatial point process is a "stochastic mechanism which generates a countable set of events x_i in the plane" (Diggle 2003, 43).

4 A point pattern is a realization of a process (Gatrell et al. 1996), and is simply a collection of points ($p_1, p_2, p_3, \dots, p_n$)

Currently, these statistical methods compare the distributions of point patterns to a reference distribution, according to the null hypothesis of a homogeneous point process (CSR), consisting of discrete random data in two dimensions whose number and location are independent of space, and whose intensity is constant (Ripley 1977, 1981, Cressie 1993, Zaninetti 2005). If the null hypothesis is verified, the distribution is random, whereas if it is not verified; the spatial structure is either aggregated or regular. In this paper, all the methods and approaches mentioned are used in *univariate* analysis, without any distinction between different types of sites within the set of *events*⁵.

1.2 ESDA methods

Extending the use of GIS in archaeology, there is more work mobilizing these methods based on measurements of distances between *events* (Orton 2005). These methods fall into two categories: local and global analyses.

1.2.1 Global methods

The global analyses are used to verify a possible general organization of space. They produce indexes or values in order to demonstrate that *clusters* exist at specific scales. In this category, the method generally used in this paper is Ripley's K function $-K(r)-$ and its Besag $L -L(r)-$ transformation. These functions consider the complete distribution of any distance in the point pattern (Ripley 1981, Cressie 1993, Zaninetti 2005). The process results are expressed as a correlogram (Fig. 1.1). Tests of significance for spatial measures are usually constructed using Monte Carlo procedures (Besag and Diggle 1977, Diggle 2003). A Monte Carlo simulation of the spatial process gives an estimate of the mean and the sampling distribution of the statistic test (Perry et al. 2006). Extreme limits are typically estimated as simulation envelopes, based on a null hypothesis of CSR, using the same intensity as the observed pattern. In the correlogram, it is expressed as a confidence interval. At a given scale, *events* sets may exhibit departure from CSR as either clustering, or regularity. Also, distributed in some region R .

5 Some of these methods are used in bivariate analysis, as KDE, with distinctions between different types of sites within the set of events (weight between points –Goerlich Gisbert 2003).

CSR acts as the ‘dividing line’ between the two (Diggle 2003): if the curve is above the confidence interval, the process is aggregated; and there are some clusters in point pattern; if it is within the confidence interval, the process is random; if it is below the confidence interval, the process is regular (Fig. 1.1). Ripley’s K function is a “multiscalar” tool, used for studying data organization, defining spatial patterns in a homogeneous space and comparing different areas. These functions are widely used in archaeology (Orton 2005, Bevan and Conolly 2006, Winter Livneh et al. 2010). For more information see Appendix 1.

1.2.2 Local and semi-local methods

These global analyses are a prerequisite for the implementation of local analyses, the second category of ESDA, which are employed for characterizing, identifying and mapping spatial shapes such as *hot spots* or *clusters*. There are two analysis types: some are based on distance measures and others are based on density calculation. We will return to some of these methods later. Some local methods derive from Ripley’s function and its algorithm of distances-calculation (Local L Function, SADIE method). One of the density-oriented methods is the Kernel Density Estimation (or KDE) (Fig. 1.2). KDE estimates the density for each cell in a grid overlaid on the map (Zaninetti 2005, Pillot et al. *in press*). It is also used regularly in archaeology (Baxter and Beardah 1997; Bertonecello et al. 2012). For more information see Appendix 1.

2. Uncertainty and Heterogeneity of Data Events

2.1 Methodological questions and assumptions - Effects of certain factors on results

Typically, point pattern analysis relies on three main assumptions: (1) the observed point pattern is a sample whose relationship with the whole population can be evaluated (completeness of the inventory); (2) the spatial pattern is influenced by the size and the shape of the study area; (3) the spatial pattern is stationary⁶ vs. the archaeological process which is considered as homogeneous, but the space has no impact in the localization of events.

⁶ i.e. the mean and the variance of the point process is constant within the studied area (Bailey and Gatrell 1995, 33)

Methods of point patterns analysis produce results that are based on a number of scientific assumptions. Often the results are generated and compared, though archaeological information is inherently riddled with incompleteness and heterogeneity which is not measurable or quantifiable. However, the assumptions mentioned above, are more or less discussed in the archaeological literature, from methodological and theoretical perspectives (Orton 2000, 2005, Bevan and Conolly 2009, Crema et al. 2010). Nevertheless, the full use of these methods in archaeological research is not questioned. In statistical approaches, these assumptions often remain underestimated or completely ignored (Pillot et al. 2012). We would therefore like to focus on three key issues relating to the use of the point patterns analysis in archaeology: the nature of archaeological data, the role of the study area and the impact of the tested process in the statistical analysis.

2.1.1 The archaeological data

The homogeneity and isotropy assumption does not hold for two reasons:

(1) Archaeological samples are intrinsically non-homogeneous and biased. An archaeological inventory is never complete. We never have access to all the aspects of a studied human phenomenon, because of the number of “filters” (conservation of sites, lack of documentation, etc.). What we study is often not even a sample. We are not always able to estimate the initial population, except by using appropriate statistical methods such as Bayesian approach (Ortman et al. 2007)⁷.

(2) The characterization and the chronology of a large part of these sites are uncertain, without any description of the chronological phases. All these added doubts create “unsuitable” datasets. Yet, there is a dependency between the results and the quality of the data used in statistical analysis (Orton 2005). Often, we are inclined to offset a lower data quality with an accumulation of information (sites poorly dated and poorly characterized), which can generate “noise”, which therefore make it not possible to detect significant levels of *clusters* (Pillot et al. accepted). Almost 10% of sites within a regional inventory of

⁷ In this study, this estimate is difficult considering the heterogeneity of the data.

nearly 5000 *events* have efficient characterization and chronological attribution. Data selection does not always help narrow the results any further. Subsequently, we compared the results of analyses on the entire data set and those of another one with better quality (Pillot et al. accepted, Fig.1). Despite considerable reduction in dataset size and area, the results show similar trends.

2.1.2 Study area

Increasingly, ESDA involves carefully considering the study area. In univariate analysis, the dataset may be of two types: point-referenced data (x-y) or area-referenced data (x-y, A). What is the most appropriate type in the case of archaeological datasets? This question may be easier to answer for documentary research. For these, the study area is often set at the beginning by administrative boundaries; however, it is different for inventories built with data obtained by surface survey. The areas could have some “voids” or they could be discontinuous. Often, Ripley’s K Function is used without any reference area, but under the effect of the change of points pattern (i.e. between two periods), the size and shape of the extent change (Fig. 2). To compare occupation over time, it is necessary to specify and to define the study area. An area larger than the points’ distribution will give clustering. Moreover, in the case of statistical methods, it is advised to apply edge corrections, because *events* at the limits of the study area have got fewer neighborhoods than the others. Several types of corrections are available, such as weighted edge correction (Ripley 1977, Goreaud and Pélissier 1999), “guard area” (Yamada and Rogerson 2003) and the method of theoretical envelope (Yamada and Rogerson 2003, Diggle 2003) that takes into account the shape of the study area. These formulas can be taken into account in the calculation of Ripley’s K function (or Besag’s L). Without edge effects correction, the values of L (r) are much less accurate (Goreaud 2000, Fig. 122g).

2.1.3 The homogeneous process

The ESDA assumption about a homogenous process cannot be realistically applied to archaeological phenomena. A homogeneous process is neutral and has no influence on the spatial distribution of points (Cressie 1993), and its spatial

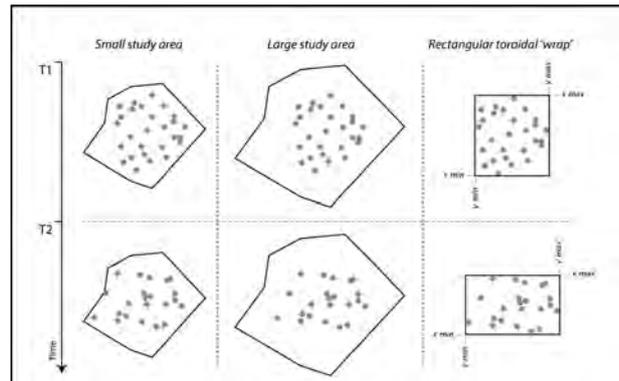


Figure 2. Differences between many types of reference area.

configuration is considered as stationary. But, logically, human occupation is not a homogeneous spatial phenomenon; it is influenced by social and environmental factors that may condition the pattern of human settlements. When archaeologists seek spatial patterns, they want to determine the evidence of a spatial non-randomness of *events* in space, while seeking to identify evidence of specific shapes. To quote an idea mentioned in the ecological field, studied space can be “patchy” space (Dale 1999), as we might expect, because randomness can imply in some studies an absence of behavior, and thus is *a priori* unlikely in the case of a space occupied by humans. Thereby with uncertain archaeological data, and with environmental and social influence on the location of sites, the distribution of archaeological places is clearly heterogeneous. In some cases, the application of the L function, does not detect *clusters*. The calculated values are above the confidence interval even for large distances: the majority of points of datasets would be aggregated in a denser area which can be interpreted as a large cluster whose size is close to the study area. This curve shape is characteristic of a heterogeneous distribution. The use of this function on the basis of this null hypothesis is irrelevant in this case, as pointed out by some authors (Pelissier and Goreaud 2001, Ripley 1981)

2.2 In terms of other disciplines

Our thinking led us to confront two syntheses of the available literature in archaeology and in ecology in which these methods are widely used (Fig. 3.1).

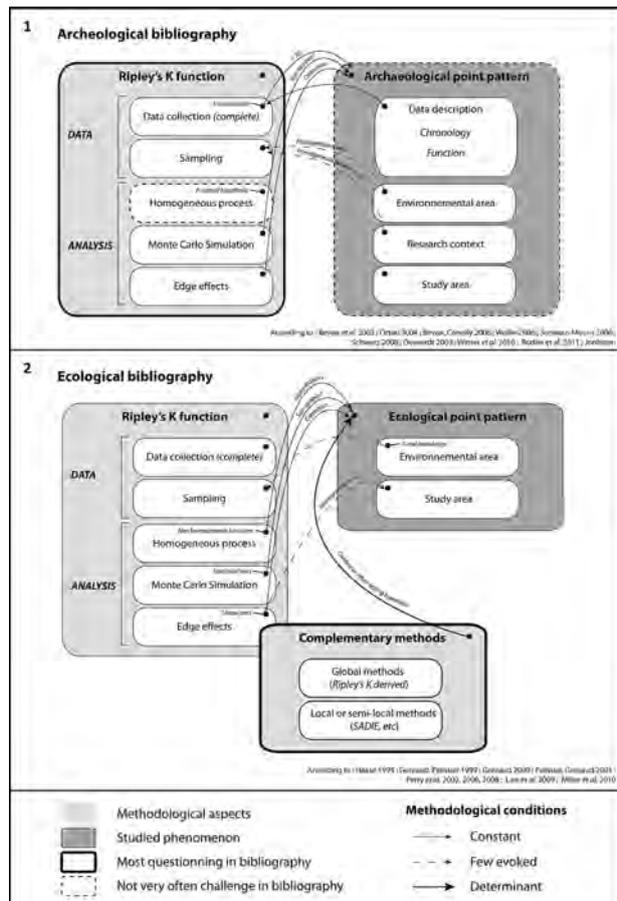


Figure 3. Between archaeological and ecological bibliographies: synthesis intent: 1: Archaeological bibliography; 2: Ecological bibliography.

The methodological requirements mentioned above are generally mentioned in the archaeological approaches, but are not treated in consistent ways. The hypothesis of a homogeneous process is always accepted, although the dataset is not, as in the case of the study area, the research setting, etc... To address this dearth, the authors suggest using homogeneous reference areas with a selection of the dataset based on a homogeneous environment or research context (Bevan et al. 2003, Bevan and Conolly 2006). However, the studied anthropic phenomenon remains innately spatially inhomogeneous (Orton 2005). Nevertheless, the results arise from statistic assumptions. This methodological shortcoming has significant implications for possible interpretations. A few authors have acknowledged these issues and provide alternative approaches (Waller 2006). Although software tools allow a large amount of swift statistical analyzes, it should be noted that

archeologists will risk using them to adopt a perilous “shotgun” approach, as K. R. Schwarz and J. Mount wrote (Schwarz and Mount 2008).

In contrast, the synthesis of the ecology literature shows that the methods of point pattern analysis are both debated and well-determined, in order to study the spatial structure of botanical or faunal species (Fig. 3.2). The various parameters are discussed, as well as their diverse impacts. In ecology, the heterogeneity of datasets is included in the analysis by a generalized use of non-homogeneous process functions (Law et al. 2009, Miller et al. 2010, Perry et al. 2006). So, we suggest using these methods the way the ecologists do, by taking into account all of the aspects of heterogeneity (spatial, studied phenomenon, etc.). With this methodological model we can recommend:

- Testing a non-homogeneous process with using a K Ripley derived function (Miller et al. 2010, Perry et al. 2006);
- Using a multi-scalar approach to observe trends (Bevan and Conolly 2006, Perry et al. 2006);
- Selecting homogeneous geographical, environmental and archaeological sectors (Pélissier and Goreaud 2001, Perry et al. 2006);
- Selecting reliable archaeological data (when the sample allows it);
- Using study area and correction of borders effects (Haase 1996, Goreaud 2000, Pélissier and Goreaud 2001, Perry et al. 2006).

On the other hand, ecologists underline the importance of weighted results, by a confrontation with other methods for analyzing point processes (Law et al. 2009, Perry et al. 2006, 2008). It is essential to apply complementary statistical methods to confirm or refute their respective results.

3. Methodological Proposition: a Case Study

This approach was applied to an example of neolithic sites in the North-East of France (Burgundy). It concerns the study of the changes that affect human occupation during the Neolithic

period. During the Middle Neolithic, around 4700-4300 BC, we observe a break with the traditions of the earlier period. It is the development of the funeral monumentality, through monumental necropolises (Fig. 4. 2). According to the distribution of these monumental necropolises, it is supposed that settlement could be structured in large territorial entities, perhaps dispersed in space.

To observe the specific characteristics of this period, it is necessary to be able to compare them with what is happening both before and after in the same sector. Inventory data for each period (Early and Middle Neolithic) has been utilized (Fig. 4.3).

3.1. Multiscalar approach and sampling

These data are very heterogeneous, despite an exhaustive inventory. They appear intrinsically and spatially non-homogeneous, because of the multiplicity of sources, research disparities, etc. Considering these degrees of heterogeneity, we recommend using various analysis levels, by data selection and by opening multiple and more consistent study areas (Bevan and Conolly 2006, Pillot et al. 2012). Three rectangular regions have been opened in the area. The whole dataset was also analyzed for comparison (Fig. 4: 1). Indeed, the multiscalar approach enables us to limit the effect of the study area. It has the advantage of allowing shifting between smaller and larger scales to refine the reading of the phenomena (Bevan and Conolly 2006, Pillot et al. 2012). This method enables us to observe and compare the distribution and structure of *events* at different scales.

3.2. Derivatives of R or L functions

The data of the study areas always remain incompletely distributed; it is difficult to consider their distribution significantly as representative of any underlying social phenomena. To improve the results and bring them closer to the archaeological realities, it seems worth adapting the hypothesis tested in the statistical analysis. It is recommended to use a method that is very commonly used in ecology: derivatives of R or L functions which allow the comparison of datasets with heterogeneous distributions⁸. These functions use the intensity

⁸ Distribution composed of discrete random variables in two dimensions, whose intensity, number and location vary in space

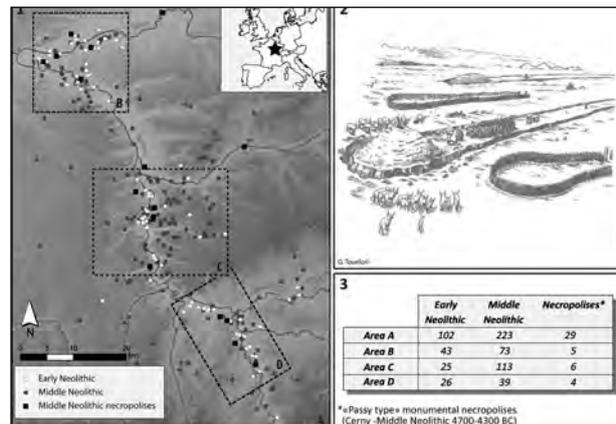


Figure 4. Studied data: 1: Location map; 2: Attempt of restoration of "Passy type" necropolises; 3: Data table.

estimation for each *event* (Baddeley et al. 2000). For more details on the use of the method, we refer to the work in ecology of J.N. Perry (Perry et al. 2006). These functions allow us to observe clustering trends over smaller distances.

In the case of neolithic sites, given the low number of gathered samples and the highly irregular data distribution, K and L functions results were invalid: the hypothesis of regularity is not assumed and therefore not verified. Using the inverse function could help solve this problem (hypothesis of non-homogeneity). The Besag function, with a hypothesis of non-homogeneous process, was applied by means of the already mentioned explicit parameters⁹. In this case, the results were more conclusive. The small scale results, however, go against those obtained on the different large scale windows (Fig. 5). There is some consistency between these three smaller areas (B, C, D), with a lower trend of data aggregation, during the Middle Neolithic, despite more important samples. The small scale results can be explained by the disjointed distribution of the data: many areas are in fact unoccupied. The increasing number of studied areas helps reinforce the identification of a trend. Of course, the aggregation distances vary because of the scale changes. But, globally, for the Middle Neolithic, the distribution seems more dispersed with the largest aggregation distances (Fig. 5).

(the number of points per unit square changes)

⁹ Methodological parameters: correction of edge effects, (correction-isotropic- Ripley 1977), using the study areas (Pélissier and Goreaud 2001) and setting the number of simulations (Perry et al. 2006, Pillot et al. accepted).

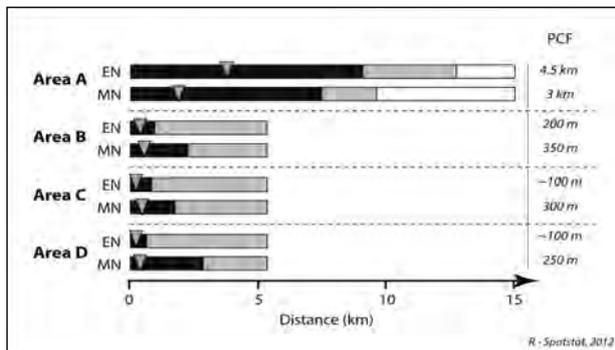


Figure 5. Comparison of Ripley's K and PCF results between all studied areas: synthesis of Ripley's K correlograms and PCF results.

3.3. Complementary analysis

The use of these tests alone is biased. Indeed, a large battery of statistics is available for analyzing spatial point patterns. The judicious use of a combination of tests is recommended. It may well be the most informative approach; both Ripley (Ripley 1981) and Diggle (Diggle 2003) state that the various statistical tests should be used to complement each other.

3.3.1 Global methods

A problem with Ripley's K or Besag L, is that of 'cumulative effects'. Because Ripley's K is cumulative, values of the statistic at a given distance may be unduly influenced by its values at smaller distances (Ward et al. 1996, Perry et al. 2006). Also, it is possible to use other methods such as the Pair Correlation Function (PCF). This method is similar to Ripley's K except that its distance classes are *annuli*, not circles (Stoyan and Stoyan 1994). The clustering trend observed *via* Ripley's, is confirmed by other statistical indicators, such as the PCF (Fig. 5). Its result correlates significantly with the peak of the Besag L function curve. The global approach enables to confirm a general trend of dispersion in space. This finding relegated one observation already made concerning the transition between the Middle and Late Neolithic in Burgundy (Pillot et al. accepted). The presence of clusters of sites during the Middle Neolithic is not excluded. The peak of L function and the PCF results might show an ability of spatial segregation of clusters of sites. It therefore calls for an examination of the place of the monumental

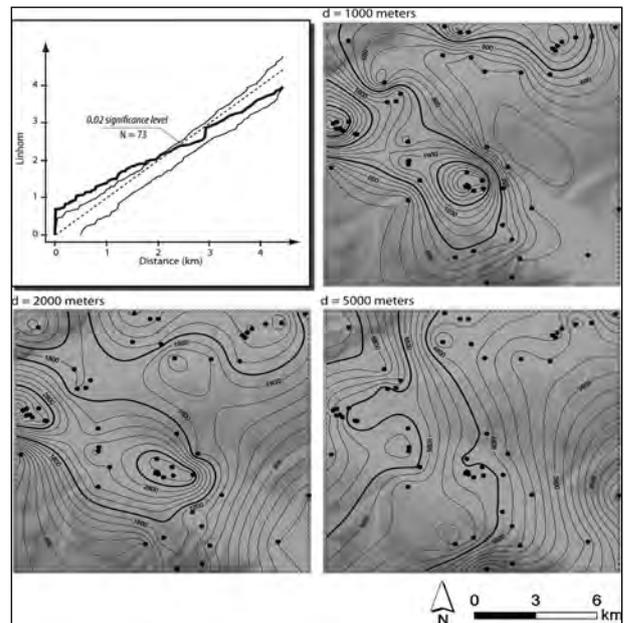


Figure 6. Contour map of Getis and Franklin's $L(d)$ analysis for the area B during the Middle Neolithic (above : Besag's L analysis for inhomogeneous data ; others: contour maps where $d = 1 \text{ km}, 2 \text{ km}$ and 5 km). Black tick lines: d value and higher.

cemeteries in space. For this, it is necessary to try to locate the aggregates with local methods. Indeed, global methods provide an average description of the spatial configuration. The problem is that it underestimates the local variations in the spatial configuration of the *events* (Fotheringham 1997).

3.3.2 Local and semi-local methods

Local statistical methods are highly complementary of the global ones. They allow to confirm or refute the observed clustering trend, and especially to locate clusters. There is a set of tests based on distances calculations.

3.3.3 Local L Function

These results are confirmed by other statistical methods such as the Local L function. Indeed, it is possible to use local methods derived from Ripley's K function such as Getis and Franklin's (1987) $L(d)$ statistic. It is similar to Ripley's K, except that a value is calculated and mapped for each individual *event*. This statistic is valid even if the point pattern is inhomogeneous. For more information about the method, see appendix 1. Thus this local form can be

	period	n	OD	RD	Pp	Ip
Area A	EN	102	19.3	23.39	0.6	0.825
	MN	223	43.6	38.34	0.2	1.137
Area B	EN	43	6	6.4	0.56	0.934
	MN	73	6.4	4.4	0.51	1.468
Area C	EN	25	NS	NS	NS	NS
	MN	113	14.6	10.8	0.5	1.347
Area D	EN	26	0.8	0.9	0.51	0.903
	MN	39	0.5	0.7	0.71	0.7

Table 1. Summary statistics of SADIE results: *n*=no. of events, *OD*=observed distance to regularity, *RD* = distance to regularity for CSR data, *Pp*=probability of as extreme aggregation under CSR, *Ip*=index of aggregation.

used to indicate local areas of non-stationarity and/or to explore the influence of individual events on the fullness of the global statistic. For our example, at large scale resolution, it is possible to discern local variations (Fig. 6). With this analysis, for each studied areas, clusters seem significant up to 2 km distances ($d = 2000$ meters). Clusters of Middle Neolithic are bigger and more densely occupied than other periods. They are larger and denser.

3.3.4 Spatial Analysis by Distance Indices (SADIE)

In the end, the application of one combined global-local approach continues to confirm our results. It is a class of methods called Spatial Analysis by Distance Indices (SADIE). The strength of these methods comes from their ability to describe and map local variation of spatial pattern and association. For more information about the method, see appendix 1. This method produces the index of aggregation I_p^{10} (Perry 1995). This index enables to evaluate clustering trends. At local scales, SADIE can produce visual plots: the 'initial and final' (IAF) plots. In our case, the trend of clustering seems to be confirmed by the I_p of SADIE analysis (Table 1). Globally, in each study area, the Middle Neolithic's I_p value is superior of that of the Early Neolithic ($I_p \geq 1$ except for area D). IAF plots show more complex patterns for the Middle Neolithic: the clustered nature of the pattern is revealed by line segments that show ranges of lengths. These

¹⁰ It corresponds to the ratio of the mean distance between events that changed over the observed period relative to simulated CSR data.

segments are larger -EN : 2.5 ± 5.6 km ; MN : 4.3 ± 5.2 km. It is possible to make the assumption of the presence of larger clusters during the Middle Neolithic, a period during which they are also denser.

3.3.5 Kernel Density Estimation (KDE)

Another local method is based on kernel density estimation: the KDE. The kernels' densities were calculated for the various zones as well as their PVC (Fig. 7). After the use of various statistical indicators, we determine a bandwidth distance of 1.5 km¹¹. Globally, for each large-scale area, we should observe the same trend: during the Middle Neolithic, occupations are more dispersed but in a rather dense way. They form large clusters (more than 3 km around sites). These aggregates are dispersed enough in space. The 50 PVC, (thick white lines in the map) pinpoint areas which represent 50% of the volume of the density surface data. Most of the 50 PVC are well-insulated from the rest of the spatial distribution data. The Middle Neolithic 50 PVC seem, in part, to be aggregated with the clustering areas, constituting larger areas with the highest density. For the Middle Neolithic, the average size of the zones surrounded by the 50 PVC, is superior to those of the Early Neolithic. Furthermore we observe larger inter-distances between PVC of the Middle and of the Early Neolithic. These areas seem more dispersed in space, but also denser.

3.3.6 Change map

The realization of a change map confirms the trend (Fig. 8). A change map between two periods, the Middle Neolithic and the previous period enables us to consider the transformations in the point patterns. We thus notice that sites are more dispersed in the Middle Neolithic, with an intensification of the human occupation. There is persistence of occupation in various areas. There is even disappearance of sites. In clustered areas, we can notice the presence of necropolises. We can suppose their attraction for other occupations. Moreover, necropolises are situated in long-occupied areas. Globally, during the Middle Neolithic there

¹¹ The estimation made subjectively, by 'naked-eye observation' is nearest to the index of Cross-Validation or to the peak of the K Ripley function, so between 1.5 and 2.3 km around sites.

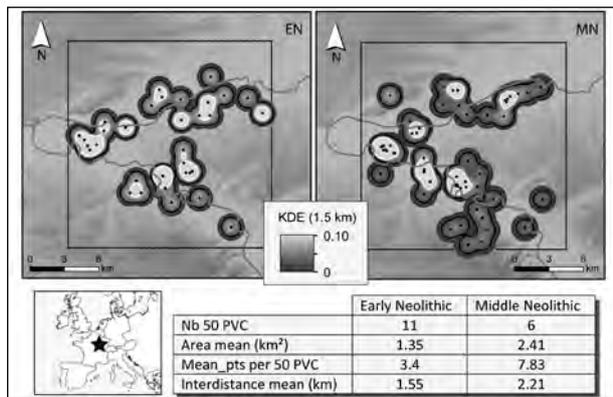


Figure 7. Kernel Density Estimation for the area B during the Early Neolithic (EN) and the Middle (MN) Neolithic (bandwidth=1.5 km).

is a reinforcement of most occupied areas (except for the area B). These sites were to have a role in structuring settlement patterns.

4. Discussion

4.1. Data heterogeneity

Heterogeneity is a characteristic of anthropic systems that can be observed both in space and time. There are a lot of processes that affect human dynamics and thus settlement spatial-structure, such as environmental influence (topography, hydrography, etc) or social behavior. Also, point processes that are behind archaeological point pattern do not meet homogeneity and stationarity criteria for the application of spatial statistics. To quote the comments made by A. Barcelo (Barcelo and Maximiano 2007), human occupation of space (intentional), cannot necessarily generate random distributions. There is more often non-random distribution (i.e., non-homogeneous) that fall between clustered and uniform patterns. Spatial point patterns that vary in a systematic way from place to place are thus called heterogeneous (Ripley 1981). Analysis of the spatial structure of heterogeneous point patterns is difficult, because the simple methods used to analyze spatial point patterns have been developed for homogeneous point patterns, i.e. for patterns resulting from stationary point processes (Goreaud and Pélissier 2001). Indeed, these methods often use indexes or functions that are averaged over the whole study area and thus only make sense for homogeneous processes. We think that using more realistic

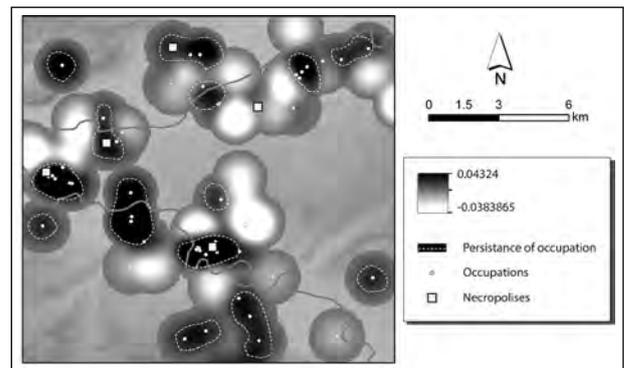


Figure 8. Change Map between Early Neolithic and the Middle Neolithic for the area B: intent of interpretation.

models for hypothesis testing (i.e. beyond the CSR) may also help (Wiegand and Moloney 2004). The methodology described above show that it is possible to partly compensate for these difficulties by using inhomogeneous forms of K or L functions: for non-stationary processes, the null hypothesis of CSR should be represented by inhomogeneous Poisson processes. The results are refined even more if you select more homogeneous areas in the dataset. This is all the more restrictive than the studied spatial pattern can be logically simplified as a series of *events* located in a two-dimensional Euclidean space. With spatial analysis, we compress observations into “temporal layers” assumed to be homogeneous (temporal information seen as ceramic typology). Nevertheless, ultimately it adds a degree of uncertainty. Indeed, temporal uncertainty adds to the uncertainty levels of the problems already mentioned. Finally, the datasets themselves are a mass of uncertainties. It becomes difficult to adopt an unbiased exploratory approach. One may thus wonder whether spatial analysis in this context is still justified.

4.2. Archaeological interpretation: spatial pattern description vs. mechanism: what can we say?

The appropriateness of spatial analysis has been challenged in other disciplines such as ecology, over two major issues (Real and McElbany 1996):

(1) The detection and assessment of spatial patterning.

(2) The identification of the mechanisms that result in the formation of different types of spatial patterns

At first, we can say that, overall, the implementation methodology used does reveal trends. This enables researchers to describe various aspects of generic non-randomness in spatial data and to compare it over time. It is possible to identify clusters that are themselves organized into spatial and territorial shapes. Secondly, statistical approaches permit the testing and the exploration of archaeological data. Indeed, ESDA are informative and useful devices for finding, verifying and visualizing spatial data patterns (Bevan and Conolly 2006). We notice that there are two main issues: (1) the use of *a priori* hypotheses which are explicitly tested *via* the analysis, and (2) the produce of testable hypotheses as an outcome of the spatial analyses. We agree with the idea developed by J.N. Perry that using point pattern analysis serves to explicitly test stated hypotheses. It is a “useful means of avoiding data-dredging and is more likely to provide useful information on mechanism than an *ad hoc* approach” (Perry et al. 2006). In our case, we have seen that there are possible territorial transformations between Early and Middle Neolithic. *A priori*, the necropolises could perform an important role in the settlement pattern. ESDA allow us to make another *ad hoc* hypothesis: the presence of bigger clusters of sites, in which necropolises have a central role.

5. Conclusions

Our approach outlines an exploratory and comparative method. We propose to take into account the heterogeneity of data and the “non-homogeneity” of the tested process. It is noteworthy that the previously used approaches can be improved by adjusted data models to more realistic assumptions (derived from Ripley’s K for non-homogeneous data). It is worth increasing the analysis levels to confirm trends (multiscalar approach). Other approaches can refine the identification of trends through selection of datasets that would be both more “reliable” and more comparable with each other. The choice of reference areas and the correction of edge effects also lead to gaps. One of the major points of this paper is to show that each result should be compared to other analyzes: this makes it more meaningful and therefore more interpretable. Eventually, it would be useful to refine the methodology by using other models of non-homogeneous processes and to study

them in terms of archaeological issues. Of course, many limitations remain, particularly regarding the incorporation of temporal uncertainties. An exploratory approach should be conducted within this framework to lead to untangling the nodes of uncertainties about the archaeological record.

Acknowledgements

We are grateful to the designers of the ‘spatstat’ Package for R-Cran and SADIE Shell (v.13). We especially thank L. Granjon and E. Le Goff for their participation in this project. We also thank B. Bohard, E. Camizuli, and F. Monna for their advice on the R software. We thank the two proofreaders (S. Ferdjani and M. Dessaigne) whose thorough reviews helped to improve the paper. The preparation of this paper was assisted by the research program “PCR Nécropoles Préhistoriques et Protohistoriques de Passy” (coord. O. Lemerrier), funded by the Ministère de la Culture et de la Communication (France).

References

- Anselin, L. 1994. “Exploratory Spatial Data Analysis and Geographic Information Systems.” In *New Tools for Spatial Analysis*, edited by M. Painho, 45-54. Luxembourg: Eurostat.
- Anselin, L. 1999. “Interactive Techniques and Exploratory Spatial Data Analysis.” In *Geographical Information Systems: Principles, Techniques, Management and Applications*, edited by P. A. Longley, M. F. Goodchild, D. J. Maguire and D. W. Rhind, 251-164. New York: Wiley.
- Baddeley, A., J. Møller, and R. Waagepetersen. 2000. “Non- and semi-parametric estimation of interaction in inhomogeneous point patterns.” *Statistica Neerlandica* 54: 329-350.
- Banos, A. 2001. “A propos de l’analyse spatiale exploratoire des données.” *Cybergeo, Systèmes, Modélisation, Géostatistiques*, 197. Accessed 1 September 2009. <http://www.cybergeo.eu/index4056.html>.
- Barceló, J.A., and A. Maximiano. 2007. “Some Notes Regarding Distributional Analysis of Spatial Data.” In *Layers of Perception, Proceedings of the 35th International Conference on Computer Applications and Quantitative Methods in Archaeology (CAA)*, edited

The Evolution of Territorial Occupation
Lucile Pillot and Laure Saligny

- by A. G. Posluschny, K. Lambers and I. Herzog, 284-289. Bonn: Kolloquien zur Vor- und Frühgeschichte.
- Baxter, M.J., and C.C. Beardah. 1997. "Some Archaeological Applications of Kernel Density Estimates." *Journal of Archaeological Science* 24: 347-354.
- Bertoncello, F., E. Fovet, C. Gandini, F. Trément, and L. Nuninger. 2012. "The spatio-temporal dynamics of settlement patterns from 800 BC to 800 AD, in Central and Southern Gaul: models for an interregional comparison over the long term." In *Settlement Patterns, Production and Trades from the Neolithic to the Middle Ages. ArchaeDyn*, edited by C. Gandini, F. Favory and L. Nuninger, 141-148. Oxford: Archaeopress.
- Besag, J., P.J. Diggle. 1977. "Simple Monte Carlo tests for spatial pattern." *Applied Statistics* 26: 327-333.
- Bevan, A., C. Frederick, and A. Krahtopoulou. 2003. "A digital mediterranean countryside: GIS approaches to the spatial structure of the post-medieval landscape on Kythera (Greece)." *Archaeologia e Calcolatori* 14: 217-236.
- Bevan, A., and J. Conolly. 2006. "Multiscalar Approaches to Settlement Pattern Analysis." In *Confronting Scale in Archaeology: issues of Theory and Practice*, edited by G. Lock and B. Molyneaux, 217-234. New York: Springer.
- Bevan, A., and J. Conolly. 2009. "Modelling spatial heterogeneity and non-stationarity in artifact-rich landscapes." *Journal of Archaeological Science* 36: 956-964.
- Crema, E.R., A. Bevan, M. Lake. 2010. "A probabilistic framework for assessing spatio-temporal point patterns in the Archaeological record." *Journal of Archaeological Science* 37: 1118-1130.
- Cressie, N. 1993. *Statistics for spatial data*. New York: John Wiley & Sons.
- Dale, M.R.T. 1999. *Spatial pattern analysis in plant ecology*. Cambridge: Cambridge University Press.
- Diggle, P.J. 2003. *Statistical analysis of spatial point patterns*. London: Arnold.
- Fletcher, R. 2008. "Some spatial analyses of Chalcolithic settlement in southern Israel." *Journal of Archaeological Science* 35: 2048-2058.
- Fotheringham, A. S. 1997. "Trends in quantitative methods I: stressing the local." *Progressive Human Geography* 21: 88-96.
- Gatrell, A.C., T.C. Bailey, P.J. Diggle, and B.S. Rowlingson. 1995. "Spatial Point Pattern Analysis and Its Application in Geographical Epidemiology." *Transactions of the Institute of British Geographers* 21 (1): 256-274.
- Gauthier, E., O. Weller, L. Nuninger, with the collaboration of M. Gabillot and B. Quilliec 2012. "The search for a common methodology for studying the spatial dynamics of material and product circulation in ancient times." In *Settlement Patterns, Production and Trades from the Neolithic to the Middle Ages. ArchaeDyn*, edited by C. Gandini, F. Favory and L. Nuninger, 65-76. Oxford: Archaeopress.
- Getis, A., and J. Franklin. 1987. "Second-order neighbourhood analysis of mapped point data patterns." *Ecology* 68: 473-477.
- Goerlich G., and F. José. 2003. "Weighted samples, kernel density estimators and convergence." *Empirical Economics* 28 (2): 335-351.
- Goreaud, F. 2000. "Apports de l'analyse de la structure spatiale en forêt tempérée à l'étude et la modélisation des peuplements complexes." Phd Diss, ENGREF.
- Goreaud, F., and R. Péliissier. 1999. "On explicit formulas of edge effect correction for Ripley's K-function." *Journal of Vegetation Science* 10: 433-438.
- Haase, P. 1995. "Spatial pattern analysis in ecology based on Ripley's K-function—introduction and methods of edge correction." *Journal of Vegetation Science* 6: 575-582.
- Hodder, I., and C. Orton. 1976. *Spatial Analysis in Archaeology*. Cambridge: Cambridge University Press.
- Kintigh, K.W., and A.J. Ammerman. 1982. "Heuristic approaches to spatial analysis in Archaeology." *American antiquity* 47: 31-63.
- Law, R., J. Illian, D.F.R.P. Burslem, G. Gratzler, C. V. S. Gunatilleke, and I. A. U. N. Gunatilleke. 2009. "Ecological information from spatial patterns of plants: insights from point process theory." *Journal of Ecology* 97: 616-628.

- Miller B.P., G.L.W. Perry, N.J. Enright, and B.B. Lamont. 2010. "Contrasting spatial pattern and pattern-forming processes in natural vs. restored shrublands." *Journal of Applied Ecology* 47: 701-709.
- Morgan, C. 2009. "Climate change, uncertainty and prehistoric hunter-gatherer mobility." *Journal of Anthropological Science* 28: 382-396.
- Niknami, K.A., and A.C. Amirkhiz. 2008. "A GIS technical approach to the spatial pattern recognition of Archaeological site distributions on the Eastern Shores of Lake Urmia, Northwestern Iran." In *the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. XXXVII. Part B4*, edited by J. Chen, J. Jiang and S. Nayak, 167-172. Beijing: Organising Committee of the XXI International Congress for Photogrammetry and Remote Sensing.
- Nuninger, L., L. Saligny, K. Ostir, N. Poirier, E. Fovet, C. Gandini, E. Gauthier, Z. Kokalj, and F. Tolle 2012. "Models and tools for territorial dynamic studies." In *Settlement Patterns, Production and Trades from the Neolithic to the Middle Ages. ArchaeoDyn*, edited by C. Gandini, F. Favory and L. Nuninger, 23-38. Oxford : Archaeopress.
- Ortman, S.G., M.D. Varien, and T.L. Gripp, 2007. "Empirical Bayesian Methods for Archaeological Survey Data: An Application from the Mesa Verde Region". *American Antiquity* 72(2): 241-272.
- Orton, C. 2000. *Sampling in Archaeology*. Cambridge: Cambridge University Press.
- Orton, C. 2000. "Point pattern analysis revisited." *Archeologia e Calcolatori* 15: 299-315.
- Pélissier, R., and F. Goreaud. 2001. "A Practical Approach to the Study of Spatial Structure In Simple Cases of Heterogeneous Vegetation." *Journal of Vegetation Science* 12: 99-108.
- Perry, J.N. "Spatial analysis by distance indices." *Journal of Animal Ecology* 64: 303-314.
- Perry, G.L.W., B.P. Miller, and N.J. Enright. 2006. "A comparison of methods for the statistical analysis of spatial point patterns in plant ecology." *Plant Ecology* 187: 59-82.
- Perry, G.L.W., N.J. Enright. B.P. Miller, and B.B. Lamont. 2008. "Spatial patterns in species-rich sclerophyll shrublands of southwestern Australia." *Journal of Vegetation Science* 19: 705-716.
- Pillot, L., and L. Saligny. 2012. "L'évolution de l'occupation humaine: l'analyse spatiale exploratoire des données appliquée à différents cas d'étude." In *Actes du Colloque International d'Archéologie et d'Histoire d'Antibes (Antibes, France, october 2011)*, edited by Bertonecello F., Braemer F., 155-173. Antibes: Editions APDCA.
- Pillot, L., L. Saligny, and C. Moreau. *In press*. "Observations of land use transformations during the Neolithic using Exploratory Spatial Data Analysis: contributions and limitations." *CAA 2010, Fusions of Cultures* (Grenada, Spain, April 2010).
- Poirier, N. 2010. *Un espace rural à la loupe. Paysage, peuplement et territoires en Berry, de la préhistoire à nos jours*. Tours: Presses Universitaires François-Rabelais.
- Real, L.A., P. McElhany. 1996. "Spatial pattern and process in plant-pathogen interactions." *Ecology* 77: 1011-1025.
- Ripley, B.D. 1977. "Modeling Spatial Patterns." *Journal of the Royal Statistical Society* 2: 172-212.
- Ripley, B.D. 1981. *Spatial Statistics*. New York: John Wiley and Sons.
- Schwarz, K. 2008. "Oak Openings Archaeology: Spatial Statistical Discernment of Late Archaic Camps." Paper presented at the 2008 OAC Fall Meeting.
- Schwarz, K.R., and J. Mount. 2005. "Integrating spatial statistics into Archaeological data modeling." In *GIS and Archaeological Site Location Modeling*, edited by M. W. Mehrer and K. L. Wescott, 167-189. Illinois: CRS Press.
- Silverman, B. W. 1986. *Density estimation for statistics and data analysis*. London: Chapman and Hall.
- Stoyan, D., and H. Stoyan. 1994. *Fractals, random shapes and point fields: methods of geometrical statistics*. Chichester: John Wiley & Sons.
- Waller, L.A. 2006. "Detection of clustering in spatial data." In *The Sage Handbook of Spatial Analysis*, edited by A. S. Fotheringham and P. A. Rogerson, 299-320. London: SAGE Publications.

The Evolution of Territorial Occupation
Lucile Pillot and Laure Saligny

Ward, J.S., G.R. Parker, and F.J. Ferrandino. 1996. "Long-term spatial dynamics in an old-growth deciduous forest." *Forest Ecology Management* 83: 189–202.

Wiegand, T., and K.A. Moloney. 2004. "Rings, circles, and nullmodels for point pattern analysis in ecology." *Oikos* 104: 209–229.

Winter L.R., T. Svoray, and I. Gilead. 2010. "Settlement patterns, social complexity and agricultural strategies

during the Chalcolithic period in the Northern Negev (Israel)." *Journal of Archaeological Science* 37: 284-294.

Yamada, I., and P.A. Rogerson. 2003. "An empirical comparison of edge effect correction methods applied to K-function analysis." *Geographical Analysis* 37: 95-109.

Zaninetti J.-M. 2005. *Statistiques spatiales: méthodes et applications géomatiques*. Paris: Lavoisier, Hermès Sciences.

	Test statistic	Description	References
Global methods	Ripley's K Function and its Besag L transformation	Considers the complete distribution of all the distances in the point pattern. The distribution is compared to a reference distribution called a homogeneous Poisson process under complete spatial randomness (calculated with Monte-Carlo Simulation) and the function tests three hypotheses: uniqueness, homogeneous stationarity and isotropic process. Method: Number of events within a circle of radius sequentially larger t from each focal event, and deviation from expectation at each t under CSR.	Ripley 1977, 1981 Besag and Diggle 1977
	Pair Correlation Function (PCF)	The average distance at which the Pair Correlation Function suggests a random distribution. It is a local measurement of the shape and configuration of point patterns (to determine the size of clusters). The PCF is also known as the Neighborhood Density Function (Ward et al. (1996)) and the O-ring statistic (Wiegand and Moloney 2004).The PCF is similar to Ripley's K except that it is non-cumulative. Method: Number of events within distance classes that are annuli, not circles.	Stoyan and Stoyan 1994 Perry 2006
Local methods	Kernel Density Estimation	Non-parametric method of estimating the probability density function based on point location. Method: Calculation of density for each cell of a grid overlaid on the map. Measures the distance between each cell and each point and determines the weight for the cell. The density estimate is the sum of each weight. The weight depends on: the distance from cell to point; the radius or bandwidth around each point, and the method of interpolation or function.	Silverman 1986
	Getis and Franklin's L(d)	Local statistics (Local form of Ripley's K) can be displayed as contour maps or, with the contours based on the values of L(d) for each individual at distance (length scale) di contours where $d < L(d)$ denote areas of regularity and those where $d > L(d)$ aggregation. The visualization advantages are clear—while the global tests above suggest there is spatial segregation and at which scale(s), the local tests can explicitly show where this is occurring. Method: L(d) is calculated for each event individually, providing information concerning local trends in pattern (e.g. areas of aggregation vs. areas of regularity in the same plot)	Getis and Franklin 1987
	Spatial Analysis by Distance indicEs (SADIE)	Based on an algorithm in which the observed events are iteratively moved until they achieve a regular arrangement. The 'distance to regularity' is assessed by summing the number of moves each event undergoes until regularity is achieved. The ratio of the mean distance moved during the observation relative to simulated CSR data is termed the index of aggregation Ip. This index enables to evaluate clustering trend. As well as this simple index of spatial pattern, SADIE provides "diagnostic" plots, which is a map showing the original and the rearranged patterns with line segments connecting pairs of points (the 'initial and final' (IAF) plots). Method: Calculates the 'distance' from the event set to regularity by moving events until a regular Voronoi tessellation is achieved; allows calculation of an aggregation index (Ip).	Perry 1995

Appendix 1

Visualising Time with Multiple Granularities: a Generic Framework

Iwona Dudek and Jean-Yves Blaise

CNRS, UMR CNRS/MCC 3495 MAP

Abstract:

When investigating the evolution of historic architecture, and putting together various pieces of information (each with its specific characteristics in terms of precision, scope and reliability), time points and intervals the analyst will identify are often inconsistent in terms of granularity. Our contribution introduces graphic solutions that combine multiple aspects of the parameter time, and particularly multiple granularities. As a first step, we initially propose a visual comparison of 25 alternative calendars, covering a wide range of historic periods and cultures or civilisations. This first result is then extended to propose a more generic framework for visualising time with multiple granularities. It is applied on two very different test cases. The contribution will present the concepts and ideas behind this research, as well as their practical applications on the tests cases and accordingly their possible benefits for researchers and practitioners in historic sciences.

Keywords:

Time-oriented Data, Information Visualisation, Historical Calendars

1. Introduction

Investigating the evolution of historic artefacts most often starts with the cumbersome task of putting together various pieces of information, each with its specific characteristics in terms of precision, scope and reliability. Naturally, time slots are among the main clues analysts expect to spot when filtering and cross-examining these pieces of information.

In order to proceed to any type of reasoning (teleological or causal) one has to place all the data and pieces of information. But due to the very nature of historic data sets – heterogeneity, uncertainty, missing data, uneven distribution in time (*etc.*) – time points and intervals the analyst will identify are often inconsistent in terms of granularity. Time intervals, typically the overall lifetime of an artefact, its periods of construction or modification, may be described by expressions like “*between the last quarter of the XIIIth c. and the middle of the XIVth c.*”, whereas some punctual events may be recorded more precisely, in cases like “*the fire that occurred on the night of November 29th 1554*” or “*the town’s siege between March 1445 and November 1445*”.

In parallel, describing an artefact’s life evolution often implies taking into consideration pieces of information that correspond to regular

Corresponding author: idu@gamsau.map.archi.fr

or cyclic events, with here also inconsistent granularities. Typically, when analysing an isolated chapel at high altitude, the analyst will need to cope with a fuzzy cyclic behaviour – the chapel is inaccessible due to snow for a certain number of weeks during the year – as well as with a well-defined cyclic behaviour – a pilgrimage is organised on the saint’s day every year.

In other words, may it be because of the nature of historic data sets, or may it be because of the heterogeneity of the events we need to report, there are very few solutions analysts can count on if they need to visualise in a consistent, insight-gaining manner the time slots they have spotted.

In historic sciences it appears clearly that the handling of multiple time granularities is one of the major bottlenecks in the analyst’s visualisation effort. Naturally, conveying properly doubts is an even harder challenge – with numerous overlapping issues, as will be shown. Our research aims at giving analysts means to combine in a single visualisation multiple aspects of the parameter time, and particularly multiple granularities.

As a first exploratory step, we initially focused on a visual comparison of 25 alternative calendars covering a wide range of historic periods and cultures or civilisations (Julian and Gregorian calendars

of course, but also Babylon, ancient Egypt, China and Japan, Incas, Inuits, Burma and Bali, *etc.*). The visualisation sums up in a synthetic way key aspects of calendars (cycles, divisions, period of validity, area of validity, the intercalation – *i.e.* correction mechanisms, *etc.*) corresponding to alternative time granularities. The visualisation helps underlining legacies in between periods and areas, alternative visions of time as linear or cyclic, common or opposing choices like solar/lunar/lunisolar, alternative divisions of the year, mechanisms to cope with intercalations, *etc.*. A prominent service offered by this visualisation is that it enables comparisons at various time granularities, “*within the eyespan*”, to quote E.R Tufte.

This first result has been extended to propose a more generic framework for visualising time with multiple granularities. It is applied on two test cases representing different scales and territories. The contribution will present the concepts and ideas behind this research, as well as their practical applications on the tests cases and accordingly their possible benefits for researchers and practitioners in historic sciences.

2. Research Context

Reasoning on the historic artefacts requires a deeper understanding of the time parameter than what we understand of it in everyday life. And if in addition we want to use visualisations to inquire into what is behind our time-oriented data, we need even more extensive understanding, analysis and control over the time parameter.

In parallel, Information Visualisation and Visual Analytics are multidisciplinary fields that are increasingly applied as a critical component in scientific research where *visual reasoning* is a relevant method (*i.e.* analytical method that employs human perception in order to inform, reveal unknowns or as a part of the thinking process itself (MacEachren et al.2005,139–160)). In other words visualisation is about : ... *pictures of numbers ... pictures of nouns ... pictures of verbs, the representation of mechanism and motion, of process of dynamics, of causes and effects, of explanation and narrative...* (Tufte 1997, 126).

This is probably the reason why the time

parameter has been carefully studied and structured notably in and around the above fields - see for example (Tufte 1990, 151), (Chardonnel 2007) (Knight 1993, 401-419), (Perlata 2004, 241-248), (Matoušek 2007, 239–254), (Allen 1984, 124-154) - although it doesn't mean that we at this stage fully understand its nature.

As demonstrated in (Aigner et al. 2008, 47-60) multiple aspects of time can be taken into account : time progression (linear time *vs.* cyclic time), time structure (ordered time, branching time, time with multiple perspectives), temporal entities and their relations (time points, time intervals), temporal scale (ordinal time, discrete time, continuous time, discontinuous time), time granularity, uncertainty of temporal position of temporal entities, *etc.*

However, as sagely remarked “*In theory, there is no difference between theory and practice. But in practice, there is. (...) Theory and practice both have limitations. Sometimes, practice proves the value of a good theory ...*” (West 2010).

2.1 Theory vs practice

Theoretical approaches are crucial in understanding the time parameter, still they are not the only solution to the problem. Another approach can be observing facts: here, historic calendars.

Calendars are basically an effort to position oneself in time, by observing and measuring objectively one or several phenomena that occur repeatedly, on a regular basis – may it be the course of planets, hunting seasons, market days in the neighbouring cities. In other words, understanding calendars means understanding how time is seen by a society, in a given space. This way calendars tell us something about societies, but in addition they also demonstrate the complexity of the time parameter, with several repetitive cycles to combine in a way that has to be understood by everyone (night & day, seasons & years, cultural events, *etc.*).

Finally, calendars are used not only to position oneself in absolute time, but also to position oneself with regards to other specific temporal slot, may they be “*how many months since my last check-up at the dentist*” or less depressing “*how many grand-fathers since Copernic's discovery*”? (account in

grandfathers, as a tribute to T. Pratchett). Naturally this applies both to past and to future events. Accordingly calendars are *de-facto* examples of (at least) three key issues when talking about time:

- positioning a time slot inside a continuum,
- handling alternative cycles, granularities, rhythms,
- ordering, comparing, analysing time slots and the way they are depicted here and there, and trying to make sense out of it.

In this contribution, we will focus on historic calendars, *i.e.* calendars that have been introduced at sometime in the past, have evolved (Julian to Gregorian for instance), ceased to be used (Coligny Gaulish for instance), or remained unchanged (Chinese calendar for instance). The word “historic” might be here a bit confusing, however we have chosen it in order to insist on two aspects that are of importance:

- Handling calendars that are not used any more implies to some extent handling something known to us through testimonies, (*i.e.* questionable data sets, and accordingly implies confidence assessments).
- Handling calendars that have succeeded to one another – may we know of it or not – implies highlighting parenthoods on one hand (Babylonian – Greek; Egyptian – French Republican) and on the other hand calls attention to the consistency of dating (shift of the new year in the Roman calendars).

Let us illustrate a classic implication of this last point with some well-known example:

To honour the date of death of Cervantes, Shakespeare and De la Vega, who supposedly died on the same day, April 23, 1616, UNESCO established April 23rd as the International Day of the Book (UNESCO 2011). But Shakespeare died on a *different* day than Cervantes and de la Vega – the date of his death (April 23, 1616) is given according to the Julian calendar, and not in the Gregorian calendar, already adopted by Spain but not by England in 1616. Since at that time the Gregorian

calendar was ten days ahead of the Julian, the Spanish authors actually died ten days *earlier* than Shakespeare, whose date of death according to the Gregorian calendar was May 3, 1616. In other words, we are officially asked to remember a fact – three main authors dying the very same day – that *never occurred*, because of a ten days shift in calendars.

2.2 From calendars to architectural changes?

Calendars are an interesting topic by themselves, but what is the connection with architectural changes? Why should we “start from the globe” – time in calendars – when what we are interested in is analysing the evolution of pieces of architecture?

The first reason is a very simple one: there is a clear relation of calendars to architecture. A well known example is the iterative reconstruction of the Temple in Ise (Japan), every 20 years – and 20 years is precisely the Japanese calendar’s main cycle (along with eras, connected with the country’s rulers). Another simple example is the orientation of Christian churches and chapels in the medieval period. These edifices are usually dedicated to a given saint, and when possible oriented with the apse towards the East at sunrise on the Saint’s day, a clear dependence of architecture to the Christian calendar.

The second reason why we chose to “start from the globe” results from an analysis of our previous chronology assessment contributions (Dudek and Blaise 2008, 349-357), (Dudek and Blaise 2011, 632-641), (Blaise and Dudek 2010, 91-100). We have introduced in the past years a number of visualisations aimed at facilitating analytical reasoning. Yet these visualisations rely on a single and common modelling choice, namely - using Aigner’s terminology – a one-year chronon time granularity.

And, although we did investigate ordinal time rather than discrete time (Dudek and Blaise 2011, 632-641), branching time rather than ordered time (Blaise and Dudek 2011, 8) we are far from having investigated in a thorough way the temporal info we have about architectural changes. And this relative thinness is neither a choice nor an accident. It is the direct consequence of how time is viewed in the

context of historical data sets where it is basically seen as a dating issue. Once you have said “circa 1st half of the XIVth century” you’re done with the temporal aspect. And so we believe that, if we are to further investigate the time parameter we need to handle historic data sets that go beyond dating – and calendars are well suited to that need.

We expect that developing better tools to model and visualise temporal aspects of historic calendars will help us to re-read our data sets and, for instance, to uncover similarities and differences, constructive innovations, etc..

2.3 Issue and content

What does it take to analyse similarities and differences between calendars? It naturally requires a modelling effort, through which common features can be pointed out. Poincaré wrote ... it is in the relations alone that objectivity must be sought, it would be vain to seek it in beings considered as isolated from one another... (Poincaré 1902).

In a way this contribution’s main issue could be seen as finding a mean to compare in an insight gaining manner historic calendars. Because calendars were designed that way, modelling will be done by extension – with a subset of 25 calendars chosen (Section 3.1). Furthermore, we will need to integrate confidence assessments in cases where the data is questionable. Given a predefined set of descriptors for each calendar, we will need to develop visualisations in order to sum-up visually a specific calendar’s descriptors, or to compare it to others (Section 3.2). Finally, we will need to check out whether or not the proposed framework helps uncovering parenthoods (preferably unexpected – otherwise the whole effort might have been vain).

An exploratory transfer of the calendar framework to the handling of multiple time granularities on architectural cases will be proposed in section 4. Remaining implementation challenges and perspectives of this research will be discussed in section 5.

One point has to be made clear here: this research is not about the parameter time in general, but about how time is handled inside calendars – i.e. discrete time, with a one-day chronon.

Result expected is a set of visualisations helping us to analyse where and when historic calendars have been used, in what they compare to one another. Finally, computational platforms helping users to translate 21st April 2023 (Gregorian calendar) into Chinese, Muslim, Hebrew or other calendars already exist, this aspect will therefore not be mentioned (see for example (Cultural-China 2010), (Philosophia Islamica 2009), (Hebcal Jewish Calendar 2012)).

3. Modelling and Visualising Historic Calendars

As a first step, we initially focused on visual comparison of 25 alternative calendars covering a wide range of historic periods and cultures or civilisations.

(Japanese (*Taiintaiyoreki*), Tibetan, Babylonian, Burmese, Chinese, Hebrew, Pre-Islamic (*Dzahilija*), Symmetry⁴⁵⁴, Incas (solar), French republican, Egyptian, Soviet, Coptic, Julian, Gregorian, Byzantine, Roman (republican), Inuit, Continental Celtic (*Coligny*), Muslim calendar (*Hijri*), Attic state calendar, Maya and Aztec calendars (*Xiuhpohualli*, *Tonalpohualli*), Badí’, Incas (lunar) and Balinese Pawukon calendars).

3.1 Modelling

At first glance a calendar is made of common notions (e.g. day, season, year) that need to be adjusted depending on the culture (day finishes with the sunset in the Jewish calendar, there are nine seasons in the Inuit Calendar, year started in November in Gaulish Coligny calendar).

In fact there also are less obvious features that need to be taken into consideration if we are to cross-examine calendars:

- What is the repetitive phenomenon the calendar bases on (sun / moon / seasons,)?
- How are conflicts between natural phenomena (intercalation) solved?
- What granules, are used (*i.e.* weeks, fortnights, months)?

YEAR DISCRETISATION	
chronon (τ)	the smallest temporal unit of a calendar
granule (level 1)	first level granule ($\lambda_1 = n\tau$) (equivalent of a week)
granule (level 2)	second level granule ($\lambda_2 = m\tau$) (equivalent of a month)
granule (level 3); leap-year sensible	calendars' basic granule ($\lambda_3 = y\tau$) (equivalent of a year)
	calendars' basic granule – relation with first level granule ($\lambda_3 = s \lambda_1$)
	calendars' basic granule – relation with second level granule ($\lambda_3 = x \lambda_2$)
	order of second level elements
intercalation mechanism	insertion of a leap granule into a year in order to bring an average year's duration closer to a length of the solar year
seasons (varying granules)	a number (integer)
beginning of the year	a given moment (chronon or granule) inside a calendar, or an astronomic phenomenon
beginning of the day	a given time point within the day/night cycle
ADDITIONAL PARAMETERS	
beginning of time countdown	originating position of temporal scale (ex. ab Urbe condita - traditionally dated to 754BC)
cycles present in calendar	types and lengths of cycles
zone of usage	geographical area
period of validity	a time interval, in history, during which a given calendar was in use
calendar type	phenomenon or calendar purpose that defines the length of a common year

Table 1. An illustration of the variety of the information / notions needed to depict and compare calendars (in terms of time span, in terms of actual formatting) – the list proposed here is not comprehensive.

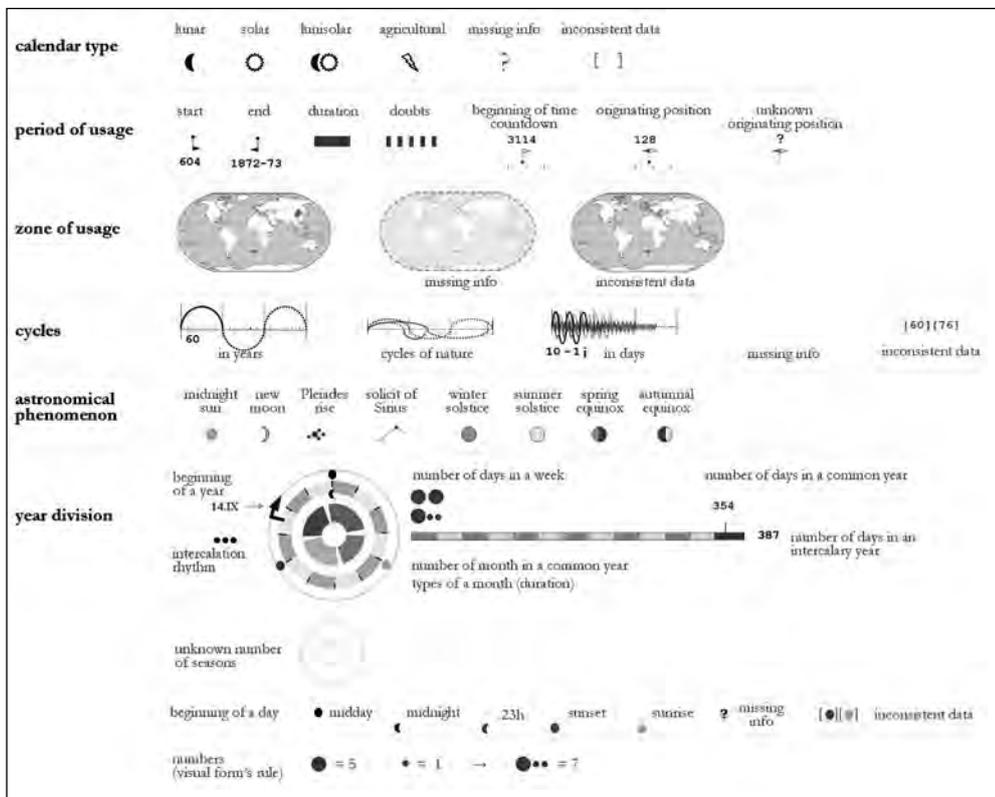


Figure 1. Visual encoding of the key descriptors taken into consideration in the comparison of calendars.

- What significant cycles are integrated in the calendar?

The conceptual model we propose combines several layers of information. Some examples are given in Table 1.

In the following section we present how these notions are actually transferred into graphics (visual cues used), and evaluated on real cases. It has to be made clear that initially we focused on the graphics – we needed to check out whether the whole idea of starting from the globe could make sense. So the implementation we propose is dynamic but poorly structured – script interpretation. Among perspectives of this research is to develop a robust OO-based structure to represent the notions we have identified here above.

3.2 Visualising

The first visualisation we propose sums up in a synthetic way fifteen key descriptors of calendars (Fig. 1) distributed inside six visual components (Fig. 2).

The most widespread calendar system today is the Gregorian calendar (solar) introduced in by Pope Gregory XIII in a papal bull signed on 24 February 1582. It was adopted in turn by different countries. Gregorian calendar modified the Julian calendar’s regular cycle of leap years (leap years are exactly divisible by four except for years that are exactly divisible by 100, but the centurial years that are exactly divisible by 400 are still leap years). A Week is the smallest granule (7 days). Each year counts 12 months (365 days in common years, 366 days in leap years). Beginning of a year was maintained on 1st January.

Until the adoption of a resolution of the International Meridian Conference, each town used its own local time. Lengths of each day-time hour and night-time hour were unequal and depended on the geographical latitude and varied according to the seasons - what is more a day could begin at noon, sunrise or sunset, reflecting the wide diversity of customs for defining and counting the hours across the European continent (Fig. 3).

Influence of a culture and climate is plainly

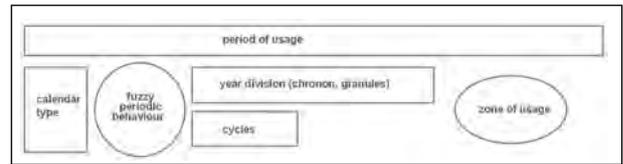


Figure 2. Distribution of the fifteen key descriptors in graphic components.

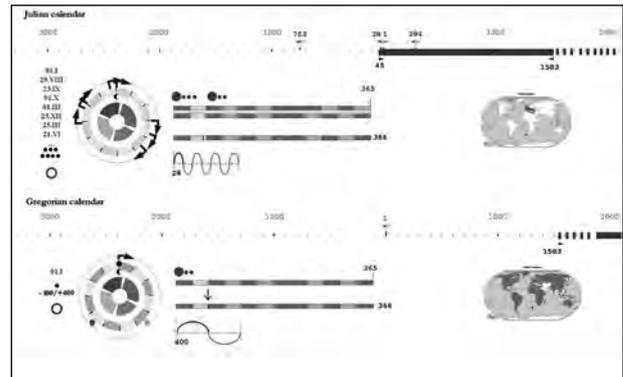


Figure 3. Visual comparison of the Julian and Gregorian calendars.

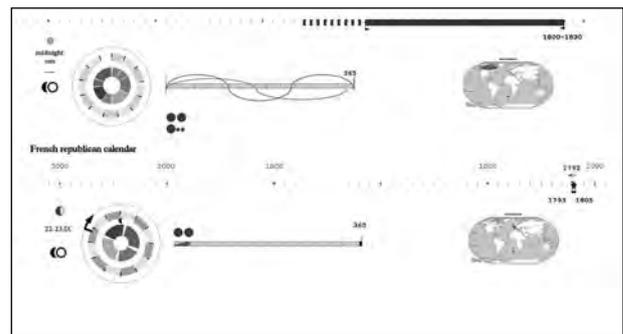


Figure 4. Visual comparison of the Inuit and French republican calendars.

visible in Inuit calendar. The absence of a notion of day or week in Inuit’s culture, as well as surprising number of seasons (related to the cycles of nature). On the other hand the French republican calendar has a clear “goddess reason” decimal structure (Fig. 4).

In order to foster comparison of the whole collection (Fig. 5), we developed a second visualisation underlining:

- legacies in between periods and areas (e.g. The Coptic Year is the extension of the ancient Egyptian civil year retaining its subdivision into the three seasons.),

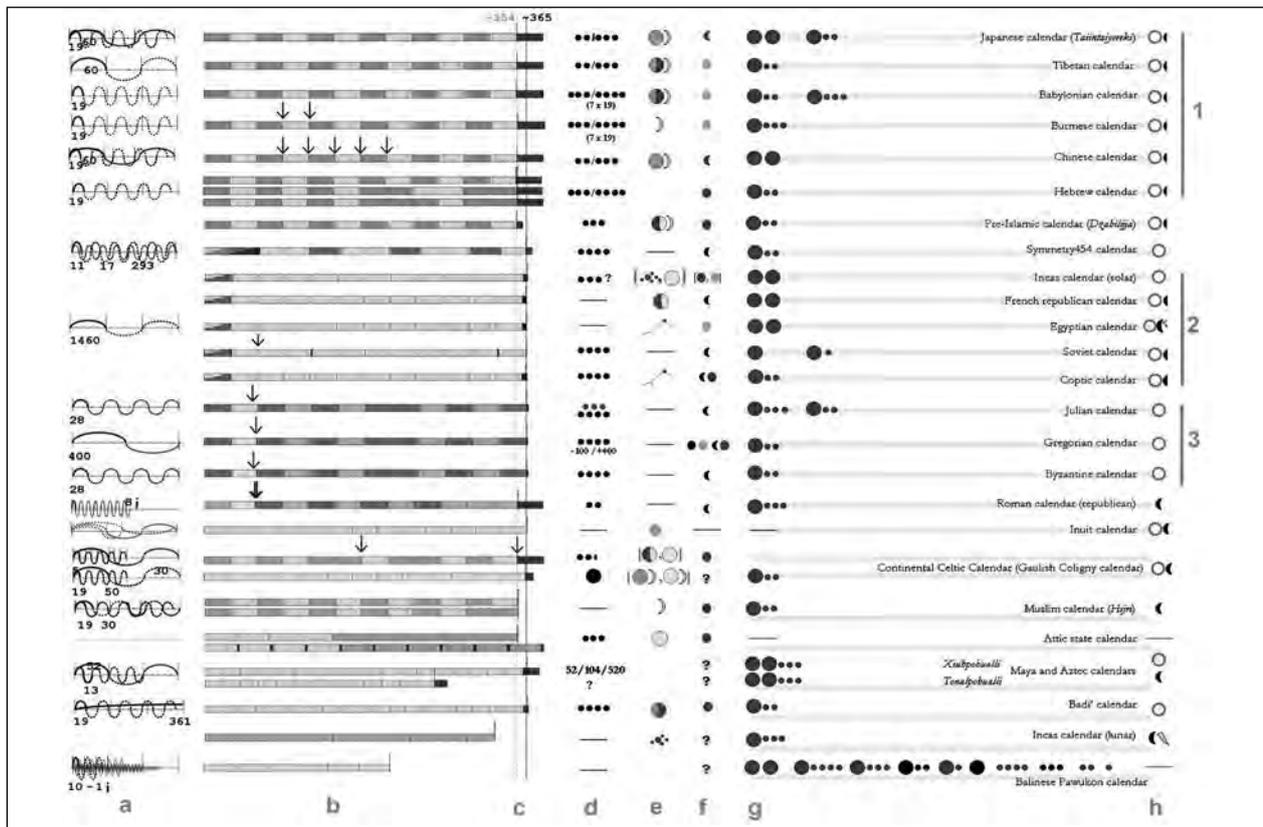


Figure 5. A visual comparison of 25 calendars. (a) natural or anthropogenic cycles present in a calendar; b) year discretisation for a common and a leap year; c) time gap between the lunar and solar years; d) intercalation rhythm; e) natural phenomenon established as beginning for the new year; f) beginning of the day; g) number of days in a week; h) calendar type, 1-3) three groups of calendars with similar patterns (year discretisation in particular).

- presence of cycles, alternative divisions of the year (e.g. Agriculture related calendars - Balinese Pawukon calendar or Incas agricultural calendar - seem quite weird compared to a calendars based on a sidereal year.),
- mechanisms to cope with intercalations, etc.

This comparison allowed us to ascertain similar mechanisms of time discretisation - day plays role of chronon in an overwhelming majority of cases and it is aggregated into widely used, standard granules (i.e. week, month, season, year, century) or some less employed ones (e.g. days outside a year). However the exact number of days in a given granule substantially differs (e.g. 12 months in a common year is not a rule but only a dominant trend).

Moreover beginning of time countdown and originating position of temporal scale strongly

varies. These parameters as well as intercalation mechanisms, beginning of a year or a day are culture dependent – showing us more about societies and their way of life than about the time.

A very common and instructive parameter is periodicity of calendars - presence of natural or calendar based cycles. Once again, cultural based cycles (e.g. ruler's lifetime, market day's cycle, unlucky day's cycle, intervals associated with different groups of deities, ...) throws light on humans and societies, zones of influences helping us in teleological reasoning, etc..

Natural cycles (climate or astronomical related) although profoundly built in calendar structure are independent from the calendar structure – even in most precise calendars they move slightly inside a sequence of a year. This relative freedom of natural cycles is reinforced by a recurrent phenomenon:

a discontinuity of time in calendars due to human manipulation - time in calendars is stretched or shortened from time to time (e.g. 80 additional days in year 46th BC, 11 days less in 1582 AD a.s. in Madrid and Cordoba, but not in Stratford-upon-Avon or in Southampton).

This first result has been extended to propose a more generic framework for visualising time with multiple granularities. It is applied on two different test cases: chapel of St Anne in Southern Alps (France) and the belfry of Cracow's former town hall in (Poland).

4. Visualising Time-oriented Historical Data

Naturally the visualisation of time-oriented data poses various types of difficulties, with a number of them generating diverse types of uncertainties:

- modelling choices of temporal variables,
- data credibility (*i.e.* heuristic accuracy and bias of analyst),
- inconsistency of data - that implies possibility of various scenarios,
- incompleteness of data,
- subjectivity of choices (amount of interpretation or personal judgment included),
- approximations and imprecision in data description (e.g. How to interpret expressions like *probably in 1567?*),

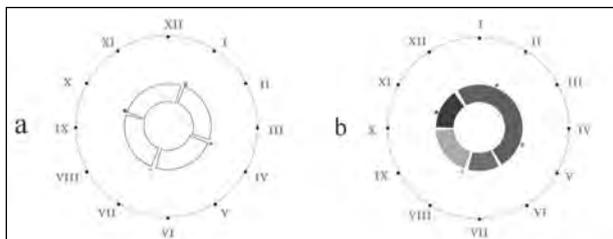


Figure 6. Cyclic time visualisation mode exploits a 'clock' metaphor – one o'clock stands for beginning of January. Beginning of astronomical seasons is marked by colour dots and meteorological seasons are represented by coloured sectors ('rose of seasons') – they may not coincide (b).

- temporal resolution of data – that relates to temporal granularity, etc.

All these elements should be integrated into any visualisations to improve the cognitive task of spatio-temporal understanding. In this paper we will limit our discussion to the last point that relates to time granularity.

We will start with a brief introduction into problems posed by visualisation of datasets characterised by varying temporal granularity. In a following section we will present a proposal framework of a method of visualising time with multiple granularities.

4.1 Visualisation of datasets characterised by varying temporal granularity

Historical data sets are not consistent in terms of temporal granularity. In other words the individual temporal resolution of pieces data may vary.

Let's take as an example three different pieces of information concerning the same object:

- A fire in the town hall's tower took place in summer 1543.
- A modification of a tower started on 15th August 1543.
- A big quantity of bricks have been bought in August 1543.

Each piece of information has its own temporal granularity. If the chronon of our visualisation (the smallest unit) is one civil year (date format YYYY), information about the events that occurred in the same year can be represented only as a group of events that took place in 1543. What is more other hints about temporal relations between these events (*i.e.* "in summer" contains "in August" contains "15th August") will disappear.

If our choice for the chronon is one day (date format DD MM YYYY), granularity of the visualisation will exceed the natural temporal granularity of the underlying information. In such situations the description and visualisation of events

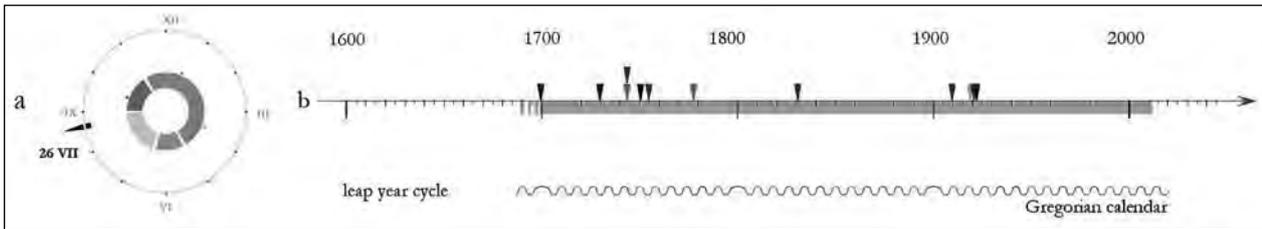


Figure 7. St Anne chapel a) cyclic time visualisation mode; b) timeline - Different colours mark different types of events. The overall evolution of the chapel is marked with the brownish line. All known transformations were described according to the Gregorian calendar.

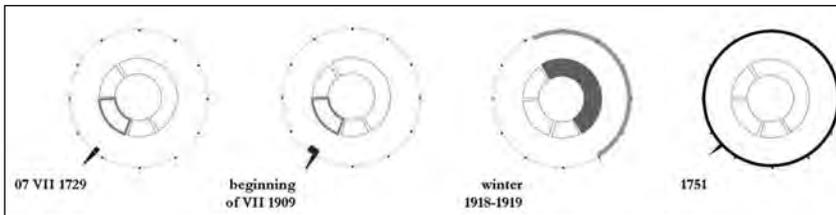


Figure 8. St Anne chapel - Events dating represented with their natural, heterogeneous; temporal granularity; From left to right : DD/MM/YYYY format, MM/YYYY format, season + YYYY format, and YYYY format.

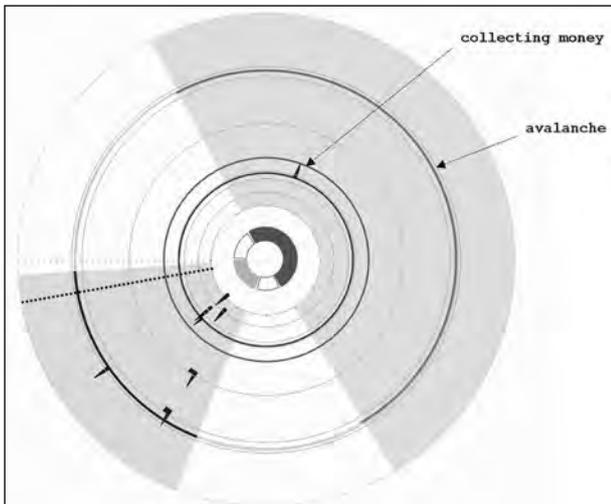


Figure 9. Visualisation produced for a whole life of the artefact. Note that all the events related to the chapel took place in summer (i.e. before st Anne's day) except of a collect of the money and avalanche – a fact very easy to consider as obvious (altitude 2400m) once the visualisation did the job for us...

compels to introduce approximated, interpretation-derived values, thereby reinforcing the global uncertainty on temporal aspects (i.e. "in summer" does no mean each and every summer day).

In other words, whatever chronon we choose, and represent, there will be pieces of information that just don't fit – either they are too "thin" and the foot moves in the shoe, or too big and the foot is shrunk to fit in the shoe.

Accordingly, and taking advantage of what we had observed on the variable granularity of calendars and on the importance of fuzzy periodic comportment of seasons, we have tried to support time-oriented reasoning tasks with mechanisms that would display in a unique visual layout several granularities at a glance. The visual formalism, first experimented on calendars and seasons, provides a simple solution to handle both a day granularity and more fuzzy intervals within a cyclic-time enhancing visualisation (Fig 6).

It is in fact a rather generic formalism (it could be applied to year and centuries, or to any other combination) that bases on two principles:

- concentricity on one hand - mechanism to segregate granularities,
- sector vs. point on the other hand - mechanism to allow switches from chronon to granules, including fuzzy granules.

4.2 Impact on representing architecture changes

The first case study is Ceillac's chapel of St Anne, in Southern Alps (altitude 2400 m).

The cyclic events related to this artefact include an annual pilgrimage to the chapel (July 26th) and naturally the cycle of seasons (Fig 7a). At this altitude meteorological seasons do not coincide with astronomical seasons. Winter is the longest

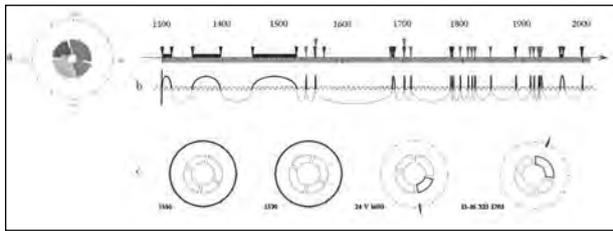


Figure 10. The belfry of Cracow's former town hall. The classic timeline indicates intervals of transformations (green colour) and fires of the artefact (orange colour). a) average meteorological seasons do not coincide with the astronomical seasons, summer is the longest season in Cracow, no cyclic event for this artefact has been found; b) rhythm of artefact transformations/ rhythm of a leap year; c) dates of known fires of the town hall tower with their specific temporal resolution.

season here but its length may vary (season's length has a fuzzy periodic behaviour), therefore it is important to point out that in our cyclic events representation we show values for an average season length – variations of this factor has to be taken into consideration by analysts.

The information about the chapel of St Anne was gathered and represented using a classic timeline with a granularity of one year (Fig 7b), but it is not the natural temporal resolution of all the data we dispose. Pieces of information we dispose are dated with a variable precision (e.g. DD/MM/YYYY, MM/YYYY, season YYYY, YYYY). Increasing the time granularity (e.g. operating with a day) makes a visualisation impractically long, what is more it introduces the unavoidable problem “shrinking the foot to fit in the shoe” with interpretation-derived values.

Our proposal is to use at visualisation time the natural temporal granularity of the dating associated with events (Fig 8).

This type of visualisations may be reorganised according to additional criterion (e.g. by events type) in order to help in reasoning. It is also possible to produce them for a selected time span (e.g. two decades, a century or a whole life of an artefact (Fig 9)).

Linear and cyclic visualisation modes underline different particularities of the data sets: it should be stated clearly that their combination

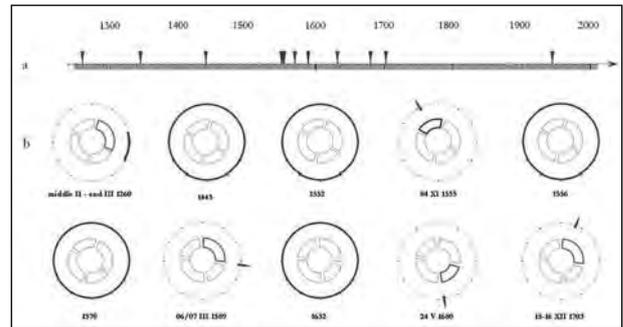


Figure 11a) A timeline listing all known fires of the ensemble of buildings that furnished the entire Main Market Square in Cracow – a fire in town is a punctual phenomenon. b) fires with their specific temporal resolution.

notably reinforces their power of assistance in reasoning. What this simple experiment shows is not that one is better than the other – the basic linear timeline is for instance very useful in underlining temporal densities – but digging into that temporal aspects requires to think outside the square in terms of visualisation.

The second case study is the belfry of Cracow's former town hall, and more precisely its fires.

It is not trivial to note that two case studies - chapel of St Anne and the belfry of Cracow's former town hall – have not the same location, therefore time measure (different calendars) and climatic conditions are different (Fig 10a). The lifetime of the belfry traverses two calendars and it had some relation with the seasons that have been moving inside year cycle (we present only a rough approximation of the changes).

Although we managed to visualise all dates of known fires of the artefact with their specific temporal resolution, it is hard to find any regularity basing our reasoning on four cases only. One could be tempted to say that : *the oldest historical sources are less precise in terms of dating.*

We have therefore verified this hypothesis on the ensemble of buildings that furnished the entire Main Market Square in Cracow (Fig 11) and it turned out to be false. Unsurprisingly, we notice no regularity, no cyclic comportment inside these datasets.

5. Conclusions

This contribution presents bases of a method we have tried to develop in order to visualise time with multiple granularities, in order to support reasoning on various aspects of temporal relations (cyclic or fuzzy periodic compartments) in the context of heterogeneous temporal data sets. In short, this experience was primarily about “shaking the tree” of possibilities when one needs to handle multiple granularity.

At the current stage of development we wish to restrict our conclusions to the following remarks:

- properly designed visualisation helps in reasoning (visual comparisons, visual thinking),
- each historical calendar has its own specific granularity,
- historical calendars are subject to change over time, including in terms of granularity,
- the day/night notion could appear as the smallest common time interval (‘universal chronon’) within reach when wanting to compare calendars – however exceptions here exist (e.g. in Inuit’s calendar that notion is absent),
- study of cyclic and fuzzy periodic behaviour may help in better understanding data sets we handle in the context of historic sciences, but periodic behaviour analysis requires specific visual tools (beyond basic linear timelines),
- depending on where changes over time occur, the analysis may require different visual instruments - ‘tailored’ to the local conditions (in particular to the succession of calendars with various - sometimes local - time discontinuities, cycles of seasons, time granulation, etc.),
- if we deal with the notion of time in our research, it is worth to take some time to think about “time” for itself, and not about time as a date stamp we need to put on a given piece of data.

It has to be said clearly that further work, especially in terms of implementation, is necessary if we wish to further investigate potential benefits and

drawbacks of this approach. Yet there is one lesson that probably deserves to be drawn in conclusion: although most often a risky bet, starting from the globe – what we did by starting from calendars in order to depict temporal aspects of architectural changes - may be a fruitful attitude sometimes, in particular when talking about things we think we know well. Talking about time, Saint Augustin said: “*If no one asks me [what is time], I know. But if I wanted to explain it to one who asks me, I plainly do not know*”. Those who investigate the time parameter probably tend to think he was not that wrong – the issue remains today a challenging one, particularly in the context of historic sciences.

References

- Aigner, W., S. Miksch, W. Müller, H. Schumann, and C. Tominski. 2008. “Visual methods for analysing time-oriented data.” *IEEE Transactions on Visualization and Computer Graphics* 14/1: 47-60.
- Allen, J.F. 1984. “Towards a general theory of action and time.” *Artificial Intelligence* 23: 124-154.
- Blaise, J.-Y., and I. Dudek. 2010. “Understanding changes in historic architecture. Can we provide tools & methods for visual reasoning?” In *Proceedings of International Conference on Information Visualization Theory and Applications*, edited by P. Richard and J. Braz, 91-100. Angers: INSTICC Press.
- Blaise, J.-Y., and I. Dudek. 2011. “Visualizing alternative scenarios of evolution in heritage architecture.” In *Proceedings of the 11th International Conference on Knowledge Management and Knowledge Technologies*, edited by S. N. Lindstaedt, M. Granitzer, 45-58. New York: ACM International Conference Proceeding Series ACM.
- Chardonnel, S. 2007. “Time-Geography: individuals in time and space.” In *Models in Spatial Analysis*, edited by Lena Sanders, 97-126. London: ISTE.
- cultural-china. 2010. “Gregorian Calendar- Chinese Lunar Calendar Converter.” Accessed February 2012. http://www.cultural-china.com/Kaleidoscope/cultural_tool/Calendar_Converter/index.html.
- Dudek, I., and J.-Y. Blaise. 2011. “Concentric Time: Enabling Context + Focus Visual Analysis of Architectural Changes.” In *Foundations of Intelligent Systems*, edited

Visualising Time with Multiple Granularities: a Generic Framework
Iwona Dudek and Jean-Yves Blaise

- by M. Kryszkiewicz, H. Rybinski, A. Skowron, and W. Raś, 632-641. Berlin, Heidelberg: Springer-Verlag.
- Dudek, I., and J.-Y. Blaise. 2008. "Profiling artefact changes: a methodological proposal for the classification and visualisation of architectural transformations." In *Proceedings of VSMM 2008*, edited by M. Ioannides, A. Addison, A. Georgopoulos, and L. Kalisperis, 349-357. Budapest: Archeolingua.
- Hebcal Jewish Calendar. 2012. "Hebcal Jewish Calendar." Last modified April 02 2012. <http://www.hebcal.com/converter/>.
- Knight, B., and J. Ma. 1993. "Time representation: A taxonomy of temporal models." *Artificial Intelligence Review* 7/6: 401-419.
- MacEachren, A. M., M. Gahegan, and E. Hetzler. 2005. "Visualizing geospatial information uncertainty: What we know and what we need to know." *Cartography and Geographic Information Science* 32/2: 139-160.
- Matoušek, K., M. Falc, and Z. Kouba. 2007. "Extending Temporal Ontology With Uncertain Historical Time." *Computing and Informatics* 26: 239-254.
- Philosophia Islamica. 2009. "Conversion of Hijri A.H. (Islamic) and A.D. Christian (Gregorian) dates." Last modified October 13 2009. <http://www.muslimphilosophy.com/ip/hijri.htm>.
- Peralta, D.N., H. S. Pinto, and N. J. Mamede. 2004. "Reusing a Time Ontology." In *Enterprise Information Systems V*, edited by O. Camp, J. B. L. Filipe, S. Hammoudi, and M. Piattini, 241-248. Dordrecht: Springer Netherlands.
- Poincaré, H. 1902. *La Valeur de la Science*, Accessed 28 November 2011. http://fr.wikisource.org/wiki/La_Valeur_de_la_Science.
- Tufte, E. R. 1990. *Envisioning Information*. Cheshire: Graphics Press.
- Tufte, E. R. 1997. *Visual Explanation*. Cheshire: Graphics Press.
- UNESCO. 2011. "World Book and Copyright Day." Accessed February 19 2012. <http://www.unesco.org/new/en/unesco/events/prizes-and-celebrations/celebrations/international-days/world-book-and-copyright-day-2012>.
- West, K. 2010. "Theory vs. practice." *The News & Advance*, September 27. Accessed May 2012. <http://www2.newsadvance.com/lifestyles/2010/sep/27/theory-vs-practice-ar-528499/>.