



UNIONE EUROPEA
Fondo Sociale Europeo



Corso di Dottorato di ricerca
in Lingue, Culture e Civiltà e Scienze del Linguaggio
ciclo 37°

Tesi di Ricerca

CHIt: a Web Application for the Valorization of the Italian Cultural Heritage with a Focus on Germanic Peoples in Italy

SSD: L-FIL-LET/15

Coordinatore del Dottorato

ch. prof. Anna Cardinaletti

Supervisore

ch. prof. Marina Buzzoni

Dottorando

Giulia Fabbris

Matricola 849291

La borsa di dottorato è stata cofinanziata con risorse del
Programma Operativo Nazionale Ricerca e Innovazione 2014-2020, risorse FSE REACT-EU
Azione IV.4 "Dottorati e contratti di ricerca su tematiche dell'innovazione"
e Azione IV.5 "Dottorati su tematiche Green"

Acknowledgments

First and foremost, I wish to thank my tutor, Professor Marina Buzzoni, for her tireless support throughout these three years (and all the years before). Her immense knowledge and passion for Germanic and Digital Philology have been truly inspirational. Thank you for guiding me with the utmost competence and kindness.

I also thank Chiara De Bastiani, with whom I have somehow shared this path. She was always available when I needed her support, and not only from a scholarly point of view. Thank you for supporting and helping me during the numerous difficult times.

I am extremely grateful to Net7, in the person of Chiara Aiola, who patiently and competently trained me in the previously unknown field of programming. Thanks for repeating a hundred times what an API was! And another special thanks to Alessandro Bertozzi for taking care of (and patiently enduring) all my requests for perfecting the app.

A special thank you goes to Giosuè, who has supported me during these three years and always listened to my complaints.

The sweetest thank you goes to my daughter Noemi, who arrived almost at the end of this journey but has filled the last and most challenging steps with endless joy.

I thank my mum, my hero, my rock. Life has been bitter with you, but you are stronger. I have never stopped in front of any difficulty: who do you think inspired me?

My siblings deserve a special thank you because they are the three fires that never stop warming my heart – even if sometimes they burn! You are the greatest gift, together with my baby girl.

Last but not least, I thank my friends Totò and Serena, the two lighthouses of my life. I know you will always be there.

Table of contents

Acknowledgments	2
List of abbreviations	5
Introduction.....	7
SECTION I: THE FRAMEWORKS OF THE PROJECT	11
1. Digital transition and adaptation for the valorization of the Cultural Heritage	12
1.1. The context of the research: Europe and Italy towards a digital transition	15
1.2. The importance of preserving digital and digitized objects	22
2. Methodological background and contextual definitions	28
2.1. FAIR principles	29
2.1.1. Keyword: <i>reuse</i>	31
2.2. Online collections: digital libraries, virtual museums, digital ecosystems and miscellaneous repositories.....	35
2.3. State of the art	37
2.4. Why a cultural aggregator.....	44
2.5. Why a web application.....	46
SECTION II: THE INCEPTION AND DEVELOPMENT OF THE APPLICATION	53
3. CHIt: a tool for the final user.....	54
3.1. The added value of being <i>findable</i>	54
3.1.1. Exposition and importance of APIs.....	58
3.1.2. The transformation within the digital paradigm through APIs	59
3.2. Components	61
3.2.1. Muruca's architecture: the backend	63
3.2.2. Muruca's architecture: the middleware.....	65
3.2.3. Muruca's architecture: the frontend.....	67
3.2.3.1. Collections.....	69
3.2.4. Pundit: the annotation tool.....	71
3.2.4.1. Collaboration, enrichment and personalization of data.....	75
3.2.5. EVT: the visualization tool	77
4. Structure and implementation of the middleware.....	80
4.1. Problems and possible solutions.....	81
4.2. Implementation of the source code.....	86
4.2.1. Importing data: the providers	89

4.2.1.1.	Europeana.....	94
4.2.1.2.	Zenodo	102
4.2.1.3.	GitHub.....	107
4.2.1.4.	Germanic cultural heritage in Italy and OntoVE.....	110
4.2.2.	Creating records	118
4.3.	Connection to the annotation and visualization tools.....	120
5.	Content organization, visualization and testing	123
5.1.	The frontend.....	124
5.2.	Testing phase	132
SECTION III: USING THE APPLICATION: CASE STUDIES APPLIED TO THE GERMANIC PHILOLOGY FIELD.....		137
6.	Application of CHIt to the Germanic cultural heritage: three case studies	138
6.1.	The Germanic cultural heritage in Italy: available digital resources	140
6.2.	Some use cases with CHIt.....	152
6.2.1.	The student.....	152
6.2.2.	The tourist	163
6.2.3.	The philologist	172
6.3.	Some preliminary conclusions	184
7.	Future developments	186
Conclusions.....		189
Bibliography.....		191

List of abbreviations

API = application programming interface

CHIt = Cultural Heritage of Italy

CoE = Council of Europe

DCMMI = Dublin Core Metadata Initiative

DESI = digital economy and society index

EC = European Commission

EOSC = European Open Science Cloud

EP = European Parliament

ERC = European Research Council

ESF = European Social Fund

EU = European Union

EUCO = council of the European Union

EVT = edition visualization technology

FAIR = findable, accessible, interoperable, reusable

GLAM = galleries, libraries, archives, museums

ICOM = International Council of Museums

JSON = JavaScript Object Notation

LOD = Linked Open Data

MiC = Italian Ministry of Culture

MUR = (Italian) Ministry of University and Research

OED = Oxford English Dictionary

PND = National Digitalization Plan

PNRR = National Plan for Recovery and Resilience

PON = National Operational Program

REACT-EU = Recovery assistance for cohesion and the territories of Europe

REST = REpresentational State Transfer

UI = user interface

UNESCO = United Nations Educational, Scientific and Cultural Organization

UX = user experience

Introduction

In the last decades, the awareness concerning the importance of the preservation of the world cultural heritage and its connection to the social sphere has considerably increased. The *Faro Convention* (2005), established by the Council of Europe (CoE), is a clear sign that goes in this direction, since through it the Parties agreed to:

- a. recognise that rights relating to cultural heritage are inherent in the right to participate in cultural life, as defined in the Universal Declaration of Human Rights;
- b. recognise individual and collective responsibility towards cultural heritage;
- c. emphasise that the conservation of cultural heritage and its sustainable use have human development and quality of life as their goal;
- d. take the necessary steps to apply the provisions of this Convention concerning:
 - the role of cultural heritage in the construction of a peaceful and democratic society, and in the processes of sustainable development and the promotion of cultural diversity;
 - greater synergy of competencies among all the public, institutional and private actors concerned (Chikobava 2023, 2).

These salient points clearly suggest a more prominent and active participation of citizens into the preservation, valorization and dissemination of the cultural heritage. Article 27 comma 1 of the *Universal Declaration of Human Rights* recites that “[e]veryone has the right freely to participate in the cultural life of the community, to enjoy the arts and to

share in scientific advancement and its benefits.”¹ The Faro Convention clearly recalls the notion of “community” and delivers to it – e.g., to the citizens – the right and the duty to engage with the cultural heritage.²

The Council of the European Union (EUCO) goes further on in 2014 when it concludes that cultural heritage is a strategic resource for a sustainable Europe and its management a strategic choice for the 21st century.³ The recognition of its economic impact on society and its importance in enhancing the social capital has led to a series of actions adopted by the Member States that aim, among other things, at its promotion and at the cooperation between private and public actors.

Moreover, after the pandemic, the *Recovery assistance for cohesion and the territories of Europe* (REACT-EU)⁴ program was created in order to sustain the European Countries and help them face the social and economic damages caused by the COVID-19. The main operational strategy is the funding of projects that have the goal to boost the economy by creating job opportunities and improve working conditions, with particular attention to the youth and healthcare. In this specific direction go the *European Social Fund* (ESF)⁵ additional resources, in which context this work is developed.

Last but not least, the domain of Digital Humanities plays a major role in the realization of this project and its intersection between different research fields is in line with the aforementioned principles (cooperation, social and economic enhancement, synergy). Ciotti (2018) asserts that the disciplinary fluid definition of this domain could be one of the reasons of its success worldwide, both in the theoretical and practical research. There is no attempt here to give a definition of this discipline, since it would be too great

¹ <https://www.un.org/en/about-us/universal-declaration-of-human-rights#:~:text=Article%2027,scientific%20advancement%20and%20its%20benefits>.

² In his article, Zagato (2015) investigates the innovative notions that the Convention contains, with a particular focus on the concept of “heritage community.”

³ EUCO (2014)

⁴ <https://www.europarl.europa.eu/factsheets/en/sheet/215/recovery-assistance-for-cohesion-and-the-territories-of-europe-react-eu>

⁵ <https://ec.europa.eu/esf/home.jsp?langId=en>

an ambition given its multifaceted nature.⁶ Nevertheless, the following chapters present several aspects fundamental for the understanding of Digital Humanities. Moreover, according to Braake et al. (2016), an interesting aspect that the intersection between the digital and the humanities offers is a change of perspective: in fact, the advantages of machine-driven data (see also section 2.1.) and, more in general, of digital tools, allow for a shift from the top-down approach to a bottom-up one. Under this perspective, larger communities of users can be actively involved in the processes of creation, dissemination and preservation of the cultural heritage. This vision is in line with the abovementioned principles emphasized in the *Faro Convention*.

The original idea for this project was the creation of two client applications: one for the visualization of resources and another for their annotation. One of the keywords has been, since the beginning, *reuse*, both of digital objects (e.g., data and metadata) and of tools. Therefore, after careful reflection, it was opted for the development of one single client application instead with two already existing tools integrated, which fulfill the two aforementioned actions. The resulting application has been named *Cultural Heritage of Italy*, or CHIt.

Chapter 1 presents a general overview of the context in which this project was developed.

The second chapter goes a step further and begins to tackle with some digital aspects of the project by presenting the principles that should be considered in every project in the Digital Humanities, in particular the concept of *reuse*, offering some preliminary examples and how the choice of this web application was born.

Chapter 3 presents in detail all the components of the application and the external tools integrated for the visualization and annotation of the retrieved resources.

Chapter 4 is the heart of the work, since it explains the implementation of the architecture of the app and the design of the code throughout all its phases. In other words, the development of the whole app with a focus on the middleware is presented,

⁶ Callaway et al. (2020) in their survey collect more than 300 definitions of *Digital Humanities*.

with a particular accent on the retrieval of resources and the standardization of metadata.

Chapter 5 is divided into two independent sections, but both dealing with the organization of the content. The first one offers an overview of the app, namely of its frontend, how it is structured and all the pages that are available. The other section presents the testing phase, which proved quite successful since the beginning.

Chapter 6 is the playground for the practical testing of the functionalities and usability of CHIIt. In fact, three user cases are presented, according to different user needs: a student, a group of tourists and a scholar. It will be seen how the app can be adapted to different types of audience.

In chapter 7 some of the possible improvements and further developments are put forth. These are desirable in order to make the application more functional and complete.

SECTION I: THE FRAMEWORKS OF THE PROJECT

1. Digital transition and adaptation for the valorization of the Cultural Heritage

The UNESCO General Assembly, in a notable decision on December 4th, 2001,⁷ declared the year 2002 as the *United Nations Year for Cultural Heritage*. Less than two years later, Member States and other entities in the UN System were called upon to develop policies related to both tangible and intangible cultural heritage.⁸ This initiative underscored the global commitment to recognizing, preserving, and celebrating the world's diverse cultural legacies. The subsequent period saw a collective effort to establish policies aimed at safeguarding cultural heritage on a comprehensive international scale.

In the premise of this work, it would be tempting to offer a definition of cultural heritage. However, this has ever been a hard task since it “has not always designated the same thing and, in recent decades, the notion has changed considerably in order to express more aptly a global and holistic approach which alone can testify to the universal nature of the human spirit in all its creations.”⁹ In later years the definition will be enlarged so as not to be limited to monuments and collections of objects, but will comprise also traditions and living expressions inherited by our ancestors. Again, on March 30th, 2017 a G7 was held and the Italian contribution begins with a section entitled *Living with multiple definitions of cultural heritage: the “Infinity of Lists”* and states that every treaty or convention gives its own definition of cultural heritage because there is not a common one.¹⁰

Therefore, a definition of cultural heritage is attempted here in order to create a clearer context, and it is drawn mainly from the aforementioned conclusions of the EUCO (2014). Cultural heritage can be referred to as tangible, intangible and digital resources inherited from the past and managed by public and private bodies. It represents and

⁷ UNESCO (2001)

⁸ UNESCO (2003)

⁹ UNESCO (2002)

¹⁰ G7 Italy (2017)

witnesses the past and as such is a tool for and of knowledge and transmission of values. It is non-renewable, non-replaceable and non-interchangeable and therefore it needs to be totally part of the transformations that occur in contemporary life. It is fundamental in creating and enhancing social capital because it promotes diversity and intercultural dialogue, it offers possibilities to develop skills, knowledge, creativity and innovation and is an effective tool for education at different levels. Its close relation to all people's lives makes it clear that *everyone* has the right to benefit from it and has the responsibility to respect and preserve it.

In the following subsection it will be shown the importance of the interplay between human capital and societies with cultural heritage for the European Union (EU). Professor Catoni writes in the Italian contribution of the aforementioned 2017's G7 meeting:

The increasing awareness of the crucial role that cultural heritage plays in contemporary societies as a driving force in fostering knowledge, innovation, conscious citizenship, respect for diversity, intercultural dialogue, peace-keeping, economic and social growth, is reflected in the growing attention governments, societies and international governmental and non-governmental organizations devote, on the one hand, to ensure the presence of cultural heritage in the national educational programs and, on the other hand, to promote the development of specific academic and technical training programs in the field of cultural heritage protection and conservation.¹¹

There is a second question that is crucial for this work and is more technical in a way: What constitutes the digit(al)ization of CH? To start this discussion, it is essential to offer

¹¹ G7 Italy (2017, 19)

a clarification. In contemporary debate, virtual cultural objects are commonly categorized into three distinct types: digitized, digital, and born-digital.

Firstly, the process of digitizing cultural objects involves the conversion from analog to digital formats. A tangible example of this is the transformation of a book through scanning, resulting in the creation of a digitized version.

Moving beyond mere digitization, the concept of digitalization involves a more profound transformation facilitated by the integration of digital technologies. A noteworthy example is the *Electronic Beowulf* website,¹² which transcends the replication of manuscript pages. Instead, it stands as a genuine digital edition endowed with numerous features that empower users to actively engage and interact with the text. This distinction underscores the evolving nature of cultural objects as they undergo digitalization processes.

Lastly, born-digital objects are inherently conceived within the digital paradigm from their inception. Examples include web applications or archives designed and developed in the digital realm from the outset. This distinct category emphasizes the native integration of digital elements in the creation and existence of cultural objects, representing a departure from analog origins.¹³

Over the recent years, the EU, represented by all its bodies, has been actively engaged in an array of projects and programs designed to facilitate the green and digital transition of its Member States. A substantial allocation of funding available for Italy has been strategically directed towards initiatives focused on enhancing human capital, as well as reinforcing the digital transformation of culture and tourism. This concerted effort reflects a broader commitment to fostering sustainable and technologically advanced growth within the Member States.

¹² <https://ebeowulf.uky.edu/ebeo4.0/CD/main.html>

¹³ For ease of writing, the difference between digitization/digitalization and digitized/digital will be pointed out only if necessary.

The following subsection endeavors to shed light on European and Italian selected projects and investments, providing an exploration that serves to establish a vital political and community context for the research at hand. By presenting an overview of these initiatives, the aim is to elucidate the multifaceted landscape in which this research is situated, highlighting the political and communal backdrop against which the study unfolds.

On the other hand, section 1.2. deals with the importance of preserving the cultural heritage in digital shape.

1.1.The context of the research: Europe and Italy towards a digital transition

At the eve of the pandemic, the European Commission (EC) released a communication in which the concern for a digital and green transition for and of Europe was expressed, recognizing the importance of digital solutions through the definition of three key objective: 1) technology that works for people, 2) a fair and competitive economy and 3) an open, democratic and sustainable society.¹⁴ COVID-19 somehow accelerated this need of “going digital” and the number of actions taken in this direction increased hugely. The rest of this section is dedicated to the review of some European and Italian programs and fundings that are in line with these goals.

To start broad, the first document considered is the *Digital Europe Programme* (DIGITAL)¹⁵ approved by the European Parliament (EP) and the EUCO in 2021 with the goal to “support and accelerate the digital transformation of the European economy and society, to bring its benefits to citizens, public administrations and businesses across the Union, and to improve the competitiveness of Europe in the global digital economy while contributing to bridging the digital divide across the Union and reinforcing the

¹⁴ EC (2020d)

¹⁵ EUCO and EP (2021), website: <https://digital-strategy.ec.europa.eu/en/activities/digital-programme>

Union's strategic autonomy."¹⁶ It is interesting to note that *whereas* 54 refers to open-source solutions, specifying that these should be encouraged so as to allow reuse, increase trust and secure transparency because this has a positive impact on the sustainability of the funded projects. This concept is crucial throughout the whole work proposed here and more European regulations and legislations will be considered in the following chapter when the topic of *data reuse* will be addressed.

The DIGITAL is part of a series of projects and fundings that have taken different shapes in the various Member States. In the following paragraphs, the Italian scenario is presented.

The year before the approval of the DIGITAL, therefore in 2020, the EU introduced the *NextGenerationEU* plan¹⁷ to face the difficulties caused by the pandemic. Within this program, the Italian Government organized the *National Plan for Recovery and Resilience* (PNRR)¹⁸ with different missions, one of which is the creation of a Nation that aims at being more innovative and digital. Of particular interest for this project is Mission 1 – *Digitalization, Innovation, Competitiveness, Culture and Tourism*, Component 3 – *Tourism and Culture 4.0* because it focuses on the relaunch of tourism and culture and at the digitalization of the cultural heritage. The strategic context was provided by the National Digitalization Plan (PND),¹⁹ a document that serves as reference for the realization of the PNRR investment M1C3 1.1 *Strategies and digital platforms for the cultural heritage*. Not only does it enhance the digitization processes of museums, archives, libraries and cultural places, but it also provides for the creation of a digital national infrastructure that gathers, integrates and preserves digital resources.²⁰ One of the most interesting ideas expressed in this work is that of *semantic capital of relationships* (section 5.1.3), according to which the digital environment finds its fundamental essence in the possibility to create and recreate reciprocal connections

¹⁶ EUCO and EP (2021, vol. 694, art. 3)

¹⁷ https://next-generation-eu.europa.eu/index_en

¹⁸ <https://digitallibrary.cultura.gov.it/pnrr-cultura/>

¹⁹ <https://digitallibrary.cultura.gov.it/il-piano/>

²⁰ Marras (2022)

among information, and therefore to facilitate the production of new meanings. In this way, new representational models of knowledge are created, free from those belonging to the institution that preserves the cultural object, but integrated and empowered by a plethora of original points of view. In the web, semantic relationships among digital resources take place and cultural assets become the nodes of a social net to which everyone can take part. Accepting this cooperative vision, it becomes clear that the data need to be organized and modeled in order to be related to each other, also automatically. In this way, the different field scopes can benefit the one from the other and enrich the digital resource.²¹

In this view, the newly created networks are required to be able to use and reuse “the integrated knowledge embedded in the data and metadata themselves while managing the increasingly complex digital data aggregates and their derivatives.”²² In other words, digital resources need to follow common schemata and standards in order to be truly findable, accessible, interoperable and reusable (see chapter 2 and the concept of FAIR data and resources).

Again, in 2021, the Italian Ministry of University and Research (MUR) announced that the EU approved the reprogramming of the *National Operational Program* (PON)²³ on Research and Innovation 2014-2020, with the inclusion of two further Actions of Axis IV – Education and Research for Recovery in order to “provide assistance for fostering crisis repair in the context of the COVID-19 pandemic and its social consequences and for preparing a green, digital and resilient recovery of the economy.”²⁴ In particular, Action IV.4 aimed at funding PhD and research contracts on innovation, digital, enabling technologies and enhancement of human capital as determining factor for the development of research and innovation in Italy.

²¹ MiC (2023)

²² Doerr and Stead (2012)

²³ <http://www.ponricerca.gov.it/pon-ricerca/programme/>

²⁴ EUCO and EP (2020)

Another key factor of these regulations is the investment on human capital and the interchange between the academic and the professional fields. In 2021, to align the regulation of the PhDs in Italy with the recent PNRR, the MUR issued a new ministerial decree in which it is stated that doctoral education allows to promote interdisciplinary, multidisciplinary and transdisciplinary educational opportunities, also in non-academic environments, such as the industry field, the Public Administration, services and cultural institutions, with the involvement of experts of the sectors in the academic activities. It also involves the acquisition of transversal skills so as to favor their transfer and development in scientific and professional environment.²⁵ The mandatory cooperation of PhD candidates with partner companies (private and/or external to the University) in the context of the PON scholarships is a strategic choice inasmuch it allows the acquisition of the aforementioned skills and guarantees the development of highly expert profiles. This aspect is fundamental for Italy to advance in the European scenario as far as digital innovation, usage of new technologies and employees' competences are concerned. There is still a mismatch in relation to supply and demand of skills, meaning that there is a wide gap between the skills owned by young scholars and those required in the world of work.²⁶ The post-pandemic world is experiencing a frantic digital transition and is bringing along the necessity to educate workers specifically on these new technologies but also on the soft skills. In fact, the swift changes that characterize our era influence also the society and workers have to be able to adapt equally swiftly. All the international and national policies described in this chapter clearly go in this direction and higher education tied to professional experience is more and more relevant in the corporate strategic view. In order to develop the PhD functions on research and education in a vision connected to the world of work and to the new professional profiles, it is required to increment the relationships with the society and the different stakeholders.²⁷ Therefore, it is necessary to promote the acquisition of

²⁵ MUR (2021)

²⁶ Boffo (2022). In this article it is possible to read about the last ministerial decree n. 226/2021 which innovates the accreditation of doctoral positions.

²⁷ Boffo (2022)

skills based on direct experience, at the workplace, so as to provide the opportunity to create a context of shared values and meanings within which it will be possible to actively develop the required skills.²⁸ This vision is clearly human-centered, the human is the first actor and it is necessary to begin from his/her needs, peculiarities and abilities to adapt to changes.

To see how the Italian situation concerning digital innovation has developed in the last years, it could be interesting to look at the Digital Economy and Society Index (DESI),²⁹ through which the EC has been monitoring the Member States' progress in key digital policy areas since 2014.

In 2015, Italy was at the 25th position out of the 28 Member States, and the situation has remained mostly unchanged until 2020. In table 1 it is possible to see the Italian situation before the pandemic (the data gathered refer to the previous year).

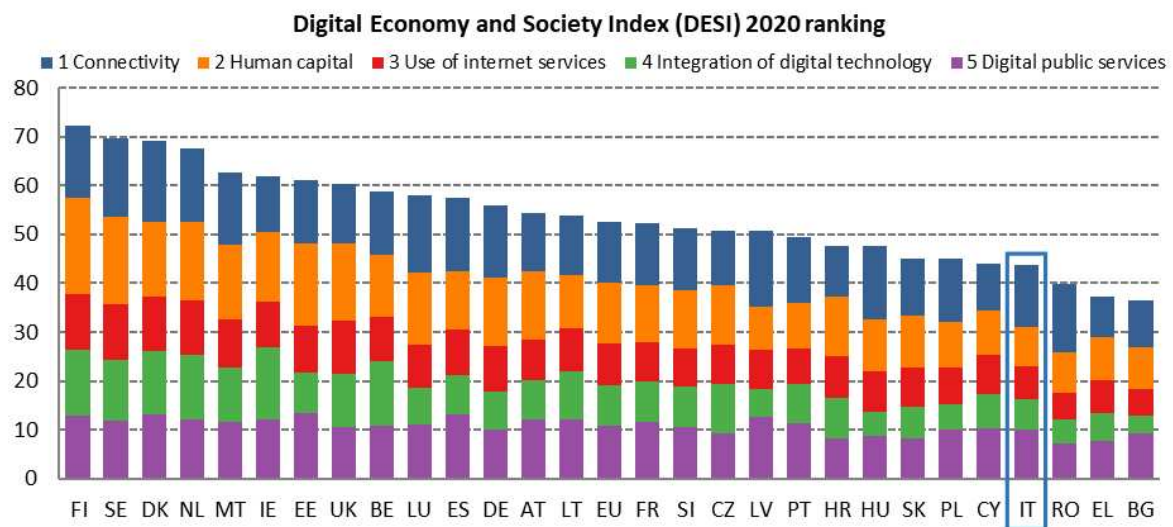


Table 1 DESI 2020

²⁸ Iannotta and Scarano (2022)

²⁹ <https://digital-strategy.ec.europa.eu/en/policies/desi>

The report reads that there is still a great need for basic and advanced digital skills to enhance the human capital and that only the 74% of the Italian population regularly accesses the Internet. What's more, also enterprises resist the use of technologies.

However, 2019 has been a year of digital strengthening from a political point of view and different actions has been taken, for example the establishment of a new Ministry for Technological Innovation and Digitisation, the presentation of the five-year plan *Italia 2025*, projects to help digitize the public administration, the plans *Enterprise 4.0* and *Transition 4.0* to sustain businesses during an innovative and green transition.³⁰

The analysis of the following years shows that these policies proved effective since the position occupied in 2021 in 21st and in 2022 18th out of 27 Countries (the United Kingdom was no longer part of Europe in 2020). Table 2 is the most up-to-date report available.

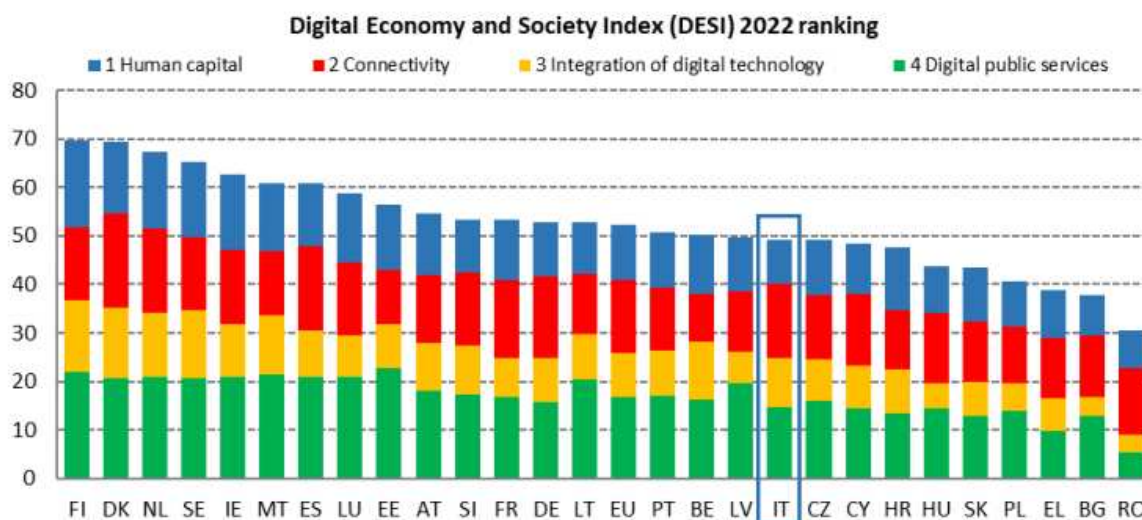


Table 2 DESI 2022

Although some gaps in the digital transformation still need to be overcome,

³⁰ EC (2020c)

Italy is catching up and, looking at the progress of its DESI score over the past five years, it is advancing at a remarkable pace. [...] Giving continuity to the initiatives undertaken and leveraging on the many assets which Italy has would allow the country to further improve its performance in the DESI. The Recovery and Resilience Plan, the largest in Europe, endows it with the necessary funds to accelerate the digital transformation. Moreover, the country has a strong industrial base and research communities in key areas such as Artificial Intelligence, High performance computing and quantum. These strengths should be leveraged to deploy digital in all areas of the economy in the full respect of the human centric approach promoted by the Digital Principles.³¹

Together with the DESI it is interesting to mention the first recently published *Report on the state of the Digital Decade*³² which controls the progress of the Member States towards a successful digital transformation as set out in the *Digital Decade Policy Program 2030*.³³ The general conclusions read that “the success of the EU’s digital transformation will require a substantial acceleration and a deepening of the EU’s and Member States’ action to make reforms, improve the business environment, create incentives and boost investment in digital technologies, skills and infrastructures.”³⁴ In particular, the Italian situation is not encouraging. It is stated that Italy’s progress in digital skills remains slow and less than half the population has basic digital skills, with severe consequences on our capacity to benefit from digital opportunities and our inclusiveness in the European scenario. Enterprises offering trainings to their employees is insufficient, the number of ICT graduates is significantly below the ambitions for the Digital Decade and the provision digital public services for citizens and enterprises is

³¹ EC (2022b, 3)

³² <https://digital-strategy.ec.europa.eu/en/library/2023-report-state-digital-decade>

³³ https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/europes-digital-decade-digital-targets-2030_en

³⁴ EC (2023a, 40)

below EU average.³⁵ However, in particular thanks to the PNRR, what has been exposed so far clearly shows that many measures have already been planned and taken in order to reach the targets set by the DESI and the Digital Decade. Therefore, it can be said that Italy has adopted an innovative vision in order to fill in the gap with respect to the other European States and in accordance with the European objectives.

1.2.The importance of preserving digital and digitized objects

To recall what was mentioned at the beginning of this chapter, it can be stated that the cultural heritage is a complex reality, fragile and threatened by conflicts, theft and plundering. The risk of losing cultural objects is daily and therefore their transmission is uncertain, which is something that cannot be afforded. The protection, presentation and transmission of cultural heritage to future generations are ethical imperatives, it “draws its substance from the past and from living cultural traditions and develops in contact with others. The heritage has to be a shared experience [...], the foremost constituent value of heritage is diversity.”³⁶ So far, the UNESCO has recognized 1199 properties in the World Heritage list out of which 59 are located in Italy, making it the first Country in the statistics.³⁷ In light of what has been exposed so far, digital transition for the cultural heritage seems a suitable solution to face a series of problems that threatens it but also a good way to ensure its promotion by reaching a global and diverse audience. Besides being worldwide accessible, digital cultural heritage allows also for interactive and engaging learning experiences, easing the contact with younger generations. Last but not least, the online availability of huge amount of data creates opportunities for advanced researches by opening new perspectives for scholars and by facilitating collaboration through the exchange of knowledge and resources.

³⁵ EC (2023b)

³⁶ UNESCO (2002)

³⁷ <https://whc.unesco.org/en/list/stat/>

All the actions taken in the last decades by the EUCO and, in turn, the Italian Government, prove clearly that the digital transition is inevitable and its importance is widely acknowledged. Many fields and different types of audience can and will benefit from this shift. A great number of actors, both public and private, have taken up the cause also before the disposal of these European funds because they recognize the importance of creating and delivering “digital representations of cultural and historical documents, artefacts and images to improve access to, and foster greater understanding of, the material they hold.”³⁸ Online availability means ease of fruition, dissemination, preservation, searchability, formation. It breaks down the barriers of distance and of time. Online resources can be used for educational purposes and research. A concrete example is the digitization of the *Arnarnagnæan Manuscript Collection* which preserves nearly 3000 items gathered by the antiquarian Árni Magnússon and is considered as the most important collection of early Scandinavian manuscripts. These resources are preserved in two different institutions in Iceland and Denmark and are part of the UNESCO since 2009. Up to now, a huge number of these manuscripts, together with others preserved in more Scandinavian institutions, is available on the portal *handrit.is*³⁹ for open consultation in the form of digitized books with rich paleographic descriptions, contents, location and bibliographic references. It is evident that their online availability is a huge convenience for the researcher or student that, for example, needs to work on a specific saga handed down by different manuscripts preserved in different institutions.

However, it is important to point out and keep in mind that a digital or digitized version of a cultural object is a *substitute*, and the original form is unique. For a paleographer, the digitized manuscript will not be enough and he/she will probably need to reach the institution anyway. Nevertheless, in many unfortunate cases the online version could be the only solution possible because cultural objects are perishable and might get lost or destroyed. A recent example is the 3D scan of the Titanic performed by Magellan, Ltd that recreates for the first time a full reproduction of the wreckage.⁴⁰ It is impossible for

³⁸ Terras (2012, 47)

³⁹ <https://handrit.is/>

⁴⁰ <https://www.magellan.gg/titanic-in-3d/>

a researcher to go down deep in the see and analyze it directly. Therefore, in this case, this latest 3D model is the best version of the ship available. In this sense, the digitization of the cultural heritage ensures the possibility to have a sort of backup of the assets in order not to lose them forever.

According to this view, it can be said that “[d]ata are no longer mere instruments to simplify administration management or to enhance the fruition of cultural heritage, but they will become digital artifacts representative of the new digital cultural heritage.”⁴¹

However, digitization of cultural objects is not the only step that has to be performed when working in this field. It has already been mentioned that other fundamental concepts are access and preservation. The first one implies the discovery and use of digital objects by the user thanks to the adoption of standardized metadata which allow the retrieval of information. The latter pertains to long-term safeguarding, assuring the continued availability of these digital assets in the future and safeguarding their accessibility, rendering, usability, and comprehensibility.⁴² It is evident that many are the factors that come into play when dealing with this topic: “advanced digitization techniques, robust data management strategies and collaborative efforts between governments, institutions and communities are essential. By adopting responsible digital preservation practices and fostering a deep appreciation of cultural heritage, we can ensure that the echoes of the past will resonate powerfully, enriching the present and inspiring the future.”⁴³

An article⁴⁴ posted on the UNESCO website perfectly summarizes the most salient reasons that should strengthen this digital shift:

⁴¹ Barbuti (2021)

⁴² Sotirova et al. (2012)

⁴³ Wagner and De Clippele (2023)

⁴⁴ UNESCO (2020)

- the role of digital technologies for the safeguarding of cultural heritage for future generations;
- expanding or opening new opportunities for collaborations and finding new audiences, especially in the creative sector;
- different way to access and participate in arts and culture, also reflected in the intersection between culture and education;
- positive steps towards the digital age from Ministries of Culture and public cultural agencies around the world;
- digital technologies are powerful tools for the transformation of societies and for the construction of open, equitable, inclusive, prosperous and pluralistic realities.

To conclude this chapter, it is here reposed the cultural value chain in the digital environment presented in the UNESCO Global Report of 2017⁴⁵ and represented here in figure 1. The traditional model has a pipeline-like configuration, meaning that each actor of the different processes – creation, production, distribution, access and participation – contribute value to a product or service before passing it along to the subsequent stage. On the other hand, the new creative chain takes the shape of a network, where the processes are nodes instead of stages and they interact with each other in real time. Felipe Cesar Londoño, Rector of the University of Caldas and Director of the *Festival Internacional de la Imagen*, contributed to this document with a message in which he states that “It is fundamental to review the concept of culture in light of today’s social and technological transformations. Digital networks and collaborative platforms have created a new way of understanding the processes of creation that empower communities, enable the expansion of new participatory models and promote the intersection of art, design, software, science and technology.”⁴⁶ However, the author of

⁴⁵ UNESCO (2017, 76)

⁴⁶ *Ibid.* (2017, 77)

this chapter laments the fact that among the Parties certain nodes of the chain remain weak because a concrete digital cultural plan is still missing. In fact, in 2021 the EC published another recommendation document⁴⁷ on a common European data space for cultural heritage that encourages Member States to accelerate their digitization and preservation efforts and to seize the opportunities created by the digital transformation in order to enhance the resilience of cultural heritage institutions. Among the other strategies, Member states are incited to collaborate with cultural heritage institutions at international level also in complementary with the UNESCO and CoE's actions. Annex I reports that on February 1st, 2021, Italy contributed to the Europeana ecosystem with ca. 5% of the European total, out of which only 2,5% was made up of high-quality records.⁴⁸ The strategy requires that Member States produce by 2030 almost the same number of records as there are now, but only according to the high-quality standards.⁴⁹

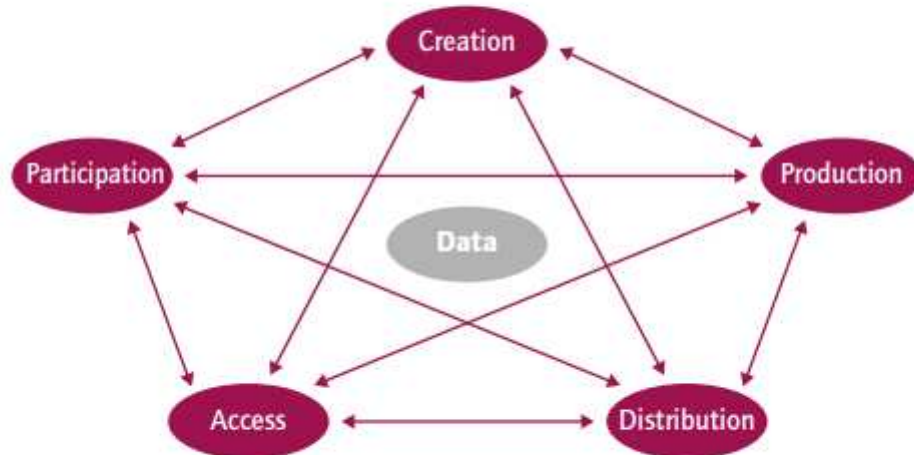
⁴⁷ EC (2021)

⁴⁸ This value is established thanks to the requirements of Tier 2 or above of the Europeana Publishing Framework for content and to the requirements of Tier A or above for metadata. See <https://pro.europeana.eu/post/publishing-framework>

⁴⁹ On February 1st, 2021 the total number of European records was 42.602.653 and the indicative target for 2030 is of 40.017.586 new high-quality records.



In the analogue/traditional model, each actor embodies a particular stage – creation, production, distribution, access, participation – where they add value to a product or service and then pass it on to the next stage in an arrangement akin to a pipeline.



The new value chain should be regarded as a network in which the links (creation, production, distribution, access and participation) are not stages, but rather nodes that interact in real time. Data are the lifeblood of the cultural system and are a key component of the creative economy.

Figure 1 Cultural value chain

2. Methodological background and contextual definitions

The issues discussed so far present the questions that were preliminary to the development of this research. This second chapter delves deeper into the implementation of the project itself by analyzing some key concepts that were taken into consideration when the idea of this web application was born.

Given the multifaceted and interdisciplinary nature of the Digital Humanities, it is difficult to establish some fixed standards and expect all the researchers to follow them. Unfortunately, this is evident in the huge amount of stand-alone projects and resources, which might be of great value for the scholarly community but are developed in absolute isolation and their usefulness is consumed within the very website of the project in question. Absence of standards, of documentation and of open licenses are still widespread and prevent these resources to become part of the digital cultural ecosystem. On the other hand, their observance ensures “the accessibility, identification and retrieval of resources. Descriptive metadata facilitate the resources’ organization, interoperability and integration, provide digital identification and support archiving. Poor quality or non-existent metadata mean that resources remain invisible within a repository or archive thus becoming undiscovered and inaccessible.”⁵⁰

In the following subsection, the FAIR principles are introduced. Nowadays, they are well-known to every scholar working in the Digital Humanities and are ever-present at conferences and in papers. FAIR, in this use, is an acronym that stands for *Findable*, *Accessible*, *Interoperable* and *Reusable*. While all these concepts are present in this work, the idea of *reusability* stands out among the four. In fact, the research started from a survey of existing resources and tools because the main operation performed by this web application is the retrieval of existing digital objects and their transfer to the frontend via the middleware. All the information is gathered through APIs, which are the emblem of exchange and integration of data.

⁵⁰ Sotirova et al. (2012, 31)

In the second part of the chapter, some concepts are introduced, namely those of digital libraries, virtual museums and online repositories as being the main containers of these digital objects. This last element is of particular interest for this research, since the resources visualized on this application are retrieved from online repositories. In chapter three it will be also pointed out that this research started from the first two types of platforms, but the direction had to be changed for reasons of, as a matter of fact, accessibility and reusability.

2.1. FAIR principles

In 2016, an influential article was published which stated that “[t]here is an urgent need to improve the infrastructure supporting the reuse of scholarly data.”⁵¹ The authors published in this occasion the guidelines that are known as FAIR principles.⁵² In particular, they leverage on the fact of enhancing machine-actionability, e.g., leaving the machine the ability to find, access, interoperate and reuse data even without human intervention.

In principles, data and metadata should be:

Findable: they should be easy to find both for humans and computers
(identifiers, rich metadata descriptions, available in searchable resources).

⁵¹ Wilkinson et al. (2016). The original idea was formulated for the first time in January 2014 during a workshop at the Lorentz Center (Leiden, The Netherlands) by a group of stakeholders – representing academia, industry, funding agencies and scholarly publishers – interested in scientific data publication and reuse. In 2016 the FAIR principles were published for the first time in *Scientific Data* 2016, 3.

⁵² <https://www.go-fair.org/fair-principles/>

Accessible: the users should be able to access them and know how to do it (open and free protocol for accessing them, metadata available also without data).

Interoperable: they can be integrated with other data and to interoperate with applications and workflows (formal, accessible, shared languages, FAIR vocabularies, references to other (meta)data).

Reusable: they should be well-described in order to be replicated or combined in different settings (richly described, accessible license, detailed provenance, domain-relevant community standards).

Mons et al. (2020) defined these principles as a “community journey”, as a

guide to create reusable research objects. FAIR is not a new standard; is not a top-down requirement; is not an all-or-nothing binary state (FAIR or not FAIR). The FAIR principles were conceived and designed as a *resource for optimal choices* to be made during many aspects of data and tool generation as well as (re)use and long-term stewardship. They serve the purpose of guiding *implementation considerations* on our *journey to machine-assisted research and innovation*.⁵³

It is important to notice that, as the aforementioned authors wrote, these concepts were no news when the 2016 article was published and that many scientific articles already mentioned the necessity of going into the direction of machine-friendly research infrastructures. The luck that Wilkinson et al. (2016) is said to be twofold. Their article hit the target because, given the popular topic and the effective acronym, it was

⁵³ Mons et al. (2020, 1–2)

massively read. Moreover, it was published and spread during the period in which the European Open Science Cloud (EOSC)⁵⁴ was being prepared and some key concepts such as machine actionability and data share and reuse were particularly encouraged.⁵⁵

2.1.1. Keyword: *reuse*

In the first part of this subsection, it is necessary to briefly go back to some European policies that deal in particular with data management. The same day as the EC released the aforementioned communication on the digital future of Europe,⁵⁶ another document called *A European strategy for Data*⁵⁷ was also made available in which the EC recognizes the importance of digital technologies in our economy and society and the fact that this transformation is driven by data, with enormous benefits for citizens and businesses. In this communication there is a heavy stress on the availability of data and it can be read that “making more data available and improving the way in which data is used is essential for tackling societal, climate and environment-related challenges, contributing to healthier, more prosperous and more sustainable

⁵⁴ <https://eosc-portal.eu/>

⁵⁵ In the *First report and recommendations* of the EOSC it can be read among the *Key factors for the effective development of the EOSC as part of open science*:

- New modes of scholarly communication (with emphasis on machine actionability) need to be implemented.
- Modern reward and recognition practices need to support data sharing and re-use.
- A real stimulus of multi-disciplinary collaboration requires specific measures in terms of review, funding and infrastructure.
- The EOSC needs to be developed as a data infrastructure commons, that is an eco-system of infrastructures.
- Where possible, the EOSC should enable automation of data processing and thus machine
- actionability is key. (Directorate-General for Research and Innovation 2016)

⁵⁶ EC (2020d)

⁵⁷ EC (2020a)

societies.”⁵⁸ Further on in the document there is a consideration on the currently insufficient availability of data for its innovative reuse – in which its value lies – and it is stated that it is a great concern for the public goods since “[d]ata generated by the public sector as well as the value created should be available for the common good by ensuring, including through preferential access, that these data are used by researchers, other public institutions, SMEs [small and medium-sized enterprises] or start-ups. Data from the private sector can also make a significant contribution as public goods.”⁵⁹ The last point in this document that is worth mentioning in the context of this project is the one regarding data interoperability, quality, structure authenticity and integrity, which are key characteristics for the exploitation of the data value. In order to gather and process data coming from different sources in a coherent and interoperable manner, it is necessary to apply standards, shared compatible formats and protocols. The European Research Council (ERC) in its Plan for Open Research Data and Data Management declares that raw data is not enough for FAIRness of stored data: proper documentation and description through informative metadata is necessary. Of course, the choice of the most appropriate metadata depends on the specific research, but a standard that is generally accepted for information description on the web does exist: Dublin Core.⁶⁰ This scheme has been adopted for this project in order to enhance interoperability of resources, as it will be explained in section 4. A useful resource to orient oneself among the metadata standards is the FAIRsharing platform,⁶¹ which works as a guide for consumers to consult resources confidently and for producers who want to improve the findability and usability of their resources.⁶²

Some months after the publication of this data strategy, a regulation on data governance⁶³ was proposed still by the EC which draws inspiration from the FAIR

⁵⁸ EC (2020a, 3)

⁵⁹ EC (2020a, 6–7)

⁶⁰ <https://dublincore.org/>

⁶¹ <https://fairsharing.org/>

⁶² Lister and Sansone (2023)

⁶³ EC (2020b)

principles and in two EUCO's Conclusions during the following year it was reiterated "the importance of better exploiting the potential of data and digital technologies for the benefit of the society and economy"⁶⁴ inviting the legislators to take work forward on specific Acts proposals in order to improve access to, sharing, pooling and reuse of data and "the importance of making rapid progress on existing and future initiatives, in particular unlocking the value of data in Europe, notably through a comprehensive regulatory framework that is conducive to innovation and facilitates better data portability, fair access to data and ensures interoperability."⁶⁵ Again, two years later another regulation was proposed on harmonized rules on fair access to and use of data, the so-called *Data Act*⁶⁶ which aims at further strengthening EU's strategy on the data economy, ensuring fairness of data value and fostering access to and use of data. The explanatory memorandum that accompanies the proposal reads as follows:

Data is a core component of the digital economy, and an essential resource to secure the green and digital transitions. The volume of data generated by humans and machines has been increasing exponentially in recent years. [...] Low trust, conflicting economic incentives and technological obstacles impede the full realisation of the potential of data-driven innovation. It is therefore crucial to unlock such potential by providing opportunities for the reuse of data.⁶⁷

This overview clearly shows that the concepts of open source, data use and reuse, interoperability and more in general digital solutions are key aspects in the European vision. Moreover, the strong accent on the importance of society and human capital is

⁶⁴ EUCO (2021b, 4)

⁶⁵ EUCO (2021a, 2)

⁶⁶ EC (2022a)

⁶⁷ EC (2022a, 1)

translated into great interest in research, as the disposal of many fundings allocated to this field has shown.

After this long preamble on European policies on data, the focus now returns to the application of FAIRness in the present project.

As previously mentioned, although all the four principles are present and relevant in this project, the concept of reusability plays a major role. In fact, no new resources or new tools were created for the development of the project. Existing ones were just gathered, reorganized and, when necessary, formalized (data and metadata of the resources) or adapted (tools). The reuse of data and tools allows scholars to save time and resources and to invest their researches in the advancement of knowledge in the field.

Even if the absence of pure creation might lead the reader to think that the implementation of this application was an easy task, this is far from being true. In fact, data retrieval, selection and formalization, as well as tools adaptation and integration are delicate processes that require scientific and technical skills, not to mention the amount of time needed to perform these operations and the orchestration of the team. All the details will be explained in chapters 3 and 4.

However, as it will be shown in detail in the next chapter, a digital object needs to be *findable* in order to be *reusable* – the lucky acronym also has it as first letter. Both humans and computers should be able to easily find it, “machine-readable metadata are essential for automatic discovery of datasets and services.”⁶⁸ Moreover, ensuring the use of standards and rules for metadata and interoperability allows users to personalize their experience by creating virtual collections and organizing and re-using them.⁶⁹ Addink et al. (2023) in the deliverable 1.3 of their project⁷⁰ have proposed a list of 29 best practices along with more than 60 recommendations for findability, re-use and accessibility of data hosted by research infrastructures. These guidelines are divided into six categories: modalities of access, building communities and trusts, technology and

⁶⁸ <https://www.go-fair.org/fair-principles/>

⁶⁹ Sotirova et al. (2012)

⁷⁰ Biodiversity Community Integrated Knowledge Library (BiCIKL), <https://bicikl-project.eu/>

standards, versioning of APIs and their data, bidirectional linking between infrastructures and API design and naming conventions. Many of these points reflect what has been done during this project, in particular the aspects concerning the adoption of open data, open formats and standards for FAIRness of the resources and of the platform in general. Appendix I of the aforementioned paper further orders these best practices by user groups, and in particular infrastructures, data providers and users. Please refer to the paper for a detailed description.

2.2. Online collections: digital libraries, virtual museums, digital ecosystems and miscellaneous repositories

As already mentioned, the pandemic of 2020 increased the awareness concerning the need of enhancing digital solutions in many working and social sectors. In the field of Digital Humanities, this shift transformation highlighted the fact that supports for research data storage needed to be improved and that previous efforts seemed to be slow and outdated. According to a review on research data management carried out by the University of Oxford in 2020,⁷¹ the research data seem to fall into three categories: active/hot data (accessed frequently and used actively), semi-active/living/warm data (accessed less frequently but still in use) and archived/cold data (rarely accessed). Living data are the most common in the Digital Humanities and, unfortunately, those with less IT solutions for file-sharing and archiving platforms. These data can be identified as open-access digital objects but which are or intended in the future to be updated and/or added according to the research results of the project they belong to in relation to files and metadata, individually and programmatically accessed via APIs.⁷² Moreover, the files are often in different formats with different metadata schemata and projects have diverse research structures and user interfaces. This is the reason why data of this kind need repositories that are actively managed, not mere archives. Researchers often

⁷¹ Chiarelli et al. (2022)

⁷² Strange, Gooch, and Collinson (2023)

implement individual solutions to store, manage and grant access their research data. However, this practice is advised against by the ERC, because even if they are “open to the public and can play a useful role[,] [...] these are generally not deposition databases, and as long as they depend on a single individual and/or funding source, long-term sustainability is challenging.”⁷³

Even if these issues have been particularly boosted in these last years within the scholarly scenario, they are not new. In the last decades, many different “containers” for digital objects have emerged. Each of them offers alternative ways of accessing and sharing data, even in the light of the fact that knowledge and cultural artifacts have undergone a massive evolution. In the different phases of this project, digital libraries, virtual museums, digital ecosystems, and miscellaneous repositories – defined here as online collections – have been addressed to varying extents. This term identifies a platform or service that gathers, organizes and provides access to digital sources (in this specific context, GLAM⁷⁴ sources), which in some cases can be drawn from different providers. These are indeed the platforms from which CHIt in turn gathers its data and metadata.

Before explaining in detail how these collections were considered for the scopes of the research, concise definitions are provided, which do not aim at exhaustiveness but rather offer a common ground for the discussion.

Digital library: digital collection of organized and curated digital works accessible through a computer-based system and networks. Among the other things, they allow to preserve, share, valorize and (re)use the data they contain for different communities.⁷⁵

Virtual museum: “A virtual museum (VM) is a digital entity that draws on the characteristics of a museum, in order to complement, enhance, or augment the museum

⁷³ ERC (2022, 8)

⁷⁴ Galleries, Libraries, Archives, Museums

⁷⁵ ICOM (2020)

through personalization, interactivity, user experience and richness of content. Both the ‘physical’ museum (PhM) and the VM share a common commitment to the institutional validation of content and quality of experience through curatorial process, inherent in the ICOM definition.”⁷⁶

Digital ecosystem: borrowed from biology, in its digital declension it refers to an infrastructure in which the components are interconnected and interdependent, they interact with each other, exchange data and continuously evolve themselves. It is a complex system whose digital objects are in relation with other parts of it and create new ways of fruition for the final user.⁷⁷

Repository: a platform that collects, organizes and stores digital materials and resources in an open-access view, where scholars can deposit their research outputs (that include diverse formats) for various purposes. Repositories are meant to support research, preservation and free access to this material, offering a way to facilitate cooperation within the scholarly community. The strong points of these platforms are ease of access, harmonization and integration of data, in order to create a centralized and re-usable ecosystem.⁷⁸

2.3. State of the art

The classical initial review on the state-of-the-art literature situation that was carried out during the first phase of this project, proved quite disappointing. In fact, a limited amount of works was found, and the search quickly turned to the analysis of web-based resources that were aligned with our criteria. This choice actually proved effective, since

⁷⁶ Polycarpou (2018)

⁷⁷ Marinelli (2020, 42–43); MiC (2023, 72)

⁷⁸ Wilkinson et al. (2016). The authors here also discuss some repositories that do not work in this way and explain how this constitutes a problem. In a way, it was experimented also during the first phase of the present project, as it will be explained further on in this chapter.

the observation of already-existing platforms gave a good overview of the features desired for CHIt.

The first step for the implementation of the web application was clear: suitable resources were needed. But resources are contained within platforms, and finding and analyzing these digital containers had to be the actual starting point. Therefore, in the pursuit of selecting appropriate collections for data retrieval, one of the first steps entailed conducting a comprehensive review of existing endeavors in the field. This led to embark on a thorough examination looking for existing Italian platforms that to some extent preserved and disseminated objects of the cultural heritage. To narrow down the field, the research started from an exploration of the landscape of regional cultural ecosystems. This research uncovered that six regions have already taken proactive measures, each establishing its own cultural aggregators. Moreover, numerous other regions are in the process of planning similar initiatives or have earmarked funds for such endeavors. Specifically, Piedmont, Liguria, Campania, Emilia Romagna, Sardinia and Lombardy have successfully implemented their respective platforms: *Mèmore Piemonte*,⁷⁹ *Puglia Digital Library*,⁸⁰ *Cultura Campania*,⁸¹ *PatER*,⁸² *Sardegna cultura*⁸³ and *Lombardia beni culturali*.⁸⁴

These platforms represent publicly accessible, cost-free resources designed to document, preserve, and disseminate regional heritages. Their primary aim is to facilitate the utilization and reutilization of both data and metadata associated with cultural assets. Notably, these platforms adhere to controlled vocabularies and international standards, ensuring consistency and interoperability. User interaction is facilitated through intuitive interfaces, complemented by user-friendly search functionalities. Furthermore, resources are meticulously organized into predefined

⁷⁹ <https://www.memora.piemonte.it/>

⁸⁰ <http://www.pugliadigitallibrary.it/>

⁸¹ <https://cultura.regione.campania.it/en/home>

⁸² <https://bbcc.regione.emilia-romagna.it/>

⁸³ <https://www.sardegnaicultura.it/>

⁸⁴ <https://www.lombardiabeniculturali.it/>

collections, providing users with structured navigation pathways to explore the wealth of available content.⁸⁵

To this kind of aggregators there belong also *CulturalItalia*,⁸⁶ the portal of Italian culture managed by the MiC through the ICCU.⁸⁷ Cultural institutions from all sectors and levels are involved in the project as strategic partners for the development of the portal, as it is the main Italian contributor of Europeana. In fact, as it will be seen in chapter 3, CulturalItalia is one of the main providers of CHIt via Europeana. The portal is developed as to be a unique access point for the fruition of the Italian cultural heritage in all of its domains (archaeology, architecture, visual arts, environment and landscape, archives, libraries, cinema and media, literature music, shows, popular traditions, human sciences). The access to all the resources made available by the Ministry and both public and private bodies is possible thanks to the adoption of shared international standards and of open-access licenses.⁸⁸ Di Giorgio (2014) summarizes some key aspects of CulturalItalia that are perfectly sharable with CHIt:

- it gathers information provided by partners via metadata, but the original data remains on the website of the provider, which the user can directly and easily access;
- it is an open and in-progress system, meaning that its growth and development are potentially never-ending because it is directly fed by the content providers;
- interoperability of resources is ensured thanks to the adoption of shared standards.

⁸⁵ For an overview of some of the platforms see Di Noia et al. (2016) for PugliaDL, Brunetti (2018) for Mèmore and the webinar at https://www.youtube.com/playlist?list=PL7Y62vJ_eSDLqC-KW2PiABbqIUvmtvcbe for PatER.

⁸⁶ <https://www.culturalitalia.it/>

⁸⁷ The Italian Ministry of Culture (MiC) and the Central Institute for the Union Catalogue of Italian Libraries (ICCU).

⁸⁸ Caffo (2008)

Continuing within this initial phase of our research, an exceptional international portal was encountered, named NetInteractive Documents (NID).⁸⁹ This innovative digital library was of high interest primarily due to its remarkable annotation tool. This tool serves as an integrated feature accessible across a diverse array of resource formats hosted on the platform. Its functionality extends beyond conventional annotation capabilities, empowering users to seamlessly embed various forms of media, including text, hyperlinks, images, and multimedia elements. The versatility of this tool enhances the interactive nature of the platform, enabling users to engage with resources in a dynamic and multifaceted manner.⁹⁰

These collections served as base to understand what direction to take as far as the general *container* was concerned. At this point, it was time to move to the *content*. At its inception, the project was meant to collect data from secondary – not for importance – scholarly works which do not benefit from the visibility ensured by digital ecosystems and risk to be lost in the vastness of the web. Therefore, a number of projects developed under strict methodological rigor that enhance elements of our cultural heritage were selected. Together with these independent works, also digital libraries and virtual museums were considered, with the scope in mind to have their data and metadata interact with each other in a new way, through a new platform that would have allowed to create new relationships and meanings.

Unfortunately, this was destined to remain just a dream, since a major problem arose: these platforms do not explain how to reuse their data, nor expose them in an open-access format. In their survey on open-access GLAM institutions, (Wallace and McCarthy 2023) examine how GLAMs expose open-access data, whether digital objects, metadata or texts, available for reuse. The research is constantly updated on a publicly available spreadsheet and, at the time of writing, Europe holds the 71.6% of the global rate. However, despite the huge percentage on the global scenario, Italy can count only 26 institutions out of the 1177 ones that have been recorded on the European continent –

⁸⁹ <https://www.nid-library.com/>

⁹⁰ Zaka (2022); Zaka and Maurer (2022)

which is slightly more than 2% (table 3). These figures are not encouraging: digital collections need to be published as data sets for computational use because they have an impact on user engagement and need to be more accessible and interoperable.⁹¹

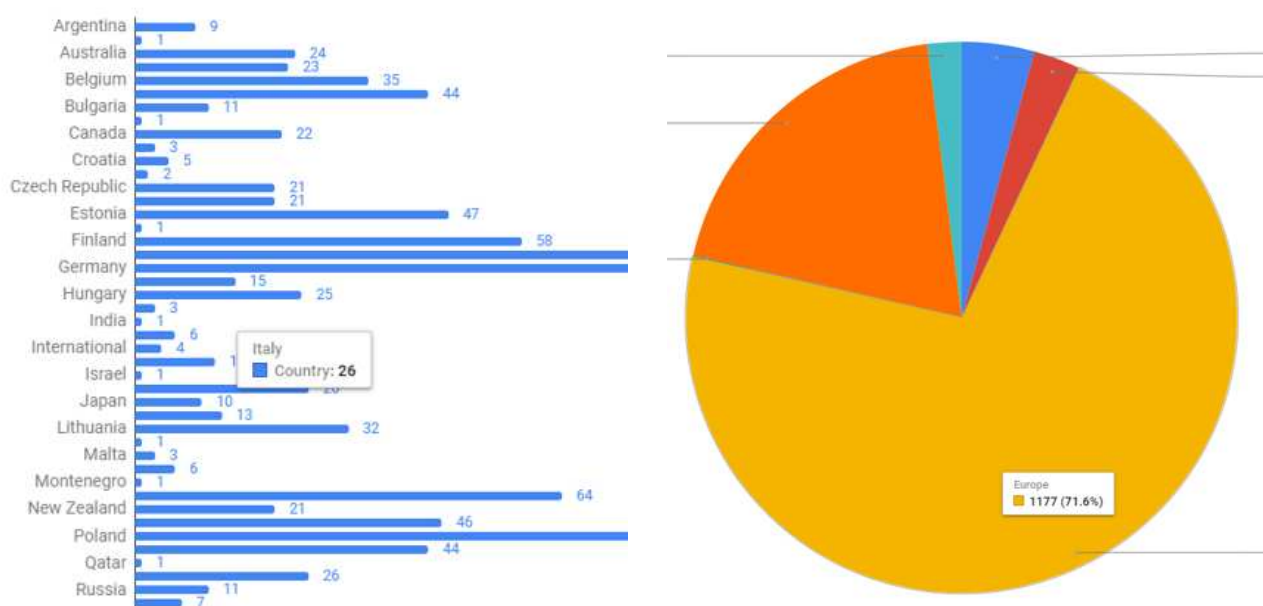


Table 3 Excerpts from the *Open GLAM survey* by (Wallace and McCarthy 2023) – *Instances by Country* (on the left) and *Instances by Continent* (on the right) spreadsheets

When it was clear that these resources could not be gathered – or at least not so easily⁹² – the attention was moved to different kinds of collections. It was decided to broaden the horizons and ended up including resources from Europeana. This turned out to be a good solution, since thanks to the exposition of detailed APIs it was possible to identify a call that allowed for the retrieval of a huge number of cultural objects belonging to the

⁹¹ Candela et al. (2022)

⁹² It was thought of contacting the people in charge of these independent projects to ask them if there was a way to retrieve their data or, even better, if they could consider the hypothesis of depositing them on an open-access repository (see below). This idea has been suggested also by the audience during a workshop where the project was presented. However, given the short time in which this research has to be carried out, this hypothesis has been suspended – yet not discarded – in favor of faster solutions. Hopefully, this path will be pursued during a subsequent phase.

Italian cultural heritage and coming from different providers, the most prominent being CulturalItalia, as mentioned above. The API call and the whole data retrieval from Europeana is better explained in the dedicated section 4.2.1.1.

The content providers exposed in the previous paragraphs deal mainly with cultural objects because they serve the specific scope of preserving and enhancing cultural heritage. However, the seeds of adding value to projects with lower visibility were still growing and, in the end, the research was focused on online repositories. Due to their archival nature, they are more inclined to data presentation, management and retrieval and, in particular, the attention was paid to Zenodo⁹³ and GitHub.⁹⁴ They had been considered as miscellaneous platforms in which users deposit their research data for different purposes. These repositories do not belong to a specific domain and are characterized by a huge variety of topics. Another aspect to point out is that they do expose APIs, and this is the key for reuse, but given their heterogeneous nature it was not possible to retrieve more resources with just one call. For example, GitHub is a cloud-base service for software development and version control via Git and its main use is code storing and managing. The APIs allow to search, for example, for repository or user name or file format, but it is not possible to search for keyword of theme, because these are not foreseen given the more technical nature of the platform. Therefore, a set of repositories (GitHub) and records (Zenodo) were manually selected through their identifiers. In this way, it is possible to overcome the initial problem related to the integration of standalone projects, since these platforms do contain such works from which the research started. Moreover, they do not offer a visualization output given the fact that it is not their aim and in this sense the advantages offered by CHIt are amplified. Even if “[i]t is imperative to acknowledge that this method resulted in a comparatively lower data volume than that obtained through Europeana [, ...] this

⁹³ <https://zenodo.org/>

⁹⁴ <https://github.com/>

approach has enabled the platform to provide exposure to scientifically significant individual resources that might not have received recognition otherwise.”⁹⁵

The last provider that was added during a following phase of the implementation was OntoVE,⁹⁶ a SPARQL Endpoint for the creation of a knowledge base of the Germanic cultural heritage in the Veneto region (Italy). More details on this project and on its integration within CHIt are in section 4.2.1.4. However, it is interesting to anticipate that for this collaboration, a different type of integration was studied, that excluded APIs and included a SPARQL query.⁹⁷ Apart from the mere integration of selected data from OntoVE within CHIt, the work brought forth thanks to this synergy hopes to create a reusable model for the connection to other SPARQL Endpoints.⁹⁸

To conclude this section, it is important to point out that thanks to these API calls it is possible to integrate new resources at need, the retrieval is not static. In fact, the CHIt platform is built as to be extensible, open and ever-growing. However, this is true for Europeana and partially Zenodo, while for GitHub the solution to be adopted turned to be a more static one, since the APIs provided were not adequate for the aims of this project. In fact, the GitHub selected resources, together with their metadata and related files, were manually inserted in the backend of the application (see section 4.2.2). This means that in this case a bit of interoperability had to be given up in order to give space to these resources.

Lastly, the OntoVE resources depend on their updates done in the Endpoint, and in this sense, they are more akin to the Europeana ones retrieved via API.

⁹⁵ Fabbris and Bertozzi (2024, 141–42)

⁹⁶ Frontend: <https://ontove-ui.chitontove.it/it/>, Endpoint: <https://ontove-endpoint.chitontove.it/#/> (at the present moment, the access to the Endpoint requires a login for security reasons but in the future open access will be hopefully guaranteed).

⁹⁷ <https://www.w3.org/TR/rdf-sparql-query/>

⁹⁸ Fabbris and Bertozzi (2024); De Bastiani and Fabbris (2023a; 2023b)

2.4. Why a cultural aggregator

In the previous section, it was pointed out that the web is populated by a large number of heterogeneous – both in content and in technical implementation – platforms that to different extent deal with cultural heritage resources. These platforms disseminate in turn a large number of digital objects, which can adhere to different formats and schemata, potentially creating a disconnected and confusing web of resources. The idea of a cultural aggregator was born with this precise concern in mind, in the hope of developing a technological infrastructure for the harmonization, processing and integration of datasets deriving from various cultural actors. A cultural aggregator has the potential of managing, analyzing and disseminating cultural information on a large scale. It can also be defined as a *digital ecosystem*, since this kind of tool allows different sources to interact with each other and this interaction benefits all of the elements that come into contact, creating new relationships and meanings.⁹⁹

When thinking about the implementation of a cultural aggregator, at least five features have to be kept in mind: data integration from different sources, scalability and adaptability, interoperability, adoption of standards and visualization of data. As already mentioned above, during the preparation of this work it was noticed that a shortage of bibliographic references hampers this topic; this situation probably highlights the fact that these argumentations still need to be done on a large scale. An interesting publication that goes in this direction is Simou et al. (2017).

The first characteristic deals with the retrieval of resources from different providers. A cultural aggregator needs to implement advanced methods for unifying data from disparate sources such as museums, historical archives, libraries, and private collections. This entails normalizing diverse cataloging systems and transforming varying data formats to create a unified, coherent set of cultural data in order to produce interoperable resources.

⁹⁹ De Bastiani, Fabbris, and Bertozzi (n.d.) (submitted paper)

As already mentioned, the volume of cultural data is growing exponentially and cultural platforms and collections are in continuous expansion. For this reason, a scalable and adaptable architecture is fundamental to accommodate to and handle the increasing demand imposed by this lively environment. In this sense, a resilient architecture is indispensable to obtain a robust aggregator that maintains optimal performance levels and guarantees a responsive user experience.

Another key feature in a cultural aggregator is the possibility to interconnect with other external systems. The absence of interoperability in the majority of the platforms that have been analyzed has proved how important it is to avoid the creation of cultural silos. For this reason, aggregators should provide API services which serve as bridges and allow for the connection with other platforms, facilitating data exchange between museums, academic institutions, and other cultural entities. In this way, these bodies can integrate their systems with the aggregator and unleash the potential of cross-disciplinary knowledge and collaboration.¹⁰⁰

Linked to the first point is the need for metadata standards. They are fundamental as they establish a common language for the whole architecture, unifying the different resources gathered from the various providers. The adoption of standards allows for the normalization and consequent alignment of the digital objects. In this way, the concept of interoperability can be referred to again since different digital resources structured with cohesive data create an ecosystem in which they can flow fluidly and facilitate collaboration and collective fruition of the cultural heritage.

The last feature – not for importance – that is mentioned in conclusion to this subsection deals directly with the final users. The mere availability of enriched, interconnected and organized cultural data would be reductive without additional tools for the investigation of the digital objects. Cultural aggregators should offer the creation of immersive and accessible experiences that encompass the needs of different types of users.

¹⁰⁰ Medium (2023)

Visualization and interactive tools serve the scope of making cultural information accessible and understandable to a broad audience and to enhance its dissemination.

2.5. Why a web application

The core of this project is the implementation of a component called *middleware*. A middleware is a piece of software that connects applications and operating systems and eases the communication among applications that were not originally developed to be connected. This is possible because it provides “services that enable different applications and services to communicate using common messaging frameworks such as JSON (JavaScript object notation), REST (representational state transfer), XML (extensible markup language), SOAP (simple object access protocol), or web services.”¹⁰¹ In this way, applications written in different programming languages can still interact with each other. The middleware is the core and middle part of the so-called three-tier architecture (see figure 2), which is also the structure of Muruca¹⁰² – used to build the architecture of CHIt – as it will be explained in the next chapter.

This tripartite structure is composed of three separate and independent modules, namely:

- the presentation tier/frontend: it is the layer that allows user interaction with the application. It presents information to the users and collects their inputs. It displays the information retrieved from the middle tier;
- the logic/application/business/middle tier: it contains the business logic and handles data processing, application rules and the communication between the other two tiers. It processes information collected by the presentation tier using business logic and can also add, delete or modify data in the data tier. Therefore,

¹⁰¹ IBM (n.d.)

¹⁰² <https://muruca.org/en/>

it is an intermediary and ensures that data is processed and exchanged appropriately;

- the data(base) tier/backend: it stores and manages the information processed by the application tier.

In this kind of architecture, the application tier works as a bridge between data presentation and data storage. In this way there is separation of concerns and modularity, scalability and maintainability in application development are enhanced.¹⁰³

The benefits of this architecture are manifold. In particular, the fact that each tier runs on its own infrastructure means that they can be developed and managed separately, on different operating systems and server platforms according to the functional requirements, without having an impact on the others. As a consequence, the development can be quicker because each tier can be implemented simultaneously and performance will also benefit from this as the best languages and tools can be adopted for each part. There are some rules that need to be followed when implementing a three-tier architecture:

- a) layers must be written in separate code files in order to be managed separately;
- b) the presentation layer can only interact with an external agent as far as incoming requests and outgoing responses are concerned and with the business tier to send requests and receive responses;
- c) the previous point applies to the business tier, but shifting one tier: it can receive requests from and return responses to the presentation tier and send requests to and receive responses from the data tier;
- d) again, the data tier can interact only with the business tier for incoming requests and outgoing responses and its requests can be issued only to the database.¹⁰⁴

¹⁰³ Matina (2023)

¹⁰⁴ *Ibid.*

The three-tier architecture is used in distributed systems, that are collections of computing systems in which the different components are physically separated but communicate with each other to reach a shared goal.

Since distributed systems are becoming more and more complex, the use of middleware to ease the integration of their components is increasing as well. Middleware provides a series of common services in distributed systems as it works as an intermediate layer of software positioned between the application and the network, streamlining communications across departments.¹⁰⁵ In the specific case of this project, the middleware was implemented thanks to the adoption of a Node.js¹⁰⁶ application based on the Express.js framework.¹⁰⁷ These technologies allow for the development of open-source and cross-platform applications. More details on them are provided in section 3.2.

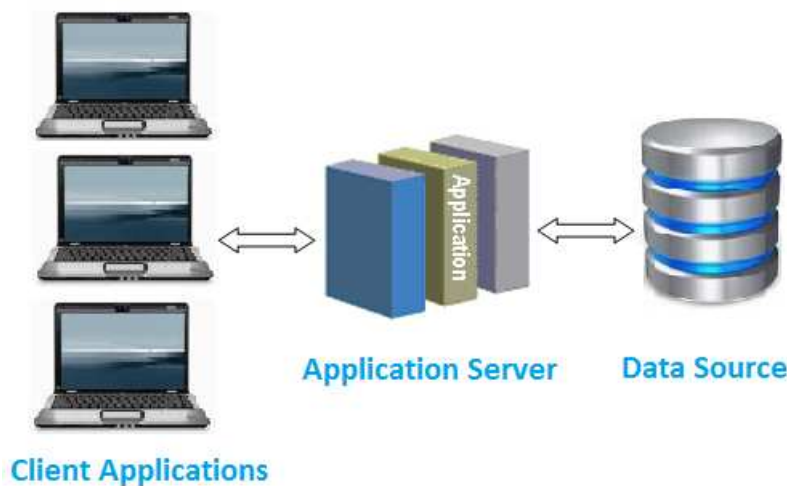


Figure 2 Model of the three-tier architecture. Source: <https://www.softwaretestingclass.com/what-is-difference-between-two-tier-and-three-tier-architecture/>

¹⁰⁵ GeeksforGeeks (2022)

¹⁰⁶ <https://nodejs.org/en>

¹⁰⁷ <https://expressjs.com/>

CHIt is composed of the three layers described above and is developed with the Muruca architecture. Muruca is a product of Net7,¹⁰⁸ the partner company of this project. It is based in Pisa (Tuscany, Italy) and has given support through all the phases of the realization of the project. The company deals with digital products under a wide perspective, but years of fruitful collaboration with the humanities field has brought it to develop a particular interest in the implementation of tools for the cultural heritage and Digital Humanities in an open-science view. One of their flagships is Muruca indeed. It is an exemplary platform that excels in managing, aggregating, and visualizing cultural data. The detailed architecture will be described in detail in section 3.2.

Figure 2 provides an overview of the different components which CHIt is made of and how they relate with each other. There is only one stage that slightly differs from the previous image, the database tier. In fact, the resources available through the frontend are fetched by the middleware in two different ways: either from the backend or directly from an external provider. The reason behind this twofold retrieval is the topic of section 3.2.1.

Besides from that, the structure is the same. On the leftmost side, there are the different providers with their resources stored in their databases and Muruca database. Through APIs and manual insertion, selected items are retrieved and sent to the middleware, where the data collected undergo a process of mapping and standardization of metadata. At this point they are uniform and interoperable and can be delivered to the frontend. Here they are shown as a collection of items and can be individually opened and therefore visualized with EVT¹⁰⁹ and annotated with Pundit.¹¹⁰ Figure 3 shows these

¹⁰⁸ <https://www.netseven.it/>. Link to the GitHub repositories of Net7:

<https://github.com/net7?language=&page=1&q=&sort=&tab=repositories>

¹⁰⁹ <https://visualizationtechnology.wordpress.com/>

¹¹⁰ <https://thepund.it/>

processes in a schematic way.¹¹¹ The specific details of these tools will be presented in the next chapter.

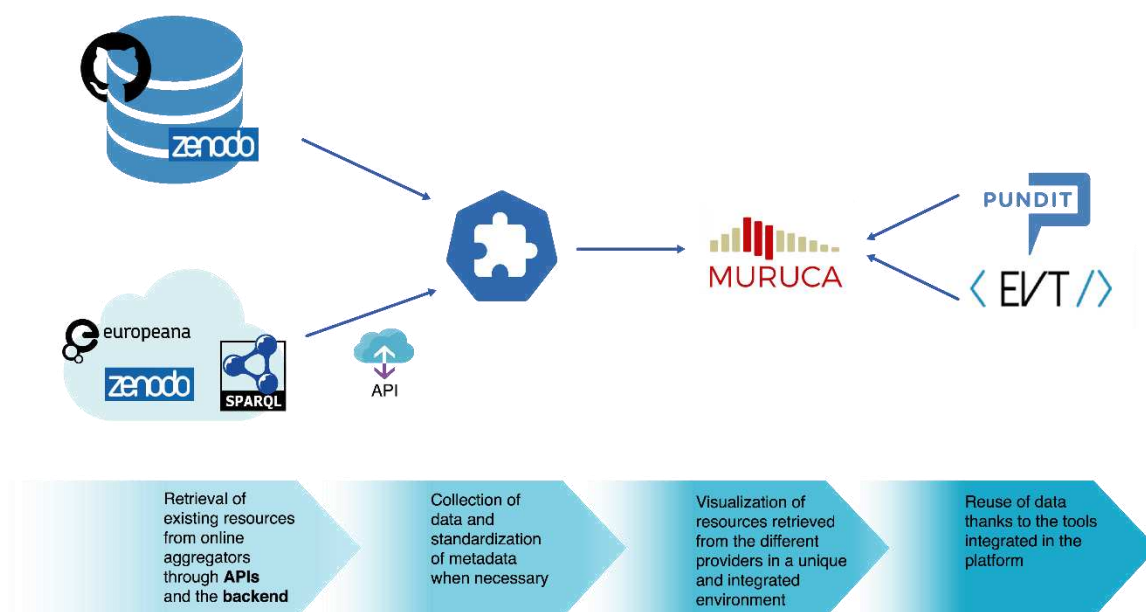


Figure 3 Schema of the application

Of course, MURUCA is not the only platform that is suitable for the implementation of cultural aggregators. Some scholars have already compared useful tools that can be used as repositories and aggregators of cultural data. Cf., among the others, Biswas and Paul (2019) Gkoumas and Lazarinis (2015), Khan (2019) and Pyrounakis and Nikolaidou (2009).¹¹² The main tools surveyed in these studies are DSpace, Greenstone, Eprints,

¹¹¹ This schema was presented during the conference DRHA2023 in Turin, but the version presented here has been adapted to the new organization of the application and therefore differs slightly from the one available in the poster (see Fabbris (2023) for the original presentation and Fabbris and Bertozzi (2024) for the related publication.

¹¹² In their papers, the authors refer to these pieces of software as suitable for the implementation of digital libraries rather than aggregators in general. However, also MURUCA is defined as such, but it is evident that this is not an obstacle since the structure is flexible and adaptable.

Fedora and Keystone.¹¹³ In these papers, careful descriptions are provided, but in the end, it is evident that each software has its advantages and disadvantages. These surveys are useful for an organization that is willing to implement its own platform and can find in these findings a guide to accommodate to its needs since they all report tabs with specific characteristics investigated and the availability or lack in each software. However, it has emerged that Dspace is the most popular software system in the field of digital and institutional repositories.

The tools that are more similar to Muruca are Dspace, Fedora and Omeka.¹¹⁴ Nonetheless, they have some limits as far as data visualization is concerned. The Muruca developing team worked as technical partner in the funded projects *Discovery. Digital Semantic Corpora for Virtual Research in Philosophy*¹¹⁵ E-content plus (2006-2009) and *Agorà. Scholarly Open Access Research in European Philosophy*.¹¹⁶ The context was that of an international consortium that set as its goal the creation of an open platform for the implementation and publishing of Digital scholarly editions.

One of the key aspects that has characterized Muruca since the conception of its architecture is the centrality given to the user experience, as these editions “typically present the need of visualizing complex scientific data, equally understandable and appealing to different types of users like domain expert researchers, school students, and people without a scientific background but still interested in the topic.”¹¹⁷ Muruca was developed with this specific goal in mind, and in fact it allows for a wide range of visualization options, partially derived from the experience gained in the field of data visualization (graphs, maps, timelines) and other specifically tailored for the Digital Humanities field (XML/TEI visualization tools, IIIF support). The greatest strength of

¹¹³ <https://dspace.lyrasis.org/>, <https://www.greenstone.org/>, <https://fedora.lyrasis.org/>, <https://www.klas.com/>, <https://omeka.org/>

¹¹⁴ <https://omeka.org/>, see Tomašević et al. (2017) for an overview of the tool.

¹¹⁵ <http://www.discovery-project.eu/home.html>

¹¹⁶ <https://cordis.europa.eu/project/id/270904>

¹¹⁷ Aiola, Lombardo, and Pedretti (2023)

Muruca is that it offers users the possibility to access data in different ways, yet remaining in the same environment.

Aiola, Lombardo, and Pedretti (2023) conclude that the integration of user experience methodologies starting from the design phase can benefit the final result of the software, “while maintaining up-to-date software standards and adherence to FAIR principles, ultimately improving the process of knowledge production and research by scholars.” ¹¹⁸

Muruca was chosen as software for CHIt also because it is undeniable that the technical support in these projects is fundamental and the training that allows a humanities scholar to become familiar with a complex tool takes time. Therefore, it was decided to exploit the collaboration with Net7 also from this point of view and use one of their products for the creation of this web application. The collaborators of the company, who perfectly knew how to work with Muruca, could train the PhD candidate with ease and could understand her needs and requests concerning the desired result. Moreover, their deep knowledge of the Muruca infrastructure allowed for an agile customization when needed (for example, for the integration of the visualization tool EVT, which will be described in the dedicated section 3.2.5).

To conclude, it can be stated that the requirement set by the PON program on the compulsory collaboration between academic and non-academic field and the consequent externship of the candidates at partner companies for the acquisition of technical skills has proven strategic and fundamental for the success of this project.

¹¹⁸ *Ibid.*

SECTION II: THE INCEPTION AND DEVELOPMENT OF THE APPLICATION

3. CHIt: a tool for the final user

The preceding chapter delineated the foundational characteristics of the infrastructure that serves as the backbone for the CHIt web application, under a very broad and general point of view. Concurrently, a comprehensive theoretical overview of the FAIR principles was provided. Building upon this groundwork, the current chapter delves into a closer examination of the distinct components comprising the platform. Additionally, the seamlessly integrated external tools are scrutinized, which play a pivotal role in enhancing the ability of the application to exploit and thoroughly investigate the diverse array of aggregated resources.

Before embarking on this exploration, a focused inquiry into one of the fundamental tenets of the FAIR principles is undertaken. Specifically, a detailed analysis of the concept of *findability* is conducted. This particular principle stands out as the linchpin that, in practical terms, facilitated the successful navigation through the extensive pool of items accessible on the web. Notably, findability emerges as the foremost characteristic that empowered the ability to pinpoint resources amid the vast expanse of online content.

However, it is important to bear in mind that the significance of the findability concept transcends the confines of the FAIR principles and the domain of Digital Humanities. Examining this from a wider viewpoint, as discussed later, shows that findability is a prevalent and all-encompassing idea that significantly impacts various types of web searches. This emphasizes its widespread influence in shaping the retrieval of information on the internet.

3.1. The added value of being *findable*

The Oxford English Dictionary defines findability as “[t]he quality or fact of being findable; capability of being found. In recent use esp. with reference to the use of online

search engines.”¹¹⁹ However, a simple search such as “findability definition” gives back numerous results related to the field of computer science, since its earliest known use dates to the 1890s in the field of Electronic Industries, according to the OED. Therefore, a more precise definition could be put together paraphrasing Jacob and Loehrlein (2009) and Morville (2005): findability is about orientation, is about the user’s ability to identify the correct website and retrieve the relevant information from it. As the first rule of e-commerce says: *if the user can’t find the product, the user can’t buy the product*. Certainly, the specific context of the research at hand is not affected from the intricacies of trading; however, drawing parallels between the findability of products in e-commerce and that of resources on the platforms under consideration is illuminating. This analogy becomes particularly relevant as the experience matured during the development of this application has already validated its applicability. By navigating through various projects, akin to a consumer searching for a product online, instances were encountered where the lack of findability hindered progress. There were cases where potentially valuable projects were set aside because it was not possible to discern how to appropriately reuse their data. Much like a consumer who doesn't find the desired product on a website and subsequently moves on without making a purchase, the research moved on to the next platform when the findability of resources proved elusive. This underscores the critical role of findability not only in the realm of e-commerce but also in the pursuit of efficient data utilization within the scope of this academic inquiry.

That being said, it has to be pointed out that findability is not just a matter of organizing the resources on the website. It integrates design, users’ behaviors, web standards just to name a few. It is therefore focused “on facilitating and enhancing the user’s overall experience with an information resource.”¹²⁰ The information architecture of a website has to be in line with the users’ expectations and should start from their needs in order to create content interactive navigation. In fact,

¹¹⁹ OED (n.d.)

¹²⁰ Jacob and Loehrlein (2009)

In a world where it's getting harder to reach consumers, shouldn't businesses make it easier for consumers to reach them? Yet, when it comes to findability, most business web sites have major problems. Poor information architecture. Weak compliance with web standards. No metadata. Content buried in databases that are invisible to search engines. From sale to support, many firms could get a great return by investing in findability.¹²¹

It is evident how relevant this keyword is in computer science in general and in Digital Humanities in particular, since through findability, products and resources can actually be discovered and therefore purchased/investigated by users. However, too often it is considered as detached from the content itself, which is something which findability is definitely about. The challenge of content findability extends beyond a mere search problem; it encompasses the broader user experience when navigating a website. Users actively seek content, and it is imperative that this content is not only present but also easily discoverable. The dynamics of user interaction indicate that reliance on advanced search tools is infrequent, and even when users do explore such tools, their proficiency in utilizing them effectively is often limited.¹²² Consequently, the optimization of content findability demands not only a robust search functionality but also an intuitively designed navigation system that aligns with users' tendencies. Addressing these aspects is crucial in ensuring a seamless and user-friendly experience, acknowledging that users predominantly rely on straightforward means of content discovery during their website interactions. These tools serve a top-down logic, they rely on the user's ability to fix and adjust their searches on their own – and here is the problem. A more suitable strategy would be to adopt a bottom-up focus, to begin with the content from the bottom, which cannot be added afterwards or from above. If the information architecture matches the

¹²¹ Morville (2005, 12)

¹²² Nielsen (2011)

users' needs and desires, findability will be inevitably improved and users will not bypass your website.¹²³

In the light of what has been exposed so far on the topic of findability, this section can be concluded with a short overview on how it is possible to create findable resources in the Digital Humanities.

First and foremost, metadata standards need to be adopted from the very beginning in order to present detailed information regarding content, context and format of the resources. The use of standards recognized and shared within the scholarly community guarantees visibility and searchability, and therefore findability. Moreover, controlled vocabularies and ontologies allow to "categorize, quantify, and/or make [the Digital Humanities projects'] objects of study findable" because "what get categorized counts."¹²⁴

One more strategy that can be adopted is thematic classification. Resources classified according to thematic and subject categories are better organized and therefore more easily found by the users because they can avail themselves of these structures to locate what they are looking for. Taxonomies and tags are useful in this context to classify consistently the resources and searches can benefit from this approach.

Still in this view of creating interconnections among data there is the Linked Open Data (LOD)¹²⁵ principle which allows to link structured machine-readable data to other data. This is achieved through the adoption of four simple principles:

1. the (digital) object needs to be identified by a uniform resource identifier (URI) which unequivocally refers to that specific item;
2. the URI needs to be connected to the HTTP in order to be machine-readable;

¹²³ Baker (2013); Dewi, Sabariah, and Effendy (2019); Morville and Rosenfeld (2006)

¹²⁴ Guiliano and Estill (2023, 1088)

¹²⁵ Ontotext (n.d.-a)

3. data should be represented formally with descriptive standards (e.g., RDF) and findable via standard query languages, such as SPARQL;
4. it is good practice to add more links between pieces of data to discover new relationships, offer more contexts and meanings and find more information.¹²⁶

One last suggestion regards the adoption of a user-centric approach, which encompasses different scenarios. To name a couple of these, gathering users' feedbacks, in particular domain experts' ones, can be a good way to improve the organization of resources within a website and enhances the users' engagement. Secondly, the implementation of interactive tools (for annotating, visualizing, sharing comments, etc.) improves the overall user experience and encourages exploration. In this way, the website can benefit from the community's inputs and reflects the actual research trends and different perspectives.¹²⁷

3.1.1. Exposition and importance of APIs

So far, APIs have been mentioned only briefly, but they are another key concept in the development of the project and need to be addressed more in detail. This acronym stands for Application Programming Interface and denotes a software that acts as an intermediary for two or more applications in order to let them communicate. They are crucial in web development as they ease complex processes and reduce code-writing time and effort because one of the main principles is *reuse*: if a desired functionality already exists, it has to be reused. This means that it is better to contribute to existing APIs first, rather than rewrite already-existing functionalities.¹²⁸ Data extraction and data access through APIs is meant for the program and the programmer, not for the end

¹²⁶ EC (2022c)

¹²⁷ Cahill (2009); Lampert and Chung (2011)

¹²⁸ SBB (2022). For a deeper presentation of other principles to take into consideration when dealing with APIs, please refer to chapter *Architecture* of this guide.

user. They abstract the information of the application and present what the developer needs, acting as mediators and allowing request, exchange and reception of data to and from programs in “flexible software ecosystems that encourage innovation and increase capabilities.”¹²⁹ Nowadays, they (should) adhere to standards such as HTTP and REST, with the result that they can be effectively monitored and managed for performance and scalability. APIs are crucial in the digital transformation scenario, since they help grow, digitize, connect and innovate digital objects and – to continue with the commercial comparison – products.¹³⁰ They must be clearly exposed in the website and logically organized to improve findability. Their names must be precise and meaningful in order to be transparent and easy to find and use. The rationale behind APIs findability can be summarized as:

1. higher reuse thanks to better transparency and findability;
2. higher collaboration between APIs consumers and providers thanks to better findability;
3. better stability and efficiency thanks to better documentation.¹³¹

3.1.2. The transformation within the digital paradigm through APIs

So far, a synthetic but exhaustive description of APIs and their exposition has been presented. But what is their role in the digital transformation scenario? How do they actually increase the growth, digitization, connection and innovation of digital objects and products? An incisive summary is provided in the concluding remarks of Chikobava (2023)’s article:

¹²⁹ Chikobava (2023)

¹³⁰ Frye (n.d.)

¹³¹ SBB (2022)

APIs play a crucial role in modern web development by encouraging innovation and simplifying procedures. They enable developers to create new features and improve user experiences by enabling integration and communication between various applications and services. APIs also contribute to the accessibility of technology by granting businesses of all sizes access to important resources and capabilities, thereby balancing the playing field for small businesses and independent developers. As the digital landscape continues to evolve, the significance of APIs in web development will only increase, influencing the future of online interaction and connectivity and ensuring that web applications remain adaptable and relevant in a rapidly evolving and expanding digital ecosystem. In addition, APIs improve productivity by reducing development time and effort, allowing developers to concentrate on their unique value proposition and primary competencies.

Standardization leads inevitably to an increase in interoperability, and a standardized way of communicating among software, systems and components enhances flexibility, scalability, usability, performance and maintenance.¹³² It also allows developers and scholars to exploit pre-built functionalities so that they can focus on creating new features, thus reducing development time. This interconnection among systems and people can be described in other words as collaboration, and, given the context, collaboration results in an expansion of the digital ecosystem. Different platforms and different devices are seamlessly compatible despite their different nature, services and data are exposed through APIs and accessible and available for the community in a viewpoint of transparency, collaboration and innovative development.¹³³

¹³² Arroyo Poleo (2021)

¹³³ For an overview on how to construct your API architecture and why you should do it, see Jin, Sahni, and Shevat (2018); Jacobson, Brail, and Woods (2012); Moilanen et al. (2019).

3.2. Components

Now that some key concepts that are fundamental for this application have been defined, an examination of its actual architecture can be conducted. First of all, it is important to define an architectural style when developing web services in order to have a pattern that helps you design your system quicker. The architecture that best fit CHIt is REST, which stands for REpresentational State Transfer and is one of the most adopted for API development. It provides standards between computer systems on the web, through which systems communicate easier with each other. REST makes an optimal use of an HTTP-based infrastructure and thanks to the HTTP methods, it allows to perform CRUD (create, read, update and delete) transactions on the resources they identify, name, address and handle on the web.¹³⁴ RESTful systems are characterized by statelessness and separation of concerns of client and server, two features that bring benefits for services on the web. In stateless systems, the states of server and client are unknown to each other, they do not track information from previous sessions but can understand all they need to know during each request/response transaction, which contains the description of the object required. On the other hand, it has already been mentioned the separation of client and server, which means that these two components can be implemented independently and changed without affecting each other's operations. Thanks to these constraints, "RESTful applications achieve reliability, quick performance, and scalability, as components that can be managed, updated, and reused without affecting the system as a whole, even during operation of the system."¹³⁵ Through REST, a client can make a request to the server to perform CRUD operations on data. For example, a GET or a POST request is sent to CHIt providers in order to retrieve their resources. The other basic HTTP verbs are PUT and DELETE.¹³⁶

¹³⁴ Jin, Sahni, and Shevat (2018)

¹³⁵ Codecademy (n.d.)

¹³⁶ A general introduction to CRUD operations can be found in this article <https://www.codecademy.com/article/what-is-crud>.

So far, the three-layer architecture and the request/response mechanism have been tackled upon only in general terms. Therefore, the remaining subchapters of this section are dedicated to the exploration of the individual components constituting the CHIt framework, in order to focus on a more granular examination of each element and to shed light on their distinctive roles and contributions within the general architecture.

As highlighted in section 2.3, the design blueprint of CHIt draws inspiration from Muruca architecture, embodying a tripartite structure that includes a backend, a middleware, and a frontend. It is a web solution that has been developed to go beyond some of the problems highlighted in the previous chapter, in particular the combination of the increasing digital market and the creation of *ad-hoc* platforms for single scholarly projects.¹³⁷ Muruca has been developed as to be easy to use, to be standard and is divided between content and design. The backend is built on WordPress,¹³⁸ one of the most popular content management systems (CMS) for web sites and is made of a series of plugins that make it particularly flexible. Moreover, data can be exported into files and stored on different platforms for their long-term preservation. The search engine integrated in Muruca is Elasticsearch,¹³⁹ a distributed search and analytic engine that allows to store, search and analyze huge volumes of data quickly and in near real-time. It is well-known thanks to its simple APIs, its distributed nature, its fastness and scalability. It is the main component of the Elastic Stack,¹⁴⁰ a set of open-source tools for the ingestion, enrichment, archiving, analysis and visualization of data.

The adoption of an existing framework is a practical solution and helps reducing time loss. In fact, if the general architecture is already present, it is necessary only to adjust the specific functions as required by the project at hand, while the others can be taken from the base library already integrated in the framework.

¹³⁷ Aiola et al. (2021)

¹³⁸ <https://wordpress.com/>

¹³⁹ <https://www.elastic.co/>

¹⁴⁰ <https://www.elastic.co/elastic-stack>

The other components – the annotator, Pundit, and the visualization tool, EVT – are integrated in and collaborate with this basic structure.

Everything has been designed in the view of openness and interoperability, both within the application itself and with other systems thanks to the adoption of international standards such as JSON and DCMI. More technical details on the implementation can be found in chapter 4.

In the field of Digital Humanities, there are other platforms that are in line with this approach. TEI Publisher¹⁴¹ e Alphabetica¹⁴² just to name a couple, which both dispose of APIs and IIIF standards. It is noteworthy to mention also CulturalItalia, which makes available all the LOD of its providers through a SPARQL Endpoint.¹⁴³

3.2.1. Muruca's architecture: the backend

As already mentioned, the backend is that piece of a technological infrastructure that works on the server-side, that is not seen on the website and stores and manages the data of the application.

In the specific context of this application, the backend ensures the seamless aggregation, organization and management of all the data retrieved from the various cultural sources. It employs both SQL (structured query language) and noSQL databases and for this reason it offers a scalable and flexible architecture that is able to handle and store large volumes of data. In fact, SQL databases, also called RDBMS (Relational Database Management Systems), use structured query language and excel in managing complex data with predefined schemas. On the other hand, noSQL databases, also called

¹⁴¹ TEI Publisher, as it will be mentioned also further on, is an application that allows to publish TEI-compliant digital editions, <https://teipublisher.com/>.

¹⁴² Alphabetica is the digital ecosystem for the Italian bibliographic services, <https://alphabetica.it/web/alphabetica/>.

¹⁴³ <https://dati.culturaitalia.it/>

distributed systems, are non-relational and have dynamic schemas for unstructured data. This dual approach ensures that the backend can effectively accommodate the heterogeneous nature of cultural information, offering an adaptable storage solution for different types of datasets.

Moreover, this level is responsible also for the processes of data cleaning, normalization, and insertion. Data cleaning is fundamental when combining multiple data sources. Thanks to this action, inconsistencies and errors are fixed or removed, ensuring the integrity of the information stored. Normalization has already been mentioned, which allows to harmonizes disparate data formats, enhancing coherence and consistency within the dataset. Insertion finalizes the incorporation of refined data into the aggregator, marking the culmination of meticulous processes that are indispensable for maintaining the dataset's coherence, reliability, and usability.

As a matter of fact, the data stored within the CHIt backend are those retrieved from Zenodo and GitHub. As already mentioned, these services do expose APIs, but given their heterogeneous nature, an API call for a massive retrieval of resources belonging to the sphere of interest of CHIt could not be formulated. Therefore, the records and repositories that were in line with the project have been manually selected. The Zenodo search interface allows to look for items, among the other things, according to the title and keywords. The keywords are assigned by the single authors when they upload their records, and this permits to find related works. On the other hand, the search on GitHub is less assisted from the point of view of CHIt because this platform was born as a cloud for code. For this reason, the parameters for searches are “limited” to the information on the repositories, users and code-related aspects such as language, extension and issue options. Therefore, to start populating CHIt with the resources of these providers, it was necessary to rely on existing knowledge of scholars and projects related to the research field of this project.

According to this mechanism, the selected resources are stored in the backend and the relevant metadata are manually added. Only in the case of Zenodo, related files are retrieved via API. In this way, when the user requests an item from GitHub or Zenodo,

the frontend sends a request to the middleware, which identifies the origin thanks to the resource ID (see section 4.2). The middleware in turn retrieves it from the CHIt backend in the way in which it was manually stored.

3.2.2. Muruca's architecture: the middleware

It has already been mentioned that the core of this application is a component defined to as middleware. It is here that the greatest development efforts have been put and it is here that the biggest part of the abovesaid FAIR principles, standards and good practices have been merged. It can be called it also *connector* or *mediator*, since it is the link between the backend, the external providers and the frontend. Chapter 4 is entirely dedicated to the exploration of this central component, so here only a general description of the workflow that characterizes its functioning is provided.

From a technical point of view, when a user selects a resource through the frontend, the middleware acts in different ways according to the provider that preserves the desired element.¹⁴⁴ In case it is Europeana, the middleware makes a request for this item through APIs; for GitHub and Zenodo, it has already been mentioned the fact that the selected resources are stored in the CHIt backend. Therefore, they are fetched directly from there. If the resource is stored in the OntoVE dataset, the metadata are retrieved directly from the SPARQL Endpoint. The collaboration that brought to the implementation of this feature and the description of it are explained in detail in the various sections of section 4.2.1.4.

At first, the metadata of the item received in the response adhere to the providers' schema. Subsequently, the Node.js scripts provide the format that allows to refactor the resources in order to make them compatible with the other parts of the platform and, thanks to the adoption of international standards, to be potentially interoperable also

¹⁴⁴ The provider is identified thanks to the ID assigned to the single resources by CHIt, see the dedicated discussion in section 4.2.

outside the platform itself. The adherence to a standardized model allows also for an internal reorganization and clean presentation of the results, in order to offer the users a smooth and pleasant navigation and fruition.

The middleware is implemented within the Node.js runtime, which is an asynchronous event-driven, open-access and cross-platform environment. It executes JavaScript code outside a browser and is frequently used to implement services like APIs. This runtime is often adopted in the development of web services because its architecture pattern is particularly suitable to create scalable applications and is highly adaptable. By being asynchronous, it allows “to start a potentially long-running task and still be able to be responsive to other events while that task runs, rather than having to wait until that task has finished. Once that task has finished, your program is presented with the result.”¹⁴⁵ With asynchronous functions, many connections can be handled simultaneously on different threads which do not block each other. In this way, the potential delay which is typical of synchronous operations can be avoided.

Node.js is also a cross-platform runtime, which means that it is suitable for different operating systems (OS) without the need to be adapted and therefore the application can run on multiple OS. This has several advantages, some of which are the reuse of code that allows to save time and costs, a broader reach of audience as their OS is not relevant and a unified user interface and user experience.¹⁴⁶

To sum up, Node.js is a good choice for the development of real-time web applications, chatbots and IoT applications, collaborative tools, single-page applications, dynamic websites, streaming applications, microservices and APIs. On the other hand, it would not be the perfect choice for CPU-intensive applications, relational databases, simple HTML and CRUD applications, large and monolithic applications and when strong object-

¹⁴⁵ MDN (2023). The advantages of this runtime are synthetized in Dziuba (2023): scalability and performance, versatility and the use of JavaScript, code reusability and sharing, rich ecosystem and microservice architecture.

¹⁴⁶ Techmango (2022)

oriented programming is required. Last but not least, it is not recommended if the developers team lacks experience in JavaScript and, of course, Node.js.¹⁴⁷

The middleware is implemented within the Express.js framework, a minimalist and flexible web framework and the most popular for Node.js whose robust features and tools make it suitable for the development of scalable applications. It enables, among the other things, to manage HTTP requests and responses. Even though it is minimalist itself, there are compatible middleware packages maintained by the Express team that address almost any web development problem.¹⁴⁸ However, the implementation of an application is not a straightforward path and there is no right way to do it. The solutions adopted by other developers and shared online could not be optimal for your application.¹⁴⁹

To sum up, it can be asserted that the middleware is an API itself, and therefore it embodies the principle of interoperability. In fact, it connects frontend and backend and mediates the communication between the two components, it is not directly accessed by the end user, rather it operates behind the scenes. Within its architecture, the middleware contains a large number of endpoints thanks to which it can perform the aforementioned HTTP requests and CRUD operations.

3.2.3. Muruca's architecture: the frontend

This last subsection dedicated to the Muruca architecture presents the user-visible part: the frontend. In general terms, the frontend is that piece of software that works as a gateway for the user to interact with the application employing both the graphical user interface (GUI) and the command line. Also called *client side* of an application, it is responsible for all the visual aspects of a website: text colors and styles, images, graphs

¹⁴⁷ Dziuba (2023)

¹⁴⁸ <https://expressjs.com/en/resources/middleware.html>

¹⁴⁹ MDN (2024)

and tables, buttons, colors, and a navigation menu. In the cultural heritage domain, It typically includes user interfaces like dashboards, interactive maps, virtual galleries, and search tools, allowing intuitive system interaction.

The primary objectives of a frontend are responsiveness and performance, aimed at ensuring a captivating user experience. By focusing on the visual aspects of the website, it goes beyond mere aesthetics. It acts as the key conduit for users to access, comprehend, engage and interact with the rich content available on the platform.

The Muruca frontend is responsible in particular for the presentation of aggregated data in an accessible and comprehensible format. This layer of the architecture is designed to be user-friendly, allowing both technical and non-technical users to access and utilize the platform's resources. The intention is to democratize access to the platform's resources, allowing a diverse audience to harness its capabilities effortlessly.

From a more practical point of view, the Muruca frontend offers a wide range of functionalities. It allows to visualize the records, to organize the elements into collections or courses, to create structured tabs with detailed content and create relationships between tabs, to perform advanced searches, to transcribe in .xml format, to manage editorial content in a customized way and to display data on maps and timelines.¹⁵⁰ These capabilities empower users to interact with the cultural heritage data in a dynamic and personalized way.

Furthermore, Muruca integrates external tools to enhance some functionality, all of which are either used and accepted as standard or developed by large communities. As an example, the reader that has been adopted is TEI Publisher, which is open-access, standard and reusable. This integration exemplifies the platform's commitment to leveraging established tools within the academic community. However, for the CHIt project, it was decided to implement the visualization of resources using Edition Visualization Technology (EVT). Despite the divergence in tools, EVT maintains a high

¹⁵⁰ Muruca has been used to create a demo on story telling in scientific communication. The website is available at <http://murucaraconta.muruca.cloud/en> and a related publication is Aiola and Lombardo (2022).

level of FAIRness, aligning closely with the standards set by TEI Publisher. A comprehensive exploration of EVT and its role in the Muruca architecture will be provided in the concluding section of this chapter.

3.2.3.1. Collections

The creation of collections on a website is crucial because it facilitates the navigation and helps users under many points of view. First of all, the grouping of resources in a logical manner allows users to easily find what they are looking for without getting lost in an unstructured jungle of information. In this way, collections serve as navigation aids that can reduce loss of time when users are looking for a specific item. By navigating well-organized collections, users can also discover new and relevant contents that would have remained hidden otherwise and this can also further engage them in their exploration path. Moreover, it is not secondary the fact that collections contribute to the design of the user interface and an aesthetically pleasant website improves user satisfaction and browsing experience.

A concrete example is Europeana. This platform has a specific section dedicated to collections, which is in turn divided into themes, topics, centuries and organizations. This feature has proven essential for our research because the *themes* facet was used to extract via API the resources that are now gathered in CHIt. It was also decided to keep them as keywords for the creation of our collections, to start with. A further subdivision for the Germanic cultural heritage in Italy was also added, and this will be the topic of chapter 6. In fact, it is dedicated to three case studies carried out from specific points of view – a student, a group of tourists and a Germanic philologist – that are interested in topics related to this kind of heritage, even if their aims and their use of the platform are divergent, of course.

As already mentioned, the Muruca backend is built using WordPress, and it allows for the creation of the so-called collections and itineraries to satisfy the abovesaid needs. Under a practical point of view, collections are pre-defined researches with a chosen

filter that regroups a series of resources adhering to desired topic. In the specific case of CHIt, a dedicated filter was added to the resources that were intended to be part of the collections. To be more precise, since the collections are four (written sources, digital scholarly editions, Germanic language islands in Italy and the Lombard and Gothic people in Italy), there are four dedicated filters assigned to the desired resources.¹⁵¹ When the user clicks on the specific button dedicated to the collection, he/she will be redirected to this pre-selected search and all the resources with the specific filter will be automatically gathered for him/her.¹⁵²

However, this solution does not seem to be the cleanest one under a technical point of view. For this reason, in chapter 7 a couple of different ways of retrieval are presented as possible future improvements.

On the other hand, itineraries are different. They are independent and more complex pages created in the backend with a more structured layout. They usually contain some structured information on the topic and, among the other things, possible external links, multimedia elements and related resources or further itineraries. In this way, users have at their disposal stories made of a selection of resources and are guided in the exploration of these paths and interconnections among resources that could have remained hidden otherwise. Figure 4 shows an example of itinerary created for the Germanic minority languages in Italy.¹⁵³

¹⁵¹ Europeana subjects are not part of this selection, as they are too many and a way to manage all of them still has to be found.

¹⁵² Unfortunately, at the present time this operation is performed only on the resources available in the backend. In fact, the items gathered from Europeana and OntoVE reach the frontend directly via the middleware and cannot be manually edited.

¹⁵³ CHIt is a platform subject to change for its very nature. Therefore, screenshots as presented in this work might change look on the web app over time.

Le lingue minoritarie in Italia di origine germanica

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.



Risorse
collegate

Walser Cultura - banca dati
La banca dati W...

CLIMAlp salvaguardia delle lingue
minoritarie in Piemonte e Valle
d'Aosta
Il progetto CL...

Dizionario del cimbrio dei Sette
Comuni

Figure 4 Example of itinerary

3.2.4. Pundit: the annotation tool

The first and second chapters contain an extensive discussion on the increase of data in the last decades and the importance for them to be, among the other things, interoperable. In this chapter another key aspect has emerged, which is users' interaction and engagement. Whether users are students or scholars, they need "to do things" on the resources they look for in order for these data to be meaningful and to acquire new meanings. Annotation of digital objects and annotations on web

applications is something that has been around for a long time now, there are different tools that allow to highlight texts and add comments and tags. However, more structured types of annotations could benefit the content itself, the user and the community in general. In fact,

if properly structured and provided with clearly-defined and machine-processable semantics, annotations can constitute themselves a primary information which can enrich the original contents and provide added value for other users as well as for third party applications. On this line, Semantic Web technologies are employed to foster the flexibility and interoperability of user created annotations, to promote their linkage with the Web of Data and to permit their reuse by other people or applications beyond the context they originated from.¹⁵⁴

Pundit was chosen as annotation tool for CHIt because it offers all the most common functionalities plus some features that improve its performance and FAIRness. It was first designed by the research group on Semantic Web and Multimedia called Samedia¹⁵⁵ (Ancona, Italy) and further developed in collaboration with Net7 for the application in different research and commercial projects. It is open source and is also available on the EOSC Marketplace.¹⁵⁶

Under a technical point of view, the architecture of Pundit is structured into four modules: the annotator, the backend, the data layer and the application. The software architecture has changed significantly since its first implementation in 2009, in particular concerning the annotator and the backend, which have been redeveloped from scratch. For further details on the architecture, see De Santis et al. (2021).

¹⁵⁴ Grassi et al. (2012, 50)

¹⁵⁵ <http://samedia.dii.univpm.it/>

¹⁵⁶ <https://marketplace.eosc-portal.eu/services/pundit-a-semantic-powered-web-annotation-service>

The Pundit application basically allows to interact with texts in four ways: by adding comments, tags, highlighting and semantic annotations. These last are a way to enrich texts with “metadata about concepts (e.g., people, places, organizations, products or topics) relevant to it. Unlike classic text annotations, which are for the reader’s reference, semantic annotations can also be used by machines. Semantically tagged documents are easier to find, interpret, combine and reuse.”¹⁵⁷ Pundit’s data model is based on named graphs and on the Open Annotation data model and its annotations are built as sets of triples. In this way, annotations reach a higher level of expression because users can use ontologies and specific vocabularies to create their semantic relations among resources. This results in non-ambiguous and precise annotations and reuse of knowledge available on other websites. Figure 5 shows how the four different types of annotations are rendered in the user interface of the Pundit application.

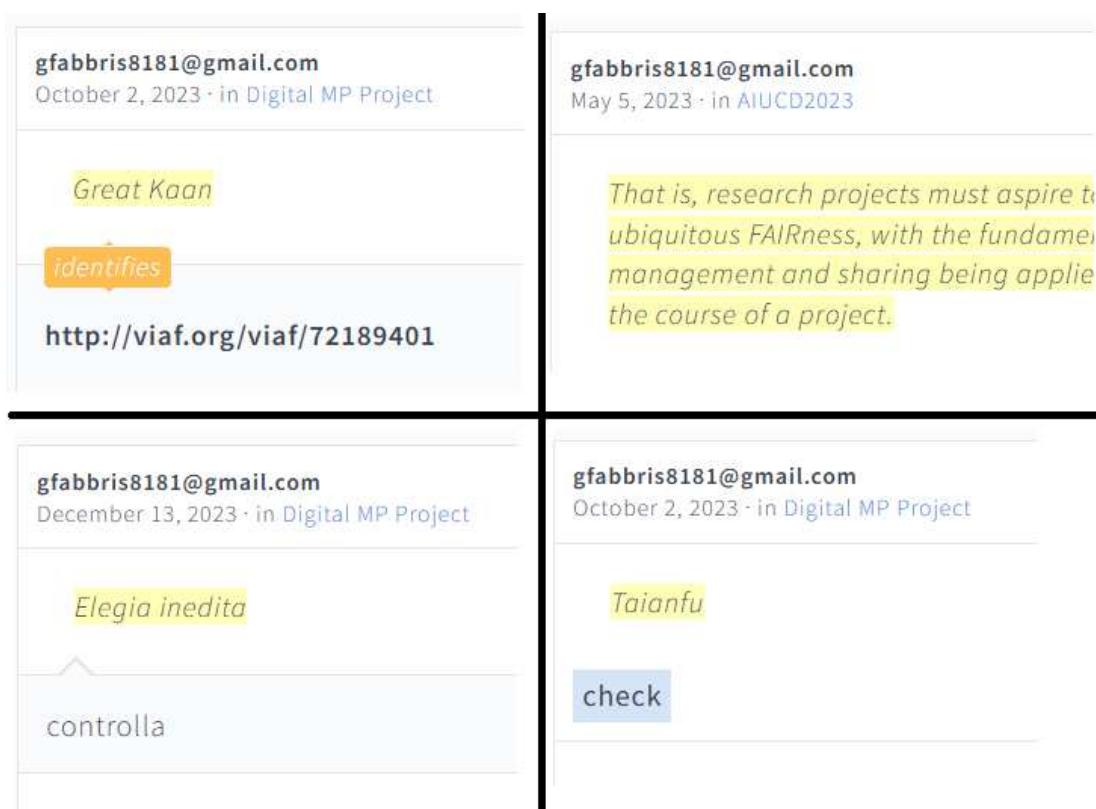


Figure 5 Possible annotations with Pundit: semantic annotation (top-left), highlighting (top-right), comment (bottom-left) and tag (bottom-right)

¹⁵⁷ Ontotext (n.d.-b)

Another interesting feature of Pundit is the availability of notebooks, which aggregate a set of annotations in a single space in order for them to be organized and easily retrievable. Notebooks can be private or public, and these last are particularly interesting for collaborative works. A public notebook can be shared through its unique URL and sent to collaborators, who can participate in it and can reply, like, dislike and even report an annotation to the Pundit administrator. Annotations can be exported in different standard or open formats and be filtered from within a notebook with a faceted interface. All the notebooks are gathered on a single dashboard, from which annotations can be easily accessed and managed. On the website there is a dedicated section for videos and tutorials that helps and guides in the discovery of the different functionalities. In figure 6 an overview of the dashboard is provided.

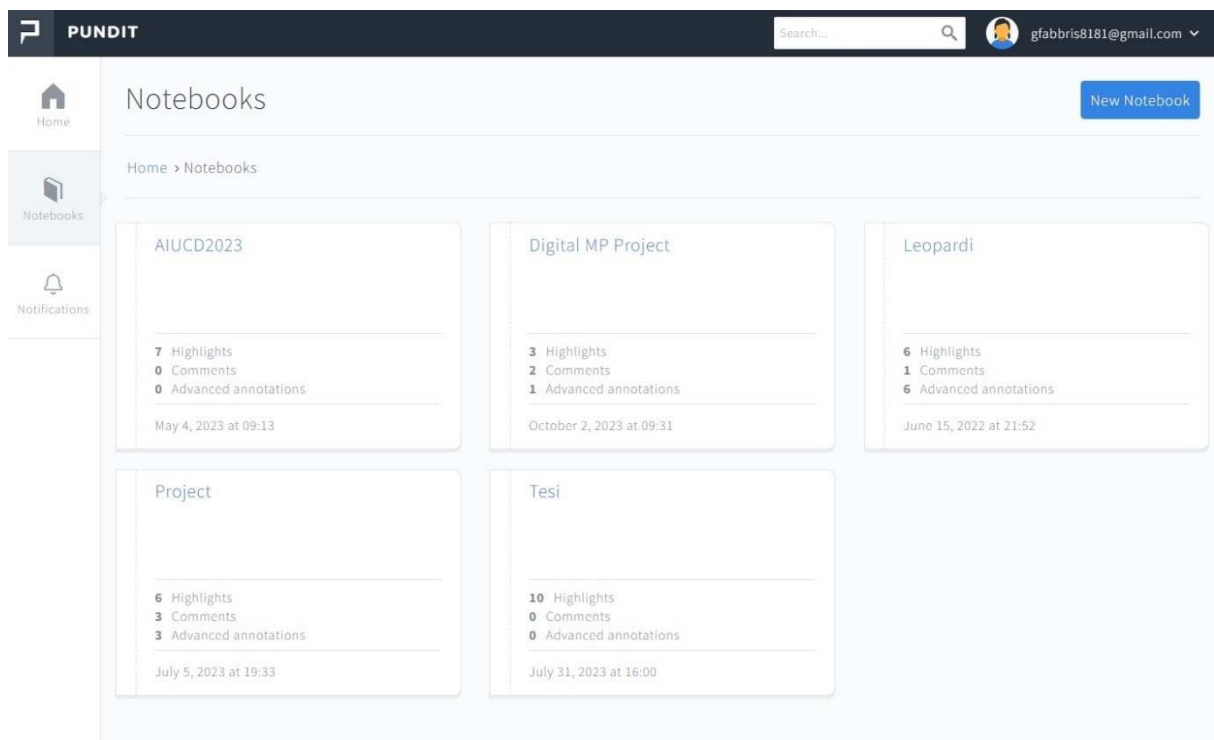


Figure 6 Overview of the Pundit dashboard.

Last but not least, Net7 is partner of the TRIPLE¹⁵⁸ project (Transforming Research through Innovative Practices for Linked Interdisciplinary Exploration), launched in 2019 and financed under the Horizon 2020 framework.¹⁵⁹ Pundit is the annotation tool on which the TRIPLE Open Annotation Tool is based and is one of the Innovation Services of the project.

3.2.4.1. Collaboration, enrichment and personalization of data

The integration of an annotation tool in CHIt was essential and one of the main concerns since the beginning of the project. Annotations are more and more prominent in the landscape of digital scholarship, since they foster collaborative engagement, enrichment of content and provide personalized experiences for users.

In this sense, Hodgson, Kalir, and Andrews (2023)'s definition of annotations is here borrowed, which are considered expressly social,

as indicated by readers who write and exchange their notes with one another, make meaning together, and use interactive media to construct knowledge about shared texts and contexts. More specifically, we borrow and build upon a definition from Novak and colleagues (2012) that defines social annotation (SA) as a type of learning technology enabling the addition of notes to digital and multimodal texts for the purposes of information sharing, peer interaction, knowledge construction, and collaborative meaning-making.¹⁶⁰

¹⁵⁸ <https://www.gotriple.eu/>

¹⁵⁹ Horizon 2020 was the biggest European Research and Innovation program which funding covered the period 2014-2020. It has been succeeded by the new Horizon Europe for the years 2021-2027 with a higher budget. https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe_en

¹⁶⁰ Hodgson, Kalir, and Andrews (2023)

Particularly in the scholarly field, collaboration is essential. Considering online platforms as mere repositories of information risks to be too simplistic and little productive. The addition of annotation tools enables users to actively engage, making them operative contributors instead of passive consumers. In this way, scholars but not only, can share insights, ask questions and get answers, and engage in dialogues that can build up new meanings because born from an innovative kind of information-making. This kind of interaction goes beyond the traditional silos of scholarly collaboration and can, hopefully, encourage a sense of belonging within a broader intellectual community. In this sense, cultural institutions play a major role since their position is shifting from content providers to experiences facilitators and, since these interactive tools are more and more common, users start to expect to find different types of experience engagement within physical and online GLAMs.¹⁶¹

The advantages stemming from annotation tools extend beyond collaboration and user interaction, comprising a spectrum of positive attributes that deeply enhance the digital experience. Narrowing down the field, annotation tools offer a nuanced approach for the enrichment and personalization of data, a dimension that can also be tailored to the individual user, with contrast to collaboration and interaction. This aspect of annotation tools is equally significant, as it introduces a layer of interactivity that is not solely dependent on communal engagement. The potential for personalized data enrichment emphasizes the various ways in which users can actively enhance their individual reading and visualization experiences.¹⁶² Consider, for example, the historical context of marginalia found in medieval manuscripts. Despite their nomenclature, which could suggest a peripheral role, they are annotations after all and have an enormous valor for the scholar since they carry information potential.

To conclude, Pundit's feature regarding semantic annotations deserves particular attention. In fact, the possibility to add semantic annotations into unstructured

¹⁶¹ Radice (2014)

¹⁶² Novak, Razzouk, and Johnson (2012); Kalir and Garcia (2021)

documents has become more and more relevant, since traditional search engines can be sometimes too simplistic as they are based on keywords that cannot go deep into the users' needs for meaning and context.¹⁶³ On the other hand, the quick development of the Semantic Web (SW) has brought forth the creation of formal representations of knowledge, like ontologies, which “are a key component of semantic association, which is the process to formalizing knowledge through the linking of words or phrases of plain text (mentions or named entities) with elements of the ontology (concepts or entities).”¹⁶⁴ In this way, users can semantically enrich their data in order to provide additional information and map a text with structured references linked to an ontology.

However, since its beginning in the late 90s, there has never been a real consensus about the SW's utility and fortune throughout time. Thanks to his survey, Hogan (2020) can discuss that twenty years on, important steps still have to be taken to align to the original version of the SW, even though these technologies are having some impact. Another interesting point of view is that presented by Lo Duca (Lo Duca 2021), where she argues that the SW has decreased progressively, but that Linked Data are trending instead, especially in the communities related to cultural heritage.

To sum up, annotations are more and more essential both under a collaborative point of view and also from an individual perspective. Whether through social interaction or personalized enrichment, annotation tools emerge as versatile instruments that contribute multifaceted benefits to the scholarly landscape and not only.

3.2.5. EVT: the visualization tool

This last Subsection presents the visualization tool, namely Edition Visualization Technology – EVT. The development of this tool started in 2003 when a new visualization software for a digital edition of the *Codex Vercellensis* (Vercelli, Archive and Chapter

¹⁶³ Pech et al. (2017)

¹⁶⁴ *Ibid.*

Library, MS CXVII) started. The team working on the project found that many of the existing tools that would have fit their needs had two or more flaws among being proprietary, closely linked to a specific software platform, too complex to install and use and too tied to specific projects.¹⁶⁵ For this reason, even if it is always recommended to exploit existing tools as much as possible, the team started to conceive a new software respecting some requirements that have guaranteed EVT an incredible success within the Italian environment of digital philology. In fact, nowadays it is used in a large number of projects and its flexibility proved essential. The first version of the software was developed specifically for the Digital Vercelli Book project, but the increasing number of requests for modification made the code too long and complex. Version 1 was “based on an XSLT 2.0 transformation chain that takes as its input one or more XML documents in the standard TEI P5 format, which are turned into a Web-based application — a mix of HTML5, CSS3 and JavaScript — that can be easily shared on the Web.”¹⁶⁶ However, the code became too verbose and difficult to understand, and the general layout too complex because of the many GUI components. For this reason, the team started to develop version 2 of the software within a completely different framework, able to support complex functionalities according to scholars’ needs. EVT 2 is built in AngularJS¹⁶⁷ and more specifically on a set of JavaScript parsers that retrieve the contents of the edition directly from the XML document(s). The retrieved information is temporarily saved in a JSON structure, which remains available until the web app is open. The style of the edition is regulated by CSS style-sheets that can be customized. However, Google stopped the development of AngularJS because of its many efficiency problems. A new Google framework was developed, namely Angular.¹⁶⁸ The EVT team decided to abandon AngularJS too and readapted the code to the new framework.¹⁶⁹

¹⁶⁵ Rosselli Del Turco (2019)

¹⁶⁶ *Ibid.* 94

¹⁶⁷ <https://angularjs.org/>

¹⁶⁸ <https://angular.dev/>

¹⁶⁹ Bioglio et al. (2023), but see also the video lesson held by Prof. Rosselli Del Turco in Pisa in 2023 in which he sums up the history of EVT with a particular focus on the new functionalities of version 3.

The schema is basically the same as EVT 2: the documents are handled by an input parser, converted into JSON and handled again by an output parser which sends information to the viewer.

EVT was chosen as visualization tool for CHIt because it has a huge pool of users, being tested many times and always proved successful. Moreover, it boasts an active group of developers and collaborators that continuously work on it in order to improve the code and integrate new features.

However, the standard version of EVT works with files stored locally. This modality does not fit CHIt architecture, as the files that EVT is supposed to read (e.g., XML ones) can be retrieved by CHIt via API. For this reason, within this framework, EVT needs to be able to catch the documents on the fly. Fortunately, this specific characteristic was already in the agenda and it has been recently implemented by another collaborating company.¹⁷⁰

The connection between CHIt and EVT is further explained in section 4.3.

Available at <https://www.labcd.unipi.it/seminari/roberto-rosselli-del-turco-levoluzione-di-evt-dieci-anni-di-sviluppo-open-source/>

¹⁷⁰ The company is called Silent Wave, <https://www.silentwave.eu/en/>.

4. Structure and implementation of the middleware

This chapter has been mentioned different times so far, and this proves its centrality. As it is clear now, the middleware is the cornerstone of the whole web application and this chapter is entirely dedicated to its implementation. The efforts spent during this doctoral project have been put mainly in this component, even if, of course, the backend and the frontend have required attention as well. These two parts of the architecture have been managed mainly by the collaborators at Net7, while the PhD candidate has worked extensively on the development of the middleware – always under the watching eye of the company tutor. The periods spent at the company as well as the work on the code done remotely have proven the usefulness of this collaboration, allowing the candidate – who works on Germanic and digital philology – to acquire a wide range of technical skills and critical thought on this transversal field. Therefore, considering the result obtained in this specific project, it can be said the decision taken during the rearrangement of the PON program of introducing these mandatory collaborations has proven absolutely effective and advantageous for the candidate and for the development of a high-quality project.

The chapter begins with a review of the major issues that have been faced since the initial phase of conception of the aggregator. Of course, every problem is accompanied by a solution, and most often the optimal one for the specific case could be found.

After that follows the explanation of how the source code was implemented. Technical details will be given for every provider, for the mapping of the metadata¹⁷¹ and how the new output model is created and sent to the frontend, for the creation of records and

¹⁷¹ In this chapter, the term *data* is used in a broad sense to refer to the information retrieved from the APIs of the various data providers. On the other hand, the term *metadata* specifically refers to the structured descriptions of resources provided by those sources. While the distinction may not always be rigidly maintained, this usage reflects the functional roles these elements play within the technical context discussed.

for how the resources can be searched. Lastly, the description of how the two external tools – EVT and Pundit – can interact with the resources is provided.

4.1. Problems and possible solutions

In the Introduction to this dissertation, it has been declared that the initial idea for this project was the development of two different applications for the visualization and the annotation of the resources. This vision was quickly reconsidered, thanks to the proposal of the company tutor, Doctor Aiola Chiara, who suggested instead to opt for a single but flexible application. One of the main characteristics that this application was supposed to have, in order to face our needs, was the possibility to connect external tools to accomplish the main functionalities of visualizing and annotating. It is well known at this point of the dissertation that the Muruca architecture perfectly fits this description. Moreover, the collaboration with Net7 ensures the availability of technical assistance for every kind of problem. After this decision, a single application with integrated tools became the choice and the adoption of Muruca as framework came along.

The next problem that arose was the identification of the providers and the selection of the resources. A complete discussion on this topic and the solution adopted has already been exposed in sections 2.3, given its centrality for the development of the project. It is not going to be tackled upon here again, so please refer to the aforementioned section for reference. However, a brief concept is recalled here for the harmonization with the next problem that is about to be expressed. The original idea foresaw the valorization of small projects, but this soon proved impossible for the lack of FAIRness of the portals that were at first considered and visited for the aggregation. Therefore, the project started from the retrieval of resources from a large provider, Europeana. From there, a bulk import through APIs could be done.¹⁷² Thanks to this operation, the platform

¹⁷² More details on these data requests will be given in the dedicated section below in this chapter (4.2.1.1).

gathered the majority of its resources, more than 200.000. However, it was soon clear that this massive retrieval would have brought along a certain level of uncertainty as far as the quality of the resources is concerned, since it is not possible to further filter this selection. Nevertheless, users always have the possibility to refer to the original source that provided the data to Europeana directly from the list of occurrences visualized in the CHIt record, and in this way, they can verify the accuracy of the resources themselves. On the other hand, the methodology adopted for the retrieval from the other providers allows for a more reasoned ingestion. Even if it is clear that manual selection from heterogeneous repositories is not the most recommended way to populate a cultural aggregator, this path was pursued to stay faithful to the initial idea and give some space to the smaller projects. This choice was also taken to add scientifically valuable resources in order to try to balance the potential uncertainty of what was imported via Europeana. Since CHIt is thought to be a tool useful both for a general audience but also for a specialized one, this choice has proven fundamental to ensure possible scholarly research. An example will be shown in section 6.2.3. Moreover, this *ad hoc* insertion can also potentially stimulate users to suggest resources that have not been noticed during the implementation, further enhancing their active participation. Again, young or independent scholars can also exploit CHIt for the publication of their own research to add them to the ecosystem and allow for interoperability and interconnection with the other objects of the platform. This would be the upgrade of what has been done so far with GitHub and Zenodo: to give the users the possibility to valorize their own researches and offer an environment where cultural data can be aggregated and come into contact in a way that would not have otherwise happened. Users will not be able to add directly new resources, but they are strongly encouraged to write an email from the *Contact* page for any suggestion and comment or for highlighting bugs and errors in general.

After some time that the development of the code had begun, there has been a crucial change that is to some extent linked to the last issue discussed above. The first stage of the project involved the aggregation and the indexing of all the metadata through Elasticsearch, used as a sort of “cache”, from which the middleware was supposed to

fetch the resources requested by the frontend. However, even if the use of a cache system has its pros, the cons have to be considered as well. In particular, this layer has to be updated constantly to ensure that the data it stores are aligned with the original ones. Therefore, there is the risk of storing obsolete data if they are not regularly checked. This is time-consuming, especially when a huge number of resources is gathered and there is not a team or part of a team assigned to this task. Therefore, it was decided to delete this passage at least for the providers that allow to retrieve data directly through APIs as Europeana, or through a SPARQL query as in the case of OntoVE. These resources are not saved in the cache but are requested by the middleware directly to the providers themselves. In this way, the storage in the backend and therefore the use of its memory is unloaded; moreover, users are sure to find the most updated data since they are gathered each time from the original source without mediation.

On the other hand, the workflow is considerably different for the miscellaneous repositories. In fact, the resources identified in GitHub and Zenodo are supposed to be requested one by one with single APIs that, in the case of GitHub, do not even give back the metadata for a general description or the link to the files – unless two or more API requests are sent, which complicates things even more. Therefore, it was necessary to find a different solution to improve the efficiency of this operation. As a result, the system adopted for these providers was mixed in order to ensure the retrieval of all the necessary metadata and improve the alignment with the other providers. Muruca backend was used to collect the references to the single resources from Zenodo and GitHub that had to be included in CHIt. The backend stores just the metadata needed by the platform for the general search and the references (unique identifiers, API references) to get the information from the original source at any time. In this way, these resources are indexed in Elasticsearch in order to be searchable in the search engine. On the other hand, once a single specific resource is required, the data are retrieved via API also in the case of Zenodo, so that what is received is always the latest version. For GitHub, the question is different. In this case, all the information – files included – is inserted manually since the available APIs did not suit the needs of this platform. In this sense, the findability principle is satisfied, allowing users to find some relevant projects

and resources possibly hidden in great and heterogeneous repositories like GitHub. However, interoperability misses, since metadata could not be accessed directly via API, but this was considered a good compromise especially because the platform always gives the direct link to the original source of the resource.

The different approaches adopted for the two types of retrieval have a major consequence, though. The resources gathered from Europeana and OntoVE are totally dependent on their availability on the original platforms, as they are called each time via API. This means that if Europeana goes down or changes its APIs, CHIt loses the access to the resources. On the other hand, the metadata stored in the CHIt backend are there and are independent from the original source. This is a compromise that has been considered and something more will be said about it when dealing with scalability and flexibility in the last paragraph of this section related to problems.

This next paragraph introduces two interesting issues that will be dealt with in more detail in the dedicated sections of this chapter, but are cited here since they posed some interesting challenges. The connection to the OntoVE Endpoint was born as the result of a collaboration between this doctoral project and another research project in development at Ca' Foscari. To start with, the idea was born because this ontology describes and preserves digital objects of the Germanic cultural heritage in the Veneto region. From the point of view of the content, the target of resources is perfectly in line with what CHIt aggregates. However, as it will be explained below, the methodology is totally different, yet the goals are shared. It is a high-quality standalone project, nonetheless integrated within the Semantic Web thanks to the use of internationally accepted standards.

Before it was decided to experiment with the connection to this Endpoint, only APIs had been taken into consideration for the retrieval of resources. Therefore, it was necessary to understand how SPARQL queries could be performed from the middleware and what possibilities were to be offered to users. In the end, it was decided to use a predefined set of *ad hoc* SPARQL queries to extract the items of interest in a way compliant with the code of the middleware, and users can access these resources as they can do with

the other providers. Users versed in the query language that need to find different or more specific information can refer to the original Endpoint through the specific link to the ontology that can be found in the section of the application dedicated to OntoVE (*Home > Il Progetto > I provider > OntoVE*, *Home > The Project > The Providers > OntoVE*).¹⁷³ Another tool that has been integrated is EVT, the visualization technology that allows to see images and XML files. The developers had to implement a separate release because this tool has been conceived since the beginning as visualization software for static pages that retrieves the necessary files from its own directory, where they are stored. However, CHIt does not store any file in its backend, only metadata. Therefore, it had to be adapted as to be able to read files dynamically.

One last due remark that is required at the end of this overview concerning different issues is a generic comment on the concept of scalability and flexibility, which are two closely linked concepts. The first one is the ability to handle increasing loads of data without compromising performance or user experience. On the other hand, flexibility allows to easily adapt to changes and needs without significant rework. These concepts have already appeared in previous chapters, but they have been tackled upon from a generic perspective and in generic descriptions. It has been said that scalability and flexibility are important in the development of cultural aggregators because they allow to adapt to the increasing demand deriving from the aggregation of data from different providers, each of them with their different systems (chapters 2 and 3). The technologies adopted for this project are all in line with these concepts and would allow the development of scalable and flexible tools. However, at the present state of affairs, CHIt can be considered as only flexible, while scalability has been sacrificed, considered the nature of the project. Scalability is not easy to achieve, it demands great efforts on the programming side and is essential for larger projects. Among the other things, it requires the harvesting of data in order to ingest and store them on a database. In this way, as it already happens in CHIt for Zenodo and GitHub, data are available locally and there is no need to send requests to the providers every time a user asks for a resource. It has

¹⁷³ As already mentioned in note 96, the access to the Endpoint is limited for now.

been said above that this mechanism risks creating obsolete data that need to be periodically updated manually. However, when a platform deals with huge numbers of resources, this is the optimal solution and a part of the team will be assigned this task. An aggregator built on a large scale that is daily used by many users, needs to be independent from continuous calls to the providers, which ensure loss of time in response and decrease in performance and user experience.

CHIt does not really need to be scalable at the moment, as the retrieval of resources is not so vast as it could be for an aggregator developed, for example, at a regional level, and the amount of users that are likely to use the platform at the same time is esteemed to be rather contained. This is the outcome of a doctoral project that had to be carried out in three years by a team composed of basically two/three people, it would have been unrealistic to consider harvesting, with all the consequences that it implies. In CHIt, there is no need to handle enormous amounts of data, and this is why the initial idea of storing everything on Elasticsearch has changed. Even so, the retrieval of data in CHIt still takes some time because it has to ask for the resources to Europeana and OntoVE, which together give back at any rate some thousands results. However, little latency has been preferred to the risk of creating a platform that is populated by obsolete resources. To sum up, it can be said that CHIt is an experimental platform which explored the crucial aspects of the development of a harvesting platform, in line with the goals of research that a doctoral project should pursue; nonetheless it is performing and it has been developed considering many aspects for its present and future functioning. The last chapter will be dedicated to the exploration of possible further developments.

4.2. Implementation of the source code

This part of chapter 4 contains the heart of what has been implemented for this PhD project. Its core component is section 4.2.1., which deals with technical precise descriptions of the various providers, how metadata are imported, formatted and sent to the frontend for the consultation. The precise output schema for each of them is

presented and examples of visualization are provided. The creation of narrative paths will be tackled upon as well, as it was previously said that these kinds of operations contribute to the valorization of the resources and ensure a better usage for users, who are guided and supported into the navigation.

Before delving into these dense aspects, it is necessary to explain how the source code was organized to import the providers' data.

The main folders are contained within `src`, and are `controllers`, `helpers` and `mappings`:

- the folder `controllers` includes the most important files, `getResource.controller.ts` and `searchController.ts`, that contain the functions for the retrieval of the resources from the providers and their return to the CHIt frontend;
- the folder `helpers` contains the different providers' connectors (i.e., `europenaConnector.ts`, `sparqlConnector.ts`, `zenodoConnector.ts` and `elasticConnector.ts`) which fetch the resources, parse them and give them back harmonized according to the CHIt metadata model;
- the folder `mappings` contains, of course, the single mappings of the providers: `europenaMappings.ts`, `githubMappings.ts`, `sparqlMappings.ts` and `zenodoMappings.ts`.

There are two main processes that take place and that involve the interaction among these files and their functions. The first one is the massive retrieval of all the resources at once, and it is handled by the `searchController.ts` file, which does not work alone. Thanks to the various connectors (in the `helpers` folder above), it sends all the requests to the providers and the backend to fetch the available resources. After that, the resources are mapped (`mappings` folder above), gathered and adapted to the visualization for CHIt. At this point, they are ready to be sent to the frontend for the

fruition. The other fundamental process is the retrieval of the single resources requested specifically by the user. In this case, the operation starts from the `getResource.controller.ts`, which decodes the incoming request to understand what the provider is thanks to its identifier. In fact, each resource has a specific and unique ID, and this characteristic is crucial for the middleware to understand its origin. Each ID is made of a prefix, assigned by the middleware which identifies the source of the resource, and a unique numeric or alphanumeric string coming from the data provider. For example, the link https://chit.chitontove.it/record/zen_10354027/castelseprio-lapide-longobarda shows that the item *Castelseprio Lapide Longobarda* is taken from Zenodo (`zen`) and its numeric string (e.g., the original ID that Zenodo has assigned to the resource) is 10354027. The two pieces of information are put together by the middleware for Europeana and OntoVE and the ID that CHIt needs to retrieve the resource is created. For Zenodo it is added manually to the backend, while for GitHub and the other resources added manually, the numeric part is automatically generated by Muruca and only the prefix is given to the backend. The prefixes are `eu` for Europeana, `zen` for Zenodo, `github` for GitHub, `sparql` for OntoVE and `mrc` (e.g., Muruca) for the resources added to the backend.

Thanks to the connectors, once again, it fetches the single resource, maps it and gives it back to the frontend for the visualization.

For the sake of completeness, it is worth mentioning some tools that have been used by the team during the development of the project. The code is written with Visual Studio Code,¹⁷⁴ a free, lightweight and powerful code editor for Windows, MacOS and Linux. APIs have been tested with Postman,¹⁷⁵ a widespread platform that allows to test, save and organize API calls. The source code was shared among the collaborators thanks to GitLab,¹⁷⁶ an open-source software for the management of Git repositories.

¹⁷⁴ <https://code.visualstudio.com/>

¹⁷⁵ <https://www.postman.com/>

¹⁷⁶ <https://about.gitlab.com/>

4.2.1. Importing data: the providers

The following subsections offer a technical description of the different providers and of the work that has been conducted on each of them.

As mentioned above, the resources are fetched via Europeana, GitHub, Zenodo and OntoVE and the modalities of retrieval are different according to the interoperable services that they offer. Basically, for the first two platforms, their APIs could be exploited, even if to a different extent. In fact, it has already been explained that the resources retrieved from Europeana are requested directly to its database, for Zenodo some information is manually stored in the Muruca backend and other is retrieved via API, while for GitHub the full storage in the backend is the solution adopted. On the other hand, OntoVE is a totally different tool, since its data are stored in a SPARQL Endpoint. As such, it was more appropriate to write a SPARQL query to gather specific resources. However, the process is more akin to that of Europeana, as there is no storage in the aggregator's database for the ontology and the items are recovered directly from the Endpoint.

After this operation, the resources are mapped according to a predefined schema that has been implemented adhering to the DCMI terms. For OntoVE, additional standards have been considered for certain entries due to the project's specific requirements and the limited metadata provided by the DCMI set. Specifically, location references and bibliographic entries have been aligned with CIDOC-CRM version 7.1.2,¹⁷⁷ the ArCo ontologies,¹⁷⁸ and Bibo.¹⁷⁹ For further details on these labels, please refer to the lower part of table 4, and for a comprehensive description of the metadata set related to OntoVE, see section 4.2.1.4. In this way, all the incoming data are uniform and can interoperate with each other. In so doing, not only is efficiency ensured, but also user

¹⁷⁷ https://cidoc-crm.org/html/cidoc_crm_v7.1.2.html

¹⁷⁸ <http://wit.istc.cnr.it/arco>

¹⁷⁹ <https://www.dublincore.org/specifications/bibo/>

experience benefits from it because thanks to the mapping the results in the CHIt frontend are organized and more pleasant. This has been one of the first decisions, as it soon became evident that the metadata of the incoming resources were rather different. Just to make an example, one of the most important metadata is the one that allows users to access the original source of the resource they are visualizing. For example, from the interaction with a resource gathered from Europeana, the user should be able to go directly to the original provider – say, CulturalItalia. The different providers refer to this external link with four different metadata: `edmIsShownAt` for Europeana, `url` for GitHub, `doi` for Zenodo and `repoLink` for OntoVE. In the mapping that takes place within the CHIt middleware, they are all uniformed to `requires`, which identifies “[a] related resource that is required by the described resource to support its function, delivery, or coherence.”¹⁸⁰ This definition, even if not perfectly fitting, is the closest one among the DCMI terms to what the external link refers to. It is not a related resource since it is the same resource, but according to the principles that guided the implementation of the platform, it is required to support its delivery. Last but not least, to each DCMI term a new label is assigned for the visualization in the frontend to ensure a clearer understanding.

In the following table, the selected metadata are shown, together with their label.

DCMI	DCMI reference and description	Label frontend
description	http://purl.org/dc/terms/description “An account of the resource.”	Descrizione
subject	http://purl.org/dc/terms/subject “A topic of the resource.”	Soggetti
language	http://purl.org/dc/terms/language	Lingua

¹⁸⁰ <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/#http://purl.org/dc/terms/requires>

	"A language of the resource."	
source	http://purl.org/dc/terms/source "A related resource from which the described resource is derived." ¹⁸¹	Istituto o autore
type	http://purl.org/dc/terms/type "The nature or genre of the resource."	Tipo
publisher	http://purl.org/dc/terms/publisher "An entity responsible for making the resource available."	Provider
mediator ¹⁸²	http://purl.org/dc/terms/mediator "An entity that mediates access to the resource."	Fornitore
rights	http://purl.org/dc/terms/rights "Information about rights held in and over the resource."	Diritti
relation	http://purl.org/dc/terms/relation "A related resource."	File disponibili
requires	http://purl.org/dc/terms/requires "A related resource that is required by the described resource to support its function, delivery, or coherence."	Collegament o alla fonte originale
title	http://purl.org/dc/terms/title "A name given to the resource."	Titolo

¹⁸¹ As the example illustrated in the body of the text above, in this case as well the choice is not perfect, but yet the most suitable one among the available terms.

¹⁸² The difference between *publisher* and *mediator* was necessary to disambiguate some information available only for Europeana, but see the dedicated section below.

identifier	http://purl.org/dc/terms/identifier “An unambiguous reference to the resource within a given context.”	/
spatial	http://purl.org/dc/terms/spatial “Spatial characteristics of the resource.”	Coordinate geografiche
bibliographicCitation	http://purl.org/dc/terms/bibliographicCitation “A bibliographic reference for the resource.”	Item bibliography
CIDOC	CIDOC reference and description	Label frontend
cidoc-crm:P54_has_current_permanent_location	https://cidoc-crm.org/html/cidoc_crm_v7.1.2.html#P54 “This property records the foreseen permanent location of an instance of E19 Physical Object at the time of validity of the record or database containing the statement that uses this property.”	Posizione
cidoc-crm:P55_has_current_location	http://cidoc-crm.org/cidoc-crm/7.1.2/P55_has_current_location “This property records the location of an E19 Physical Object at the time of validity of the record or database containing the statement that uses this property.”	
ArCo	ArCo reference and description	Label frontend
arco:ArchaeologicalProperty	https://dati.beniculturali.it/lodview-arco-onto/ontology/arco/ArchaeologicalProperty.html “This class represents an archaeological cultural property.”	Bene culturale archeologico

context- description:completeBibliographicReference	https://dati.beniculturali.it/lodview-arco-onto/ontology/context-description/completeBibliographicReference.html "This property represents the complete bibliographic reference."	Bibliografia completa
Bibo	Bibo reference	Label frontend
bibo:handle	https://dcmi.github.io/bibo/#:handle	Bibliografia completa ¹⁸³

Table 4 DCMI terms adopted for the CHIt output model

This table represents the whole set of metadata that have been mapped in the four providers but, of course, not all of them are present in each of them and not all of them are necessarily evident in the frontend. In fact, the provider and the ID are not shown in the list of metadata, but can be retrieved by looking at the web address, since it is an information that is required for the visualization of the resource but from the point of view of the retrieval (see above, section 4.2).

However, a set of these metadata only belongs to OntoVE. In fact, the specific field for which this ontology is developed requires a set of information that is not needed to the resources gathered from the other providers and, as a matter of fact, not available.

Now that a general description has been provided, the single providers are listed in detail below.

¹⁸³ *bibo:handle* can refer to any handle, but in the mapping through OntoVE it is used to retrieve the bibliography of a resource available online.

4.2.1.1. Europeana

The overview starts with the provider that was tested as the first one and that gave the first satisfactory results, and needs no introduction. As already mentioned, Europeana was chosen after a sort of change of direction in the conception of the project, since there has been a phase of disappointment when the team first tried to retrieve resources from smaller projects. After that, it was decided to start testing the system with a bigger provider, and Europeana was considered almost immediately. It confirmed to be a good solution, as it proved that the workflow actually worked and was promising. Therefore, the first phase of massive resource retrieval began with a careful selection of the different filters. In particular, it was decided to exploit the already existing collections to narrow down the pool from which to draw resources. The selected collections are: art, archaeology, manuscripts, maps and geography and world war 1. Europeana has available a good documentation of its API, so after the kind of queries to be performed on the providers identified was tested,¹⁸⁴ a connector to Europeana API was developed. The Europeana content providers are: CulturalItalia, Carare, 3D Icons, European Library of Information and Culture, Museu, European Library of Information and Culture, Central Institute for the Union Catalogue of Italian Libraries, The European Library and New Manuscript Library, Veneto Region. All of them, except for the last one, are *providers*, while the New Manuscript Library is a *data provider*. According to the Europeana Organization Profile,¹⁸⁵ *providers* are organizations that deliver data directly to Europeana, while *data providers* can submit data directly but more likely via the services of a provider. In our case, the New Manuscript Library delivers data directly. This difference is kept in the metadatation of the CHIt output model but with a conceptual difference, to adapt to the available DCMI terms. The data provider is labeled as *publisher*, and the provider as *mediator*. Tables 4 above and 6 below account for this

¹⁸⁴ Of course, in this phase the reference to *providers* is to the ones of Europeana, not of CHIt.

¹⁸⁵

https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_profiles/EDM_Organization_Profile_042016.pdf

division. The reason for this choice is that when the data providers are present, they are the actual providers of the data, and they use providers to *mediate* the delivery of their data.

From the Europeana Pro web page,¹⁸⁶ it is possible to explore all the documentation related to the use of APIs. Europeana offers REST (search, record, entity, annotations), IIIF and SPARQL APIs, LOD and methods for downloading and harvesting metadata for purposes that require scalability.

For the scope of this project, the search API was used, which allows to search for metadata records and media on the Europeana repository, similarly to how the Europeana website does. An API key is required, which can be easily obtained, and there is a console with a practical interface that permits to test your APIs.

It was said above that APIs are HTTPS requests, and as such they need a URL signature. The one required by Europeana is

<https://api.europeana.eu/record/v2/search.json>

In the dedicated web page, all the possible parameters are enlisted and described and the EDM data model is available with detailed descriptions of each metadata.

In table 5 below, the six APIs are provided. What is provided are the address, a mandatory field query which in this case is completed with a wildcard * since the retrieval is not meant to narrow down by keyword as the following field, themes, already functions as filter. After that, in a qf field (query refinement) the provider(s) is/are given and, where necessary, the country as well. Last but not least, the API key is provided. The query refinement field can be repeated as many times as necessary, and the syntax is built with a keyword that identifies what the refinement is about (in our case *(data_)provider* and *country*) and the filter within inverted commas after a colon.

¹⁸⁶ <https://pro.europeana.eu/>

[https://api.europeana.eu/record/v2/search.json/?query=*&theme=manuscript&qf=PROVIDER: "CulturaItalia"&wskey=xxxx](https://api.europeana.eu/record/v2/search.json/?query=*&theme=manuscript&qf=PROVIDER:)

[https://api.europeana.eu/record/v2/search.json/?query=*&theme=art&qf=PROVIDER: "CulturaItalia"&qf=PROVIDER: "CARARE"&qf=COUNTRY: "Italy"&wskey=xxxx](https://api.europeana.eu/record/v2/search.json/?query=*&theme=art&qf=PROVIDER:)

[https://api.europeana.eu/record/v2/search.json/?query=*&theme=archaeology&qf=PROVIDER: "European Library of Information and Culture"&qf=PROVIDER: "3D ICONS"&qf=PROVIDER: "MUSEU"&wskey=xxxx](https://api.europeana.eu/record/v2/search.json/?query=*&theme=archaeology&qf=PROVIDER:)

[https://api.europeana.eu/record/v2/search.json/?query=*&theme=ww1&qf=PROVIDER: "CulturaItalia"&qf=PROVIDER: "Central Institute for the Union Catalogue of Italian Libraries"&qf=PROVIDER: "The European Library"&qf=COUNTRY: "Italy"&wskey=xxxx](https://api.europeana.eu/record/v2/search.json/?query=*&theme=ww1&qf=PROVIDER:)

[https://api.europeana.eu/record/v2/search.json/?query=*&theme=map&qf=PROVIDER: "CulturaItalia"&qf=PROVIDER: "European Library of Information and Culture"&wskey=xxxx](https://api.europeana.eu/record/v2/search.json/?query=*&theme=map&qf=PROVIDER:)

[https://api.europeana.eu/record/v2/search.json/?query=*&qf=DATA PROV
IDER: "New Manuscript Library, Veneto Region"&wskey=xxxx](https://api.europeana.eu/record/v2/search.json/?query=*&qf=DATA PROV
IDER:)¹⁸⁷

Table 5 Europeana APIs

The query in the middleware is built in the `europeanaConnector.ts` file.

Once the request is sent, a JSON response is received with the list of the resources that were available through that call. These results are used by CHIt to create the search interface, of which a screenshot is provided in figure 7. In this phase, a short list of metadata is selected in order to show a preview of some essential information related to the resources. As it can be seen in the picture below, a preview of the visual content is given, together with the title, the original provider and the keywords associated.

¹⁸⁷ As mentioned above, this is a data provider and has a different label according to the EDM data model. Moreover, since by its nature it contains only manuscripts, it was not necessary to specify the theme.

Risorse

Filtra i risultati 226550 Risorse Ordine Ordine alfabetico (A-Z)

Cerca nel titolo

Filtri attivi: **Europeana**

Provider

- Ontove
- GitHub
- Other
- Zenodo
- ☒ Europeana

Autori

Temi

- filologia digitale
- medioevo
- filologia germanica
- lingue minoritarie
- Codifica XML-TEI
- early middle ages

Epoca

Luogo d'origine



"A la fiera de Mast'Andrea" | Madrigale Napoletano- con solo di soprano | con accompagnamento d'orchestra | ottavino | flauto | 2 corni | 2 trombe | Arpa | Camp.Ili | batteria | e Quintetto a corda | Renato Parodi
Provenance: Library of the Cherubini Conservatory of Music
Subjects: Printed music, Manuscripts



"Anch'io a casa ho una bimba come te"
Provenance: Central Institute for the Union Catalogue of Italian Libraries
Subjects: World War, 1914-1918, World War, 1914-1918 -- Art and the war, World War I, Photograph, War and society



Animale carnivoro forse della famiglia dei mustelidi
Provenance: University Library Of Pisa
Subjects: Drawing, Egyptology, Egyptian Studies, drawings (visual works), field archaeology



/ ANO // Amstelodami, Sumptibus Henrici Hondii)
Provenance: Marciana National Library
Subjects: Maps



"*A voi, fanciulli...." : (per il fronte interno) / Rachele Ferrari
Provenance: National Library of Rome - Italy
Subjects: World War, 1914-1918, Book, World War I, World War, 1914-1918 -- Propaganda



"Batteria Amalfi. Impianto binato da 381/40"
Provenance: Central Institute for the Union Catalogue of Italian Libraries



"Bosco di Gallio", trincee di resistenza
Provenance: Central Institute for the Union Catalogue of Italian Libraries



"Caproni" abbattuto nei pressi di Pola
Provenance: Central Institute for the Union Catalogue of Italian Libraries

Figure 7 Resources retrieved by Europeana and visualized on the CHIt frontend

So far, the massive retrieval of resources had been implemented and the code developed until this point starts to populate CHIt frontend. However, once the user selects a single resource, a different request is sent to the middleware, and a different response is given back. The following step that is described now is crucial for the selection of the single resource. From the search interface presented above, the user can choose the desired item and send another request from the frontend to the middleware. In turn, the middleware requests the item to Europeana through a different

API call. It is obviously impossible to enlist all of these kinds of calls performed by CHIt in this phase, since there are thousands of them – as many as all the gathered resources are. The URL is the following:

[https://api.europeana.eu/record/v2/\[COLLECTION_ID\]/\[RECORD_ID\].\[FORMAT\]](https://api.europeana.eu/record/v2/[COLLECTION_ID]/[RECORD_ID].[FORMAT]) and, taking as example the first item of the previous image, a practical example would be as follows:

<https://api.europeana.eu/record/v2/930/RMSE073930.json?wskey=apikey>, where 930 is the identifier of the collection and RMSE073930 is the identifier of the single record, requested in JSON format. In this way, the middleware has access to the whole set of metadata associated to the requested item. At this point, the much-cited mapping takes place and from the input model acquired by Europeana, the output model is generated. In particular, the metadata taken and transformed for Europeana are resumed in table 6. The input model is not provided here, as it would be too long given the enormous number of metadata originally available. This selection clears the result and selects the most relevant ones.

Europeana input model CHIt output model

dcDescription	description
dcSubject	subject
dcLanguage	language
dcSource	source
edmType	type
edmDataProvider	publisher
edmProvider	mediator
edmRights	rights
About	relation

edmIsShownAt	requires
dcTitle	title

Table 6 Europeana input model into CHIt output model

As it is evident comparing tables 4 and 6, the output model was modeled upon the Europeana schema since it provides the most complete one among the four providers of CHIt. As it will be seen below, the other three do not fill in all of the rows.

After this process of mapping, the resource is sent back to the frontend with its new guise and results in the visualization of figure 8. An image-viewer is embedded in the page, so that when images are available, they can be visualized directly from the CHIt window. When the resource is a PDF, it can be easily opened in the web browser of the user and when the resource is an XML file it will be visualized with EVT.

"Casotto Alto," Val d'Astico, "Casotto Alto", Val d'Astico



Metadati

Soggetti	World War I, World War, 1914-1918, Photograph
Organizzazioni	Central Institute for the Union Catalogue of Italian Libraries, The European Library
Istituzione o autore	Museo della Guardia di Finanza
Tipo	IMAGE
Fornitore	http://data.europeana.eu/organization/1482250000001267025
Provider	http://data.europeana.eu/organization/1482250000004516062
Diritti	http://rightsstatements.org/vocab/InC/1.0/
Collegamento alla fonte originale	http://www.14-18.it/foto/MUSEOGdf_01_022B

Figure 8 Visualization of Europeana single resource in CHIt and integrated image-viewer

Of course, the metadata visualized for each resource depend on their availability in the original source. For instance, the example above shows that this item does not contain keywords, and the field is missing in the output model.

From here, users can now access the related files and annotate the resources according to their needs, but this will be the topic of a dedicated section where an in-depth analysis of the two integrated tools will be given.

4.2.1.2. Zenodo

The second provider that was taken into consideration was Zenodo, since by its nature it contains interesting data related to the academic field. At first, an approach akin to that experimented with Europeana was applied and good results were expected. The expectation was to retrieve a smaller number of resources but of higher quality, given the fact that Zenodo is an open-access repository populated by researchers themselves. This does not mean that all the resources are reliable *a priori*, but that the average of scientific accuracy should be finest and that, not less important, it is easier to control each item manually. With the retrieval from Europeana, an individual check is impossible.

The first step was, of course, the exploration of the available APIs.¹⁸⁸ There are three possible actions: deposit (to upload and publish items), records (to search for published records) and files (to upload/download files). The base URL is <https://zenodo.org/api/> and, since the aim is to look for records, the [records/](#) field has to be added. Under the section *Entities > Records* there are two interesting subsections, namely *List* and *Retrieve*. The first experiments were done with the first one, as it was hoped to retrieve a good number of resources with the right API call. All the parameters are enlisted, but it soon became evident that the only useful one for the scope of this project was *q*, search query. Numerous tries were done, but the results were not satisfactory as the number of items retrieved was small and, moreover, the greatest majority were scholarly articles. This material is undoubtedly fundamental in the academic field, but is not pertinent with the CHIt platform. The research was narrowed down using the *type* parameter, so that it was possible to filter for dataset, image and similar, but this choice nullified all the results. For example, one of the tries was [https://zenodo.org/api/records?q=\"\(longobard*\)\"](https://zenodo.org/api/records?q=\), but the results were disappointing since the query string catches information from the whole set of metadata. Therefore, the results comprised records that deal with biodiversity,

¹⁸⁸ <https://developers.zenodo.org/>

geodiversity and astronomy, because authors, bibliographic references and other fields contained surnames or titles with *Longobard**. Moreover, trying to filter for dataset, gave no results. For the sake of completeness, also another solution was tried. The same research (*longboard* + dataset type*) was performed from the user interface, e.g., from the Zenodo web page. It turned out that the results were totally different and out of ten results, two were actually adequate for the CHIt target.

Once it had been ascertained that this kind of API was not suitable for the needs of the platform, a mixed modality of insertion was implemented, e.g. an API call plus manual insertion. On the one hand, the API call, in this specific case, requires the identifier of the record (provided by Zenodo) instead of the query. For example, the API call <https://zenodo.org/api/records/4745084> retrieves the record identified by the numeric string 4745084, which is *Codice diplomatico longobardo 1-2: the original charters and coeval copies written in Tuscany between AD 714 and 774* by Timo Korkiakangas, containing a TEI/XML edition of a Lombard charter. This call retrieves the information displayed on the single resource in CHIt, as shows figure 11 below.

On the other hand, the manual insertion is necessary to show the basic metadata (title and subjects) to be presented in the research page, as shown in the box in figure 9.

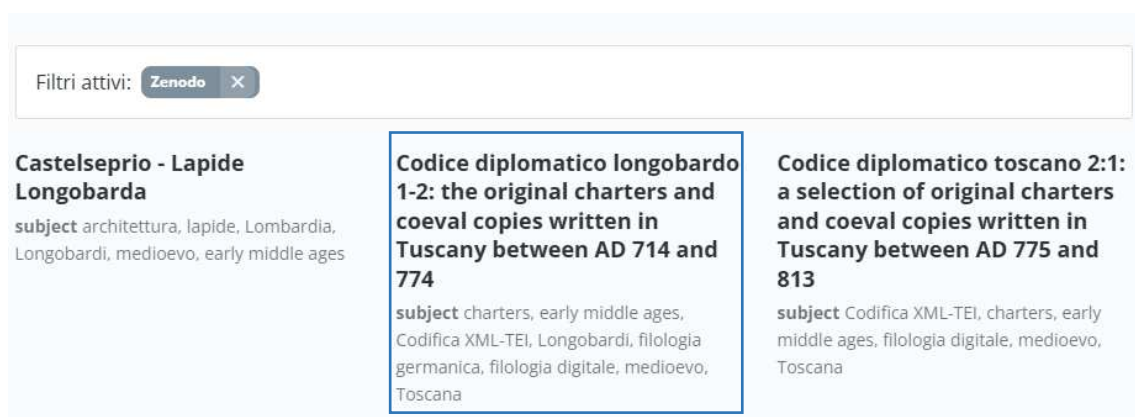


Figure 9 Presentation page with Zenodo resources

The ID is also added to allow CHIt to locate the resource, as explained in section 4.2.

This double retrieval (e.g., the metadata for the presentation page added manually plus the metadata for the visualization of the single resource retrieved via API) was necessary to quicken the whole process. In fact, for every single resource, a single API call is performed, unlike for Europeana, which has available a detailed set of filters that allow to retrieve, in this case, more than 200.000 items with only 6 API calls. Retrieving via API every single Zenodo resource for the presentation page would require a load of data that might slow down the loading of the page.¹⁸⁹

The following step is analogous to the one described for Europeana above. Once the user selects a specific resource, the middleware sends an API request to the appropriate provider, in this case Zenodo. The incoming resource is therefore mapped according to the CHIt output model. The selected metadata for Zenodo are resumed in table 7.

Zenodo input model

CHIt output model

description	description
keywords	subject
language	language
creators	source
resource_type	type
"Zenodo"	publisher
/	mediator
access_right	rights
links	relation
doi	requires

¹⁸⁹ At the present moment, this solution could still be manageable given the low number of resources. However, the ambition is to add more and more resources, and a more scalable process was adopted since the beginning.

title	title
-------	-------

Table 7 Zenodo input model into CHIt output model

The following figures show a record on Zenodo (Figure 10) and how the same item is rendered on CHIt after the mapping (Figure 11).

The screenshot shows the Zenodo interface for a dataset. At the top, there's a blue header with the Zenodo logo, a search bar, and links for 'Communities' and 'My dashboard'. Below the header, a light blue banner contains a message about a planned intervention on Tuesday, March 18th. The main content area has a dark blue background with the text 'Published May 10, 2021 | Version 1.0' and buttons for 'Dataset' and 'Open'. The title of the record is 'Codice diplomatico longobardo 1-2: the original charters and coeval copies written in Tuscany between AD 714 and 774'. The author is 'Korkiakangas, Timo'. Below the title, there's a paragraph of text describing the dataset as a TEI XML edition of original charters and coeval copies from the Codice diplomatico longobardo 1-2, edited by Luigi Schiaparelli. The text mentions that the digitized version comes from the site of the Institut für Mittelalterforschung of the Austrian Academy of Sciences and that expansions of abbreviations and indications of lacunae are preserved. It also mentions that the tags 'abbr' and 'gap' are used for unexpanded abbreviations and unwritten spaces respectively. The text concludes with a reference to a paper by Korkiakangas, T. and Passarotti, M. (2011) in the Journal of Language Technology and Computational Linguistics. Below the text, there's a section titled 'Files' which shows a list of files. The first file is 'CDL_TEI_XML_base_text_for_LLCT.xml'. The file content is displayed in a code editor, showing the beginning of a TEI XML document. The code includes a DOCTYPE declaration, a public ID, a base URI, and a series of entity declarations. The main content of the file is a <teiHeader> element with a <fileDesc> subelement. The <fileDesc> element contains a <titleStmnt> subelement with a <title> element containing the text 'Codice diplomatico longobardo 1-2' and a <sub> element containing the text 'The original charters and coeval copies written in Tuscany between AD 714 and 774'. The <titleStmnt> element also contains an <editor> element with the text 'Luigi Schiaparelli'. The <fileDesc> element also contains an <extent> element with the text 'about 787 kb', a <publicationStmnt> element, a <sourceDesc> element, and a <biblStruct> element. The <biblStruct> element contains a <monogr> subelement with a <title> element containing the text 'Codice diplomatico longobardo 1-2', an <editor> element with the text 'Luigi Schiaparelli', and an <imprint> element.

Figure 10 Single record visualized on Zenodo

Codice diplomatico longobardo 1-2: the original charters and coeval copies written in Tuscany between AD 714 and 774

Metadati	Descrizione
	TEI XML edition of the original charters and coeval copies written in Tuscany between AD 714 and 774 from the Codice diplomatico longobardo 1-2, ed. by Luigi Schiaparelli, Roma: Tipografia del Senato, 1929-1933. The digitized version comes from the site of the Institut für Mittelalterforschung of the Austrian Academy of Sciences. Expansions of abbreviations ("expan"), indications of lacunae ("damage") as well as all diplomatic information provided in CDL are preserved only if relevant to morpho-syntactic analysis. The tags "abbr" and "gap" are used for unexpanded abbreviations and unwritten spaces respectively. Line and page information omitted. The ChLA numbers added in order to enable comparison with the facsimiles. For details, see Korkiakangas, T. and Passarotti, M. (2011), <i>Challenges in Annotating Medieval Latin Charters</i> , in «Journal of Language Technology and Computational Linguistics», 26, pp. 103-114.
Lingua	lat
Parole chiave	Early Middle Ages; charters
Collegamento alla fonte originale	https://doi.org/10.5281/zenodo.4745084
Istituzione o autore	Korkiakangas, Timo
Tipi	dataset
Diritti	open
File disponibili	https://zenodo.org/api/records/4745084/files/CDL_TEI_XML_base_text_for_LLCT.xml/content

Figure 11 Same resource of figure 9 in CHIt after the application of the output model.

A quick comparison of figures 8 and 11 shows the uniformity of the visualization of the retrieved resources that the mapping ensures. This is important for the user experience, since it guarantees a pleasant and smooth navigation and the user has the feeling of actually browsing an ecosystem, where items are in real connection with each other, as if all the resources were created in that place and had ever been in communication. Moreover, figure 10 shows a raw XML file on Zenodo, which is not comprehensible for the majority of average users, and also expert users could find this presentation insufficient. In this sense, the integration of a visualization tool such as EVT offers a great opportunity for the consultation and valorization of the files related to the selected records. On the CHIt frontend, once the single resource is selected, from the section *Available files* (Figure 11) the user can open, in this case, the XML document and visualize it thanks to the EVT software. The detailed functioning of EVT within CHIt is the topic of section 4.3, but figure 12 shows how the XML file is visualized with this tool.

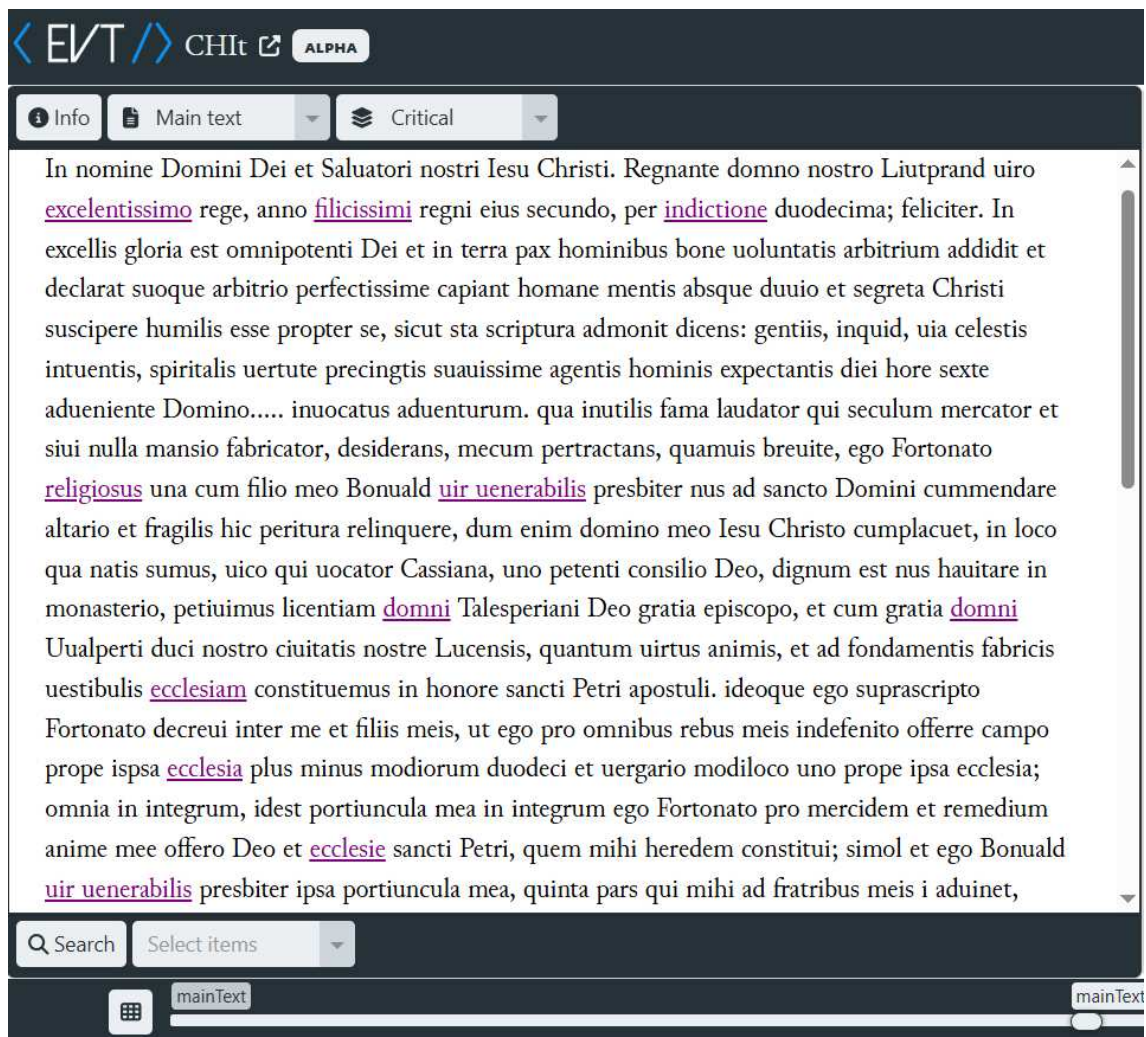


Figure 12 Visualization with EVT of Zenodo resource

Thanks to this integration, the initial goal which involved the valorization of smaller projects is supported, as it offers a visualization tool to resources that do not have one.

4.2.1.3. GitHub

The third provider, GitHub, was chosen almost simultaneously to Zenodo because they were selected for the same general reason, namely the valorization of smaller projects.

However, this is (almost) the only thing that they have in common, as far as CHIt is concerned. In fact, the target items are different, the search of the resources has been operated in a different way, and the behavior of the single resources within CHIt is different as well.

GitHub is a platform for developers, where they can store, manage, create and share their code. It is free and open-access, based on Git¹⁹⁰ and therefore it allows to keep versioning under control. It is widely used worldwide, especially in collaborative projects.

Given its nature, the benefits that GitHub can bring to the collection of items of CHIt are rather specific and it seems evident that this provider is the one that goes the closest to the original – re-elaborated – idea of addition of scholarly projects.

As for the retrieval, this platform obviously provides APIs¹⁹¹ and the documentation is also very detailed, and this is expected since it is a service for developers. However, as already anticipated, the available APIs could not serve the scope of this project adequately since the kind of search that they perform is a technical one, is based on the code language and on the repository structure. Less concern and attention are dedicated to the content of the repositories, and this penalized the retrieval of resources for CHIt. For example, it was not possible to filter for keywords related to the topic of repositories, because they are not assigned at all. The advanced search can filter for repository, code, issue, user and wiki options. Therefore, the operational step in this case was the same as for Zenodo: manual selection. Resources relevant for the topic of Italian cultural heritagewere identified starting from known projects and/or scholars, and the single repositories can be retrieved through a single API call. The base URI is <https://api.github.com/>, and to perform a call that addresses a specific repository the parameters that have to be added are [repos/{owner}/{repo}](#). In this way, the general metadata are retrieved, many of which are useless for the scope of CHIt. What is more important, this API call does not retrieve the files stored in the

¹⁹⁰ <https://git-scm.com/>

¹⁹¹ <https://docs.github.com/en/rest?apiVersion=2022-11-28>

repository. The retrieval of files is performed with a different call, in which it is necessary to add [contents/{path}](#), but still what is obtained is a list of the files available at that path. To sum up, APIs seemed not to be a good solution in the case of GitHub. For this reason, the resources aggregated from this third provider are inserted manually in the CHIt backend both for the search phase and for the consultation phase. The selected metadata are at any rate much less than the other two providers and are enlisted in table 8.

GitHub input model	CHIt backend insertion
description	description
login	creator
url	references referencedBy
name	title

Table 8 GitHub input model into CHIt backend insertion

More metadata are inserted without being part of the original input schema, but are needed to create harmonized resources and can be added with a certain degree of confidence. To supplement some of the metadata available for the resources taken from the other providers and align the outputs, the following information is further added to the GitHub resources (table 9).

Metadata added manually to the backend for GitHub resources

subject
language
publisher

references referencedBy¹⁹²

Table 9 More metadata for GitHub

In this way, all the relevant information regarding each resource is present. The way in which CHIt behaves with GitHub resources is therefore different once again, but the result on the frontend is the same. In fact, for this provider the middleware extracts information directly from the backend. The fetching, mapping and returning are the same nonetheless, as it is evident in figure 13.

Edizione digitale del <i>Devisement dou Monde</i>		
Metadati	Descrizione	File XML per la realizzazione del progetto DEDM.
	Parole chiave	Codifica XML-TEI, Digital Scholarly Edition, filologia digitale, manoscritti, medioevo, Marco Polo, oriente, Veneto, Venezia
	Lingua	it, eng, lat, fr
	Istituzione o autore	Buzzoni Marina, Fabbris Giulia, Rosselli Del Turco Roberto, Burgio Eugenio, Simion Samuela
	Collegamento alla fonte originale	https://github.com/vedph/dedm-evt https://vedph.github.io/dedm-evt/#/readingTxt?d=tr&s=c106_tr106&e=critical

Figure 13 GitHub resource visualized on CHIt frontend

4.2.1.4. Germanic cultural heritage in Italy and OntoVE

The fourth and last provider from which the resources that populate CHIt are gathered is a Knowledge Base, named OntoVE.¹⁹³ It is another PON project in development at Ca' Foscari University of Venice that benefits from the same type of fundings of CHIt, with

¹⁹² At the present moment, the files are inserted in the same field as the link to the original source, but a new dedicated field will be added as soon as possible.

¹⁹³ <https://chideba.github.io/ontove/ontology.html>

the difference that this is a research project, not a PhD one. The aim of this work is the development of two ontological modules for the description of items belonging to the Germanic cultural heritage related to the Veneto region in Italy. Even if the domain of this dataset is very specific and restricted, the items of which it is made of are heterogeneous, ranging from archaeological findings to written documents.

The Knowledge Base is divided into two modules: *OntoVE_Archeo*¹⁹⁴ and *OntoVE_WrittenRecords*.¹⁹⁵ The first one describes archaeological items, while the second bibliographic ones, but the two will be visible as one single dataset at the end of the project. The two ontologies can be used independently, but are merged in OntoVE for the description of the heterogeneous items collected.

As already mentioned, the resources that compose these datasets are heterogeneous, and so are the metadata describing them. In fact, after determining which items are going to be included in the dataset, their metadata need to be formalized according to established ontological models. For this scope, abstract domain ontologies have been aligned, and in particular the abstract classes provided by CIDOC-CRM, v 7.1.2 (Bekiari et al. 2022) and LRMoo, v 0.9.6 (Bekiari et al. 2023) with finer grained domain ontologies that provide a standardized vocabulary, such as ArCo,¹⁹⁶ Bibframe¹⁹⁷ and Bibo.¹⁹⁸ For more details on this alignment cf. De Bastiani (2023; 2024).

This ontology is being developed not only for a specialized field of academics, but also for citizens, since the resources described testify local history. Tourists as well might find interesting information related to locations of their interests. For this reason, a double solution for the visualization of the objects was considered so as to enhance a democratic fruition of the resources, in line with the principles of the Faro Convention, cited at the beginning of this dissertation. In fact, the consultation and interrogation of

¹⁹⁴ https://github.com/ChiDeBa/ontove_archeo

¹⁹⁵ https://github.com/ChiDeBa/ontove_writtenrecords

¹⁹⁶ Gangemi et al. (2019)

¹⁹⁷ Library of Congress, <https://www.loc.gov/bibframe/>

¹⁹⁸ <https://www.dublincore.org/specifications/bibo/bibo/>

a SPARQL Endpoint might be privative for a huge section of audience. The ontology will be publicly available through a SPARQL Endpoint built on Apache-Jena-Fuseki-4.7.0,¹⁹⁹ from which the consultation can be personalized as long as the user knows the SPARQL language, which is the most powerful language to work with LOD and RDF databases. However, since this knowledge is rather exclusive, a user interface has been implemented as well, using the Sampo model (figure 14).²⁰⁰ Thanks to this interface, it is possible to present research data through pre-defined SPARQL queries, which is performed at the backend level. In this way, the user only sees the results and can still filter them according to some search criteria. The visualization comprises a subset of the whole ontology, but makes it available to a wider public.

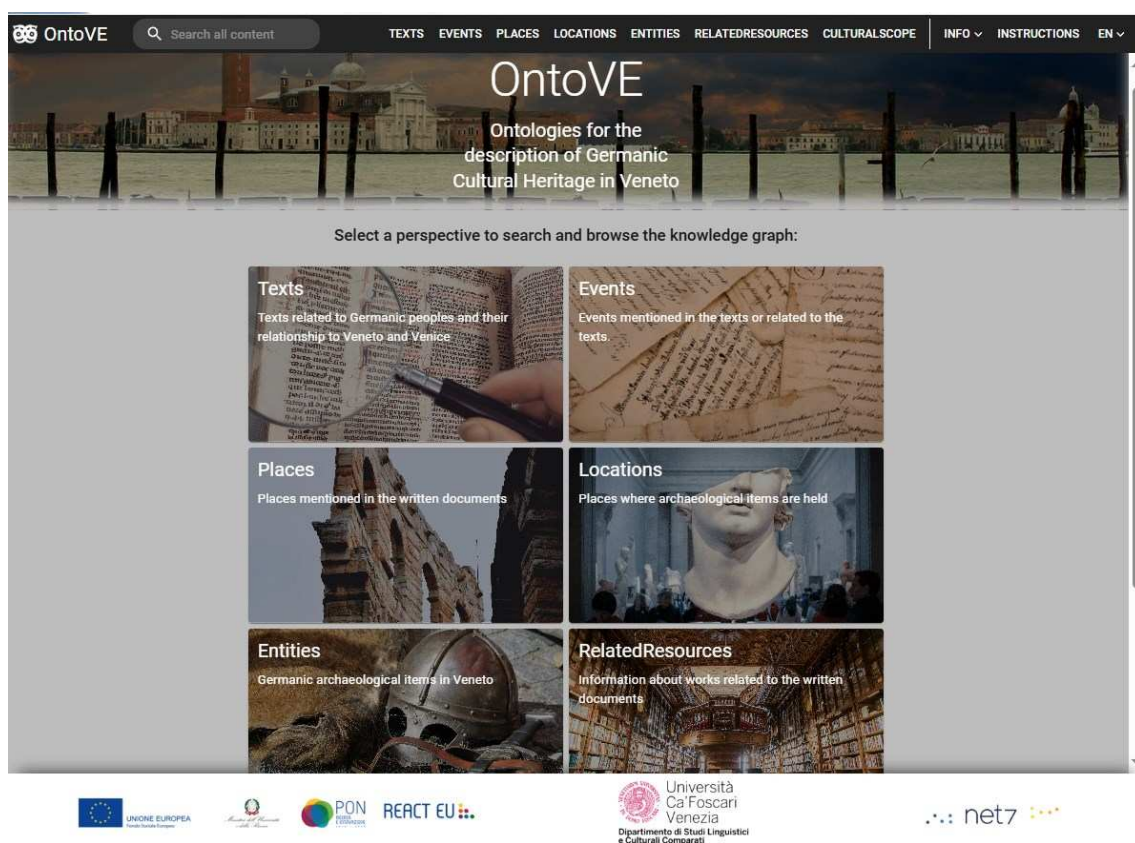


Figure 14 User interface of OntoVE with the Sampo model

¹⁹⁹ <https://jena.apache.org/>

²⁰⁰ Ikkala et al. (2022), GitHub repository <https://github.com/SemanticComputing/sampo-ui>

On the other hand, the full ontology is always available at the Endpoint and users skilled in SPARQL can refer to it in case they had more specific research questions.

This brief introduction of the OntoVE project shows that it shares much with CHIt, even if the workflow and the results are different. The idea of a collaboration was brought forth precisely by these similarities. Starting with, the same funding sustains them, and for this reason the accent on innovation, technology and enhancement of human capital is present since the beginning in both of them. For the same reason, also the collaboration with Net7 is a shared point. Another guiding principle is the valorization of the Italian cultural heritage in the digital paradigm and its availability online through easy-to-use platforms, in open-access format and suitable for a varied audience. Last but not least, the stress posed on standardization and reuse is very strong in both projects, as FAIRness in the Digital Humanities field – but not only²⁰¹ – is a cornerstone nowadays. In fact, also the resources added to the OntoVE Endpoint are all already available on the web and are gathered and aligned in this new digital environment.

This is the reason why in 2023 a collaboration of the two PON projects started, inspired by the topic of the ADHO DH2023 conference *Collaboration as Opportunity*, held in Graz (Austria).²⁰² The communication between these two projects has brought to a series of advantages for both of them, and to a further valorization of the Italian CH, with a specific focus on the Germanic resources available in the Veneto region. On the one hand, the resources belonging to OntoVE and visualized on CHIt will be inserted in a new context, which is a third visualization output for OntoVE, but this time in a different environment. In fact, as they are formatted with the CHIt output model, they become interoperable within the aggregator. This means that they can be put in relation with the other items of CHIt, retrieved from Europeana, Zenodo and GitHub. Moreover, they can be annotated with Pundit as every other digital object on CHIt. On the other hand, CHIt gains more scholarly resources that populate and enrich the collection of Germanic

²⁰¹ Cimiano et al. (2020); Kush et al. (2020)

²⁰² Alliance of Digital Humanities Organizations, <https://dh2023.adho.org/>

resources, retrieved from an independent project that is open-access and scientifically valid.

Last but not least, another *desideratum* of this collaboration was the possibility of designing a model that could be applied to other RDF databases. In this way, more users could have access to databases that would otherwise require knowledge of SPARQL.

The collaboration was presented the first time at DH2023 as a poster²⁰³ and subsequently at a workshop held in Venice some months later.²⁰⁴ A more structured description has been proposed in De Bastiani, Fabbris and Bertozzi (submitted paper) and the final result of this first phase of experimentation will be visible on the CHIt platform at the end of the project.

From a technical point of view, the workflow for the retrieval of the cultural objects is, once again, different but akin to the previous ones. In this case, there are no APIs, since the metadata are provided by the *ad-hoc* SPARQL query, which is presented in table 10.

²⁰³ Abstract (De Bastiani and Fabbris 2023b), poster (De Bastiani and Fabbris 2023a)

²⁰⁴ <https://diptext-kc.clarin-it.it/cultural-heritage-in-the-digital-paradigm/>

```

SELECT ?archprop ?label ?labelLoc ?externalLink

      (GROUP_CONCAT(DISTINCT ?bibliography; SEPARATOR=', ') AS ?bibliographies)
      (GROUP_CONCAT(DISTINCT ?reference; SEPARATOR=', ') AS ?references)
      ?id ?labelId ?repoLink

WHERE {
  ?archprop a arco:ArchaeologicalProperty;

    rdfs:label ?label;

    (cidoc-crm:P54_has_current_permanent_location | cidoc-
    crm:P55_has_current_location) ?currentLocation.

    ?currentLocation rdfs:label ?labelLoc;

    ontove:externalLink ?externalLink.

    ?archprop context-description:hasBibliography ?bibliography.

    ?bibliography (context-description:completeBibliographicReference | bibo:handle)
    ?reference.

    ?archprop ontove:hasWebPage ?webpage.

    ?webpage bibo:handle ?repoLink.

    ?archprop cidoc-crm:P1_is_identified_by ?id.

    ?id rdfs:label "0500590886"@it.

    ?id rdfs:label ?labelId

    FILTER(LANG(?label) = "it" && LANG(?labelLoc) = "it" && LANG (?labelId) = "it")
}

GROUP BY ?archprop ?label ?labelLoc ?externalLink ?id ?labelId ?repoLink

```

Table 10 SPARQL query for the retrieval of items from the OntoVE database

This request retrieves archaeological artifacts, their current location and their bibliography from the *OntoVE_Archeo* module, which is made of the combination of the CIDOC-CRM, v. 7.1.2 and the ArCo ontologies. Bibliographic references are assigned classes and properties from the Bibo ontology. Some properties were created appositely

for the project and have the prefix *ontove* (cf., for example, *externalLink* in table 10). In a few words, this query searches for the individuals belonging to the *arco:ArchaeologicalProperty* class, and their current location. Furthermore, with the query the bibliographic entries assigned to each individual are requested, which can have the form of a bibliographic reference or a handle, the link with the geographical coordinates assigned to the places where the items are currently stored, and the handle of the digital repository where metadata about the items are found.

The query is requested directly from the middleware to the Endpoint, and once the resources are gathered, the workflow is similar to the other cases. In particular, the resources undergo the same process of mapping according to the CHIt output model, even though the set of metadata is quite different in this case, given the different nature of the retrieved objects (cf. table 11).

OntoVE input model	CHIt output model
[See note] ²⁰⁵	subject
"OntoVE"	publisher
?webpage bibo:handle ?repoLink.	requires

²⁰⁵ The subjects for OntoVE are retrieved thanks to another query that was added in order to have this information also for this provider. The query is the following and selects all the cultural scopes attributed to the resources.

```

SELECT  ?archprop ?labelarcheo ?culturalscope ?label
WHERE
{
  ?archprop a arco:ArchaeologicalProperty;
    context-description:hasAuthorshipAttribution ?authorshipAttribution.
    ?authorshipAttribution context-description:hasCulturalScope ?culturalscope.
    ?culturalscope rdfs:label ?label.
    ?archprop rdfs:label ?labelarcheo.
    FILTER(LANG(?label) = "it" && LANG(?labelarcheo) = "it")
}

```

rdfs:label riferita al soggetto ArchProp	title
context- description:completeBibliographicReference	context- description:completeBibliographicReference
bibo:handle	bibo:handle
rdfs:label riferita all'id della risorsa nell'ontologia (quindi la sua URI)	identifier
rdfs:label riferita alla località di conservazione, individuata con le seguenti proprietà (cidoc- crm:P54_has_current_permanent_location cidoc-crm:P55_has_current_location)	cidoc- crm:P54_has_current_permanent_location cidoc-crm:P55_has_current_location
ontove:externalLink²⁰⁶	spatial
a arco:ArchaeologicalProperty	a arco:ArchaeologicalProperty
[Cf. table 10]	[the whole resource is the result of the query presented in table 10]

Table 11 OntoVE input model into CHIt output model

The output in the CHIt frontend is similar to the other resources, even if the starting point was totally different. An example is provided in figure 15.

²⁰⁶ *ontove:externalLink* can be used with links to external authority records referred to people and places. In this specific query, it retrieves places.

Casteldardo struttura di fortificazione

Metadati	Fornitore	Ontove
URI nell'ontologia		https://w3id.org/ontove/ontology#6a938d2f-1d0d-4634-b464-ab69d2006cf6
Titolo		Casteldardo struttura di fortificazione
Posizione		Trichiana
Collegamento alla fonte originale		https://catalogo.beniculturali.it/detail/ArchaeologicalProperty/0500591273
Coordinate geografiche		https://www.geonames.org/3219557/trichiana.html
Bibliografia dell'oggetto		https://w3id.org/ontove/ontology#075db212-3a28-4e0f-abe9-1f86d5487775
Bibliografia completa		https://catalogo.beniculturali.it/detail/ArchaeologicalProperty/0500591273
Archaeological Property (ArCo)		https://w3id.org/ontove/ontology#d1d23b9b-a5bc-4e61-9ce2-6b28bd1c5a2c

Figure 15 Visualization of OntoVE resources into CHIt frontend

4.2.2. Creating records

It was already mentioned in the dedicated section 3.2.3.1 that an interesting aspect of cultural – but not only – aggregator is the availability of collections, stories, narrative paths that work as guiding tools for the user in what can sometimes be an intricate jungle of resources. The Muruca architecture foresees the possibility of creating such collections, and in CHIt this has obviously been exploited.

As already mentioned in section 3.2.3.1, thematic paths can be implemented in different ways: they can be narrative paths with resources selected from the archive and external references linked to specific parts of the text, they can be a sort of pre-compiled

research with selected resources that automatically creates an independent collection, or there could be a section with short narratives dedicated to different topics.²⁰⁷

It has been specified in the aforementioned section that the solution adopted for CHIt is twofold: collections and narrative paths (e.g., itineraries).

The first itinerary that was implemented in CHIt is related to the Germanic cultural heritage in Italy. For an explanation of the reasons that guided this decision, see chapter 6 where the choice is explained in detail, together with the exposition of a case study that aims at testing the functionalities of the aggregator.

However, it was noticed that the number of resources available through the four providers belonging to the selected field was not satisfactory. OntoVE is the only one that contributes with a good amount of resources in this sense, being entirely dedicated to the Germanic objects present in the Veneto region. Some of the selected items of Zenodo and GitHub are valuable resources as well, but the number remains low nonetheless. For this reason, it was decided to exploit the potential of the backend. Previously in this chapter it was illustrated how the metadata of the resources gathered from GitHub and Zenodo are stored in the backend, even if to different degrees. It was decided to adopt a procedure analogous to that used for GitHub to add resources related to the Germanic cultural heritage in Italy that are not FAIR and, therefore, cannot be integrated as wished. This means that the websites that host them do not expose APIs or other ways for freely reusing their data even if it is potentially allowed according to the license under which they are published. The consequence is a drawback both for the source, as it reduces or nullifies the possibility of interoperability, and for CHIt, as the FAIRness of the platform is shelved in order to have access to these resources to some extent.

In a practical way, a survey of available resources was performed and a selection of their metadata was added to the backend according to the information displayed on the

²⁰⁷ On the website of the Muruca demo (<https://murucaracconta.muruca.cloud/>), there is the possibility to explore the tales selected according to specific themes (*Home > Motivi*) or to follow narrative paths (*Home > Percorsi*) that are related to the topic.

websites. In so doing, no files are available directly from the CHIt frontend, but the link to the original source is given. Users can thus go back to it and consult the files directly from there. As a general rule, each resource of this type displays a description, the author(s), some keywords, the main language(s) and the link to the external website. When a resource is contained in another one, an internal reference to that record is given. Let us take as an example the website of the Institute of Cimbrian Culture of Roana.²⁰⁸

4.3. Connection to the annotation and visualization tools

The last section of this central chapter is dedicated to the integrated tools, the annotator and visualization technology. As already mentioned, these are Pundit and EVT, the first one developed by the partner company Net7 and the second one by a team of researchers and programmers coordinated by Professor Rosselli Del Turco and operating mainly from Turin (Piedmont) and Pisa (Tuscany).

The structure of Pundit and of EVT has already been explained in detail in sections 3.2.4 and 3.2.5, and will not be proposed again here. The scope of this short section is to specify how these tools are integrated in and interact with CHIt.

The first tool discussed here is Pundit, and the preliminary remark that has to be made is that it comes as a browser extension and is freely available on the Chrome Web Store.²⁰⁹ The download is easy and immediate, Pundit is ready-to-use in a handful of seconds. A login is required since the annotated data need to be linked to an account in order to be stored, retrievable and reusable. It is possible to register via Google, Facebook or EGI check-in, or simply with a personal email. Once these simple operations are completed, the Pundit web annotator can be enabled on every web page from the

²⁰⁸ <https://cimbri7comuni.it/>

²⁰⁹ <https://chromewebstore.google.com/detail/pundit-annotator/eilojhemeemnpfmailfeiggpgilblaco?pli=1>

dedicated icon that can be found among the collection of the browser extensions. From their personal area, users can create and organize notebooks, as already said above. To start annotating, it is necessary to select a portion of text after activating Pundit. When the portion of text is selected, a window appears from which the user can select the type of desired annotation: highlighting, comment, tag or semantic annotation. Once the annotation is processed, it is automatically saved in the last used notebook. The user can easily move it into another notebook by simply clicking on the notebook assigned by default and selecting the desired one. From the private area, the user can further organize and access the saved data.

As for EVT, its integration was possible since it has been recently released as a service (EVT SaaS),²¹⁰ providing a web-based environment for hosting and visualizing XML-based digital editions. The application runs in standalone mode on a dedicated server with its own URL. Deployment requires downloading the source code from the repository, compiling it and hosting the generated site on a web server.

EVT SaaS allows to display a digital edition in a dynamic way passing a configuration JSON as a parameter in which the remote location of the XML file is specified, along with additional configuration parameters such as image folders and UI settings.

In the frontend application, when information about XML files is present, a dynamic link is generated by combining the base URL of EVT with a parameter pointing to a JSON file. This JSON file dynamically returns the necessary configuration for visualizing the XML resource within EVT SaaS.

To integrate EVT SaaS into an existing digital edition project, the configuration file must be hosted on a web-accessible domain and referenced in the service URL using the syntax:

²¹⁰ At the time of writing, the source code is hosted in the following repository <https://github.com/SilentWave/evt-viewer-angular/tree/develop>. However, it will be moved to the official EVT repository (<https://github.com/evt-project/evt-viewer-angular>) as soon as the documentation will be ready.

```
https://{Domain_Hosting_EVT}#/readingText?RemoteConfig=true&fileConfigUrl={ENCODED_CONFIG_FILE_URL}
```

For security reasons, EVT SaaS restricts the use of remote configuration files to a predefined list of trusted domains, managed through the `host_config.json` file. This ensures that only authorized sources can load edition configurations, mitigating potential security risks associated with arbitrary file access.

5. Content organization, visualization and testing

This concluding chapter of section II wants to be a first and general glimpse on the actual output of the implementation phase described in detail in the previous chapter. Even if the writing of the code for the middleware took the biggest part of the three years in which this project has been developed, during the last year great efforts were put also in the finalization of the frontend and the testing of its functionalities.

Different parts of this work stressed how important UX and user engagement are and how a bottom-up perspective can increment the usability of an application. For this reason, these aspects could not be set aside and occupied a prominent role in the last phases of the implementation.

In section 3.2 it was said that choosing to adopt an already existing framework is successful for different reasons, among which the reduced use of time. The Muruca architecture was already standing on its own and had already been tested for other projects, and this proved extremely helpful in this last phase. In fact, it was also possible to have a visual and therefore concrete idea of what could be done with it and what the possible outputs were. On the one hand, the code of the middleware was reworked more extensively, since it is the main layer and handles the business logic. Moreover, it is pure code, and handling it requires sophisticated development skills. On the other hand, it is possible to work on the backend more intuitively, thanks to the use of WordPress. Of course, behind the backend there is much code as well, but thanks to its high configurability there is no need to change it deeply. For example, the insertion of contents and images, the creation of pages, collections and itineraries can be done via WordPress also by non-developers. Once again, the expertise of the Net7 team was essential also in this phase in order to guarantee the best visualization results and to adjust the frontend – also acting from the code, as a matter of fact – to satisfy the expected views and the overall experience.

The frontend and its structure are briefly introduced in the first subsection of this chapter, but readers can enjoy the application on their own by accessing the website at <https://chit.chitontove.it>.

The second and last subsection is dedicated to the presentation of the testing phase, which was performed manually given the nature of the project. In fact, when a project is highly concerned with UX, the presence of a human to test all the functionalities is required. A human user can give an automated set of actions to testing tools such as Selenium,²¹¹ but these tools are not able to test feelings and perceptions related to the navigation within the application.

5.1. The frontend

The homepage of CHIt has been studied in order to give access to all the different parts of the application in an intuitive way. The presentation page is divided in three main horizontal blocks: the upper one contains the title and a short introductory sentence, the central one gives direct access to the collections and the block at the bottom gives a general yet concise idea of what CHIt is and how it can be used (figure 16 below).

A quick overview of the collections is now provided, not from a technical point of view, since it has been the topic of section 3.2.3.1, but as far as the content is concerned.

The first one, *Risorse manoscritte*, gathers all the written sources available in the platform, of which manuscripts are the majority. However, to this category there belong also letters, laws and charters, inscriptions.

The second collection, *Edizioni scientifiche digitali*, is about scholarly digital editions and their relevance is prominent given the context in which this whole project is developed. The creation of this collection is the ultimate attempt to valorize the smaller scholarly digital works from which the idea of CHIt was born. To it there belong some of the resources added manually from Zenodo and GitHub, as well as some external ones that could not be made properly interoperable.

The third (*Isole linguistiche germaniche in Italia*) and fourth (*I Goti e i Longobardi in Italia*) collections are dedicated to the valorization of the Germanic heritage in Italy, in particular the minority languages and the Lombard and Gothic people. Once again, the

²¹¹ <https://www.selenium.dev/>

reader is redirected to chapter 6 for clarifications regarding this choice or to the specific itineraries of the applications, of which something more will be said below.

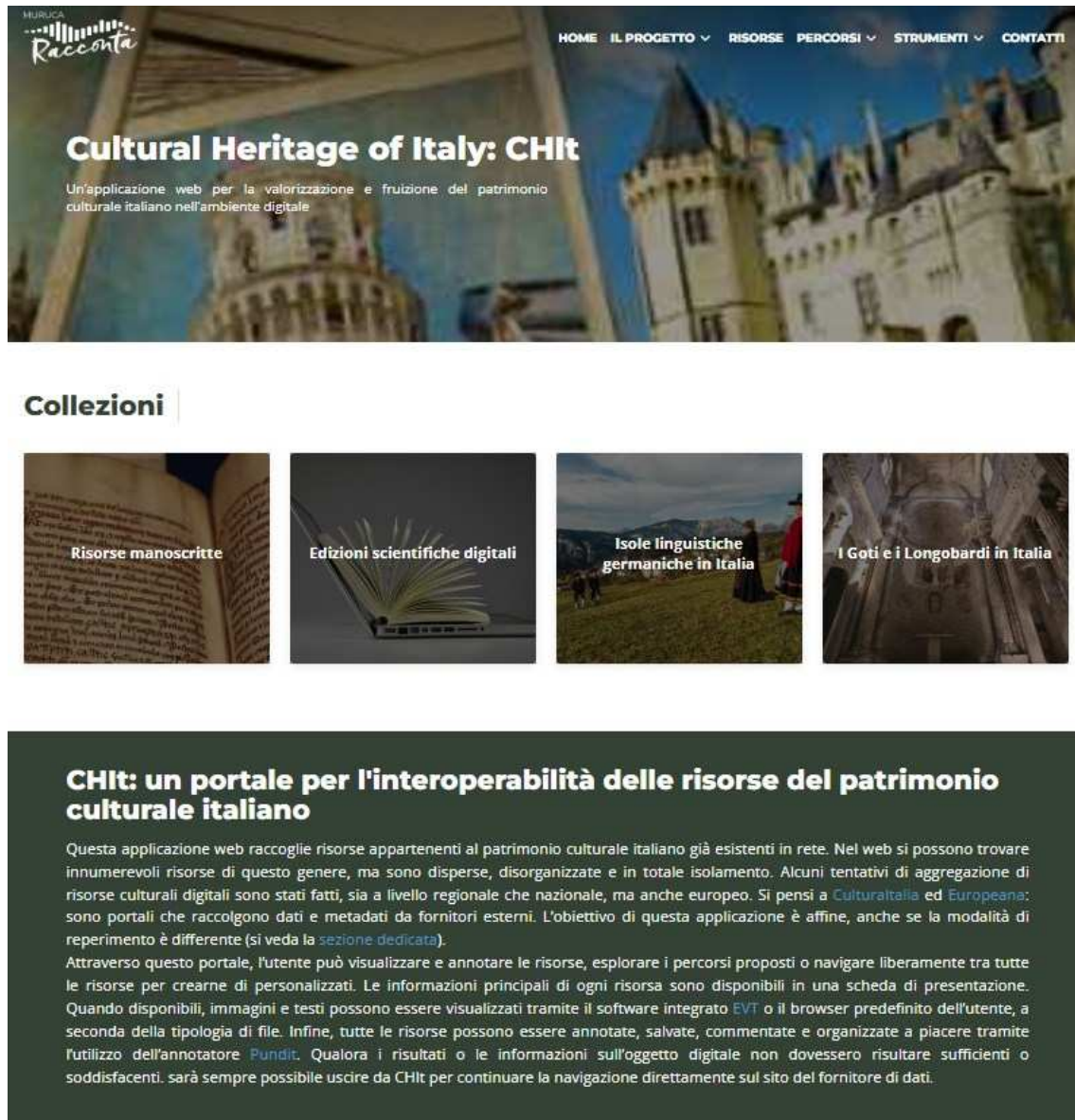


Figure 16 CHIt homepage

Above the upper block of the homepage, there is the menu with all the other sections available on the website. Some are informational, others are for the organization of the contents. The map of the website is the following:

- Home
- Il progetto
 - I provider
 - Europeana
 - Zenodo
 - GitHub
 - OntoVE
- Risorse
- Percorsi
 - I Goti e i Longobardi
 - I Longobardi
 - I Goti
 - Le isole linguistiche
 - Walser
 - Cimbri
 - Sappadino
 - Saurano
 - Timavese
 - Mòcheno
 - Ontologia
- Strumenti
 - Annotatore
 - Visualizzatore
- Contatti

In particular, in the section *Il progetto* (the project) there is an in-depth description of the whole project and a further subcategorization with a general overview of the different providers. At the bottom, each provider is further described in its own section.

The third button of the menu leads to the whole catalogue of resources. From here, users can scroll all the items, which would be quite useless and confusing, or filter them according to provider, institution or author and keywords. Of course, free-text search is also supported. By clicking on one of the items, a separate window opens, from which the user can read the metadata and see related files. Examples of this kind of visualization are given in the single sections of the different providers in chapter 4.

The next label is the one dedicated to the Germanic people in Italy. This is the first thematic path presented to the users, and is further divided into three areas: Lombard and Gothic people, minority languages and the OntoVE ontology, which is also a provider but since it aggregates exclusively Germanic resources, it seemed fitting to add it also in this part of the website. Of course, the deeper the exploration into the structure of the itineraries, the more specific the information becomes. For example, from the section dedicated to the minority languages, users can read about some generic considerations on the topic: what these languages are and where they are spoken, how they are still preserved, where the first settlers came from and when they settled in these areas. At the bottom of the page, as it should be clear by now, there is the section dedicated to the related resources (figure 17).

HOME IL PROGETTO ▾ RISORSE LE POPOLAZIONI GERMANICHE IN ITALIA ▾ STRUMENTI ▾ CONTATTI

Le minoranze linguistiche in Italia di origine germanica

Le lingue germaniche presenti sul territorio italiano sono principalmente sette, distribuite in maniera molto differente lungo tutto l'arco alpino. Queste varietà sono il walser, il mòcheno, il cimbri, il sappadino, il saurano, il timavese e la varietà della Valcanale. Sono presenti in cinque delle otto regioni settentrionali della penisola: Valle d'Aosta, Piemonte, Veneto, Trentino-Alto Adige e Friuli-Venezia Giulia.

Le varietà più utilizzate su un territorio più ampio sono il walser e il cimbri, che coprono quattro di queste cinque regioni e contano diverse parlate. Il walser è l'unica minoranza germanica parlata nella parte occidentale dell'Italia ed è distribuita tra alcuni territori della Valle D'Aosta e del Piemonte. Il cimbri è parlato, come le restanti cinque varietà, a est e si trova tra il Veneto e il Trentino-Alto Adige. Le altre varietà sono collocate in singoli territori.

Risorse
collegate

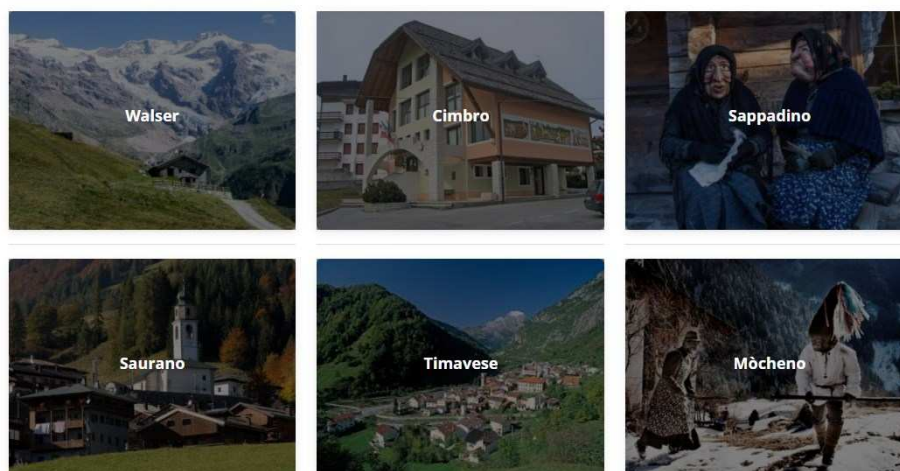


Figure 17 View of the Minority languages itinerary

From the related resources, users can access the specific minority language of their interest and discover even more detailed information. In the specific case of this topic, each language is enriched with an organized list of online resources that fall into the categories of *Cultural tourism*, *Media*, *Publications and other* and *Linguistic resources*. The survey of the available resources was presented at the conference *Small Languages, Big Ideas*, held in Uppsala (Sweden) in June 2024.²¹² One last time, the bottom of the page provides access to all the related single resources of the language under exploration. Figure 18 shows the example of the Walser language.

²¹² Fabbris (2024)

Walser

Il walser è l'unica varietà germanofona presente a ovest del nord Italia, e si estende in alcuni territori della Valle d'Aosta e del Piemonte. In particolare, in Valle d'Aosta è parlato a Gressoney/Greschoney e a Issime/Eischeme, mentre in Piemonte a Campello Monti/Kampel, Rimella/Remmalju, Formazza/Pomatt, Alagna Val Sesia/Im Land e Carcoforo/Chalchoufe. I dialetti variano da zona a zona.

Il walser è una minoranza di origine alemanna, portata da alcune popolazioni attorno al 1100 che si stabilirono alle pendici del Monte Rosa.

Turismo culturale

- [Walser Cultura \(*\)](#)
 - [Beni etnoantropologici d'epoca](#)
 - [Toponimi](#)
- [Sportelli Walser \(*\)](#)
- [Walser-alps \(*\)](#)
 - [Cultura](#)
 - [Le regioni Walser](#)
 - [Musei Walser](#)
- [Centro Studi e Cultura Walser / Walser Kulturzentrum di Gressoney e Issime \(*\)](#)
- [Associazione Augusta Issime \(*\)](#)
- [Gruppo Walser Campello Monti / Walsergemeinschaft Kampel \(*\)](#)
- [Centro Studi Walser Rimella \(*\)](#)
- [Due sentieri culturali a Rimella](#)
- [Rimella Walser \(*\)](#)
- [Il territorio di Alagna Valsesia](#)

Media

- [Video sui Walser della Valle d'Aosta](#)
- [Fotografie – Walser Cultura](#)
- [Galleria fotografica Walser del comune di Alagna Valsesia](#)
- [Museo virtuale](#)

Pubblicazioni e altro

- [Pubblicazioni](#) Gressoney e Issime
- [Rivista Associazione](#) Augusta Issime
- [Capitolo](#) su Gressoney del libro *Isole di Cultura*
- [Capitolo](#) su Issime del libro *Isole di Cultura*
- [Capitolo](#) su Campello Monti del libro *Isole di Cultura* x
- [Capitolo](#) su Rimella del libro *Isole di Cultura*
- [Rivista Remmalju](#) (Rimella)

Risorse linguistiche

- Pronuncia
 - Lessico di Gressoney e Issime (1) e (2)
- Dizionari
 - [Walser-alps](#) (*)
 - [Walser - Tedesco](#) (*)
 - [Gressoney](#) (*)
 - [Issime](#) (*)
- Altre risorse linguistiche
 - [Giornale](#)
 - Architettura, società e lingua: Homepage Centro Studi e Cultura Walser > I Walser > menù a tendina

Risorse collegate

Walser Cultura - banca dati
La banca dati W...

**CLiMAIp salvaguardia delle lingue
minoritarie in Piemonte e Valle
d'Aosta**
Il progetto CLi...

AlpiLink - Lingue alpine in contatto
AlpiLink è un a...

Dizionari di CLiMAIp
Dizionario tits...

Figure 18 Walser itinerary

The last type of visualization is the one already cited for the single resources, to which users can be redirected from the bottom of each section. Figure 19 complements the visualization outputs presented in chapter 4. In this case, the resource is from an

external provider and what the user can see on the screen are the metadata manually inserted in the backend. Thanks to this method, the presentation of the digital object is in line with all the other retrieved items in order to guarantee a smooth navigation. The uniform presentation available for all the resources helps to recreate the feeling of a unique environment that hosts items from disparate websites. Figure 19 shows how the visualization of this external resource looks like the ones retrieved from Europeana, Zenodo, GitHub and OntoVE.

Last but not least, from here the user can take advantage of Pundit to annotate the resources, while for the external providers EVT plays no role since potential multimedia files have to be opened directly from the source.

Walser Cultura - banca dati		
Metadati	Descrizione	La banca dati Walser Cultura, nata nell'ambito del progetto Interreg IIIB Walser Alps, è concepita quale centro informatizzato di documentazione finalizzato alla raccolta, alla divulgazione e alla valorizzazione del patrimonio culturale Walser ed alla nascita di un sistema informativo capace di salvaguardare e valorizzare un'eredità di rilevanza universale e promuoverne la diffusione e la conoscenza. Sezioni del sito: documenti, fotografie d'epoca, canti della tradizione orale, beni etnoantropologici, toponimi, bibliografia.
	Parole chiave	walser, lingue minoritarie, Valle D'Aosta, Piemonte, filologia germanica, archivio digitale, strumenti digitali
	Lingua	it, walser
	Istituzione o autore	Walser Alps Interreg III B Alpine Space
	Collegamento alla fonte originale	https://www.walser-cultura.it/

Figure 19 Visualization of single resource from external website

The second last label of the menu contains the references to the integrated tools, Pundit and EVT. Once again, the structure of the section is hierarchical and users can access the dedicated pages from the *Related Resources* section or from the drop-down menu.

Last but not least, the *Contact* section contains the email address to which users can refer for any reason: help, suggestions, bugs, recommendations, new resources etc.

This is the general layout of the application.

The Muruca architecture allowed to create an intuitive, graphically pleasant and easy-to-use structure suitable for a wide and varied audience.

5.2. Testing phase

Chapter 4 has shown in detail how the application was developed and has presented the main structures that lie behind it. However, to ensure the correct performance of CHIt, it was necessary to plan a phase in which the actual functionalities of the application were tested against the specified requirements.

Functional testing in software development is highly practiced and consists in checking that each functionality of the application, given the appropriate input test, gives back the expected output. In this context, UI, APIs, database, security, client or server application are tested.²¹³ Its importance is paramount to verify whether the application works as intended or not and that its functionalities, usability and reliability are ensured. The steps that are usually involved are:

1. The identification of the test input, which consists in singling out the functionalities that need to be tested.
2. The computing of the expected outcomes by creating the input data based on the specifications of the function and determining the output based on them.
3. The execution of the test cases and the recording of the output.
4. The comparison of the actual output with the expected one in order to verify how much the first one deviates from the last. This step is fundamental to check if the system works as expected.²¹⁴

Under a practical point of view, it is possible to choose between two approaches when dealing with functional testing: the manual one and the automated one. They both have

²¹³ GeeksforGeeks (2023a)

²¹⁴ *Ibid.*

pros and cons, and, as it is often the case, the choice depends on the type of application being tested and, more specifically in this case, on what the developer(s) need(s) to test. In functional testing, there are several factors and aspects that come into play and need to be considered when selecting the type of testing. The following table summarizes these characteristics.²¹⁵

	Manual testing	Automated testing
<i>Accuracy</i>	Less accurate as it is conducted by humans, but handles tests that require human thinking and judgments better.	More accurate when repeated tasks need to be tested, but less accurate for tasks that require human thinking and judgments.
<i>Cost-efficiency</i>	Less expensive: no need to spend on automation tools and processes.	Low cost for predictable tests, high cost for automation tests.
<i>Test coverage and scalability</i>	Good for smaller and non-complex scenarios, in particular those that need human intervention.	Suitable for a wide range of scenarios, included large and complex ones and repetitive tasks, as long as human intuition and judgment is not needed.
<i>Test cycle time</i>	The development of test cases requires a lot of time	Faster.
<i>User Experience</i>	Effective as it requires judging the feelings and perceptions of the users.	Ineffective.

²¹⁵ GeeksforGeeks (2023b); McMillan (2023)

Human resources/skills

No great skills required, anyone can test.	Programming skills required because using the tools requires trained people.
--	--

Table 12 Manual VS Automated testing

To sum up, manual testing is more suitable for testing functionalities, UI/UX, application behavior and user acceptance and is particularly indicated for short-term projects in which flexibility is required and end-user usability holds a high role. One of the main concerns in this project is usability indeed, the required functionalities are not many and the scenarios to be tested are not particularly complicated. For this reason, manual testing was chosen, also considering the fact that some type of automated testing is already performed thanks to the unit tests integrated in the libraries used for the development of the code of the application. Unit tests is a piece of code “that verifies the accuracy of a smaller, isolated block of application code, typically a function or method. The unit test is designed to check that the block of code runs as expected, according to the developer’s theoretical logic behind it.”²¹⁶

In order to perform an effective testing, a structured approach is required. There are some phases that are generally planned before both automated or manual testing but, once again, everything depends on the specific project. In general, these steps are:

1. the understanding and analysis of the functional requirements of the software;
2. the creation of a detailed test plan based on these requirements, the identification of the functionalities that need to be tested, the definition of the testing environments and tools required and, in case of automated testing, the selection of a perfect test automation platform that fits your development pipeline;

²¹⁶ AWS (n.d.)

3. the identification of detailed test cases that cover all the functional aspects of the software;
4. the creation of test data that will be used for testing;
5. the execution of the test cases according to the test plan;
6. the comparison of the actual test results with the expected test results;
7. the creation of a detailed test report outlining the test cases' status.
8. the fixing of the reported defects;
9. regression testing, e.g., the repetition of the tests to ensure that the bugs are solved.²¹⁷

The manual testing of CHIt was quite in line with all these points. First of all, functional requirements were identified. The application is supposed to:

1. give access to the collections from the homepage;
2. show the general description of the project and the providers;
3. show all the resources at once;
4. show the thematic itinerary dedicated to Germanic culture in Italy and all its subcategories;
5. show information on the integrated tools (Pundit and EVT);
6. show the contact page;
7. filter resources according to: provider, themes and show only the filtered results;
8. allow for the use of Pundit;
9. open XML files with EVT;
10. give access to all the external links;
11. show possible images in the box above the *Metadata* section;
12. when indicated, refer to the related records from the *Metadata* section.

²¹⁷ TestSigma (2023)

Given the fact that the requirements are not a huge number, the methodology chosen for the testing of them has been a straightforward use of the application in order to explore all the expected features. With reference to the last list above, resulted that:

- points from 1 to 7 work smoothly;²¹⁸
- the use of Pundit, point 8, created no problem since it is an extension that has simply to be downloaded and enabled. Once the log-in is done, the annotator works perfectly;
- EVT: see section 3.2.5;
- it is not possible to manually test all the available links in the platform, since the resources are thousands and each of them has at least one link (that redirects to the original source). In particular, the doubt persists for the Europeana resources, as the others are manually inserted and their operativity is guaranteed. Anyway, users are invited to report for non-working links;
- point 11 works only for Europeana, but the integrated image-viewer works for the tested items. However, images can be stored in different parts of the Europeana API call and in some cases only the cover one is retrieved. Zenodo resources are visualized on the browser and GitHub ones are opened directly in the repository they belong to. This threefold way of image visualization will be object of revision;
- last but not least, point 12 could be implemented only for the resources added manually to the backend since it allows for the connection among resources. It works smoothly.

To sum up, the manual testing proved efficient and satisfactory, and the few errors found could be detected and fixed quickly.

²¹⁸ See sections 3.2.3.1 and chapter 7 for further considerations on point 7. In fact, the actual logic behind the creation of collections has brought forward some considerations.

SECTION III: USING THE APPLICATION: CASE STUDIES APPLIED TO THE GERMANIC PHILOLOGY FIELD

6. Application of CHIt to the Germanic cultural heritage: three case studies

As already stated, one of the main concerns that have guided the conception and implementation of CHIt was the idea of creating a tool that could be suitable for a varied audience with different levels of expertise in the navigation of digital platforms. Thanks to the user-friendly interface of Muruca and to the annotation and visualization tools that can be used to support the navigation in the aggregator, this goal seems to have been reached in a satisfactory way.

Together with a more general kind of consultation, as can be the case for a user who is looking for some generic information, e.g., a tourist, also academic material relevant for a scholar can be found among the CHIt resources.

The aim of this chapter is to present three case studies with the goal of showing the potential use of CHIt as a tool used for different research purposes. The field chosen for this experiment is the Germanic cultural heritage in Italy, and the three users into question are a student, a group of tourists and a Germanic philologist. To non-experts, the Germanic tradition and people may seem far from the domain of the Italian cultural heritage and therefore not well-suited for the platform exposed here and the aforementioned study. However, Italy has had contact with Germanic people ever since the early middle ages. In fact, some tribes of the Western and Eastern branches of the Germanic family – namely the Gothic and Lombard people – came to Italy and settled in some parts of the peninsula already in the 5th century.²¹⁹ Unfortunately, these populations are still labelled “barbarians” and “invaders”, and many people miss to acknowledge the immense inheritance they left to Italy.²²⁰ Eventually, these people fused with the local population, and their genealogical heritage faded. However, archaeology, toponymy, anthroponymy and the juridical system clearly testify their long

²¹⁹ More information on the origin and movements of these people is found in the next section.

²²⁰ Gasparri (2014)

dwelling among the local communities of Italy.²²¹ This variegated heritage has not to be forgotten or lost, but instead preserved and valorized among the modern inhabitants, since it is a crucial part of the State's history.

Moreover, some centuries later, other Germanic people coming originally from Alemannia and Bavaria settled in the Northern part of Italy, and started to dwell all along the Alpine arc. These populations are still there and they have preserved, to some extent and to different degrees, their cultural and linguistic heritage, enriching the varieties of those border places and, by extension, of all Italy. The hidden and harsh valleys and remote territories in which they established their small villages preserved the culture and language,²²² and nowadays they are protected language minorities still spoken in some of these areas and many cultural associations were born to preserve their identities. These institutes organize or support language and university courses, educational activities in schools and museums, the production of journals and dictionaries and cultural meetings to spread and increase awareness of these traditions. More information on this topic can be found in Fabbris (2024), where on occasion of the conference *Small languages, big ideas* held in Uppsala (13-14/06/2024) a poster on minority languages in Italy was presented.

The material related to the Germanic cultural heritage that is available on CHIt comes from all the different providers and, as seen in section 4.2.2., also external resources have been manually added to the backend, and the nature of these resources is highly different. From simple pictures to digitized manuscripts, from dictionaries to the metadata retrieved from the OntoVE SPARQL Endpoint.

For the case studies at hand, the users and their specific topics chosen are:

²²¹ Francovich Onesti (2007); Christie (1999)

²²² These languages have preserved some Germanic linguistic features that are lost in the other modern cognates, becoming precious for scholars that aim at reconstructing older stages of the Germanic languages (Angster, Benedetti, and Geyer 2014, sec. Introduzione/Einleitung).

1. a student attending a foreign languages high school who has to prepare a short paper on languages spoken in Italy with Germanic origin;
2. an educational trip for graduands in archaeology for their course on medieval archaeology. Under this perspective, the case study is targeted towards cultural tourists;
3. a Germanic philologist who is researching on the Edict of Rothari, a juridical text that belongs to the collection of the *Leges Langobardorum*.

The case studies are presented in detail respectively in sections 6.2.1, 6.2.2 and 6.2.3, dedicated to the explanation of the experimental researches conducted to test the potential of CHIt in this field under both an academic and non-academic point of view. On the other hand, the following section offers an overview of selected resources related to these topics that are available in some digital format on the web. Some of these are found in the providers of CHIt, while the others have been added to the backend.

While the free search for keyword is always available, the resources belonging to this field have been organized also into dedicated collections and specific paths to facilitate the exploration for the user. In particular, the collections are accessible directly from the homepage of the application, as seen in section 5.1. On the other hand, the dedicated paths are in the section on the top-right called *Percorsi > I Goti e i Longobardi* and *Le isole linguistiche* (Paths > Goths and Lombards and Language Islands). Here, users can read a general introduction to the topics with scientific references and access the selection of related items.

6.1. The Germanic cultural heritage in Italy: available digital resources

The Germanic cultural heritage in Italy is related to two different periods: the settlement of Lombard and Gothic people during the Middle Ages and the still current presence of

populations that came from the Southern parts of Alemannia and Bavaria starting from the beginning of the first millennium.

The original settlement of Gothic people is still matter of debate. Some scholars believe that they moved from Scandinavia,²²³ but this theory does not seem to find confirmation in the oldest sources. In fact, Greek and Roman authors attest that their origins can be found on the mainland.²²⁴ What everyone agrees upon is that in the 2nd century, they reached the Vistula, and from there they moved East. After the 3rd century, they split into different tribes, and those who reached Italy were the Ostrogoths and the Visigoths. In particular, the Ostrogoths established a kingdom in Italy between 493 and 553, after they defeated Odoacer, who in turn deposed the last Western Roman Emperor in 476. On the other hand, the Visigoths arrived in Italy earlier, but established peacefully, until they sacked Rome in 410, after which they left Italy.²²⁵

The Gothic language belongs to the Eastern branch of the Germanic family, but it extinguished during the 6th century with the fall of the Ostrogoth empire in Italy.²²⁶ Its decline started quite soon, yet there are some documents that allowed scholars to reconstruct its linguistic aspects and this is crucial because Gothic is the only attested language of this branch, as well as the oldest written testimony of a Germanic language (besides the earliest runic inscriptions, which are however often too short to gain insights into the earliest stages of Germanic languages). The most famous and longest

²²³ Robinson (1992, 43); Mierow (1908, 7). The sources for this theory are mainly historiographies (Jordanes' *De origine actibusque Getarum sive Gotorum* and Procopio's *Bellum Gothicum*) and the presence of Gothic-Northern isoglosses that would suggest that their original place is Scandinavia (cf. Francovich Onesti 2011, 69–70 and Scardigli 1964, 48–49). However, the scarce number of isoglosses and the absence of support from archaeology tend to weaken this option (Saibene and Buzzoni 2006, 38, 46).

²²⁴ Dolcetti Corazza and Falluomini (2020)

²²⁵ Saibene and Buzzoni (2006, 46–47)

²²⁶ Francovich Onesti (2011, 80). However, a more recent variety of Gothic was recorded by the Flemish diplomat Ogier Ghielin de Busbecq in the 16th century. During a stay in Constantinople, in 1562, he came into contact with some representative of the Gothic of Crimea, and he recorded around eighty words and the lyric of a song that were afterwards published in one of his letters in 1589. See Fulk (2018, 3:20) for further references.

witness is the translation of the Bible commissioned by the bishop Wulfila in the 4th century.²²⁷

The other Germanic tribes that closely wove their path with the Italian history were the Lombard people, a group of Germanic tribes that from the 2nd to the 5th century moved from their original territories along the Southernmost part of the Elba river.²²⁸ Their migration lasted some centuries, and in this period of time they went through Germany, Hungary and eventually Italy in 568, under the guidance of Alboin. However, their occupation was not particularly violent, and not for political reasons.²²⁹

At the beginning, the Lombards remained rigorously separate from the local population and gave birth to numerous independent duchies. Only later, thanks to, among the other things, the promulgation of laws written in Latin and to the conversion to Catholicism, they got integrated within the local people and a new culture rose.²³⁰

However, since the very beginning of their occupation of Italy, the Lombards had to fight against the Byzantine empire, which ruled over the peninsula at that time and had recently reestablished its power after a long war against the Ostrogoths. Nonetheless,

²²⁷ The earliest manuscripts available are of the 5th/6th centuries.

²²⁸ Again, historiographies (Jordanes, Paolo Diacono's *Historia Langobardorum*, *Origo gentis Langobardorum*) suggest that Lombards as well originally came from Scandinavia, but archaeological findings do not confirm this hypothesis (Menghin 1985; Christie 1999).

²²⁹ Gasparri (2014, 4); Zanella (1991)

²³⁰ Francovich Onesti (1999) writes that it is excluded to think that in the 8th century Longobard language was still spoken in Italy, except for some expressions used by a minority. Moreover, the author suggests that in the 7th century, the Lombard people could have been still bilingual (Latin being the other language), at least in some social classes, and that their language definitely died with the conversion to Catholicism by the end of that century. In the Edict of Rothari – belonging to the earliest law code, but see section 6.2.3 – entries of juridical nature in their native language seem to be inserted as memory and national coloring, to recall traditions, national identity and monarchy (*ibid.* 48). On the other hand, Haubrichs (2010) supports a thesis according to which in the mid-8th century this language was still spoken, as proven by the phonetic evolution of some nouns. Francovich-Onesti (n.d.) believes that there are not enough proofs to support this thesis and confirms hers of 1999.

the Lombard people founded an empire that lasted until 774, year in which they were definitely defeated by the Franks of Charlemagne.²³¹

As far as the Longobard language is concerned, not much is known, unfortunately. From the few sources at our disposal, most scholars have argued that it belongs to the Western branch of the Germanic family, and in particular to the group of languages characterized by the Second Consonant Shift (SCS), which affected the German varieties below the Benrath line (a former village that is now a district of Düsseldorf).²³² This change conditioned Germanic stops, which reflected as affricates or fricatives according to the phonetic environment, and voiced stops turned into voiceless stops in all the environments. Similarly to the First Consonant Shift,²³³ the stops were not affected by the shift in consonantal clusters such as /sp, st/. The realization of this phenomenon is not regular, unlike, for example, the shift of /d/ into /t/, but a detailed description of the isoglosses belonging to the SCS and to what extent they influenced the different dialects can be found in Fulk (2018) and Saibene and Buzzoni (2006). Other isoglosses support the identification of the Longobard language with the Western branch of Germanic languages,²³⁴ while the characteristics that would put it into continuity with Gothic belong to a period following the origin of the language, but preceding the Lombard placement in Italy. These aspects faded and the language incorporated the linguistic tendencies of High German, following its natural predisposition given the common

²³¹ Gasparri (2021). However, the Lombard duchies of Spoleto and Benevento remained independent for a longer period.

²³² The SCS consists of a series of linguistic phenomena that characterize a subgroup of Germanic languages, the Old-High German dialects. It did not take place homogeneously, but its effects can be observed to different extent in different territories. It presumably affected Upper German dialects first and spread northwards after, but bringing along fewer sounds. However, other theories postulate different movements of the shift, cf. Fulk (2018, 3:135–36) and Saibene and Buzzoni (2006, 164–65).

²³³ Fulk (2018, vol. 3, chap. 6.4)

²³⁴ The result of the vowel Germ. *ē₁ > Long. [a], Germ. *b > Long. [b], Germ. *ō > Long. [d] just to cite a few. See Francovich-Onesti (1999) and (n.d.) for a more detailed analysis.

origin.²³⁵ Moreover, the presence of a specific phenomenon, rhotacism, clearly categorizes Lombard as a Germanic language of the Western branch instead of the Eastern one.²³⁶ The studies on the Lombard languages are based on anthroponyms, the toponymic heritage and the common lexicon that has endured throughout the centuries in the Italian language.²³⁷ For this reason, no evidence regarding syntax has come down to us.

Even though the Lombard and Gothic linguistic evidence is scarce, their tangible and intangible heritage in Italy is not. As already mentioned, Latin texts written by these people have survived, with important information on different aspects of their lives. Besides that, they have left behind them architectural finds and, as far as Lombard people are concerned, many objects have come down to us, from weapons to combs. All these findings need to be preserved before human activity or natural phenomena might eliminate them forever, as stated at the beginning of this dissertation (section 1.2).²³⁸

²³⁵ Francovich Onesti (1999, 46). Lombard, Alemannic and Bavarian people (these last two speaking Old High German) are grouped under the term *Irminones* by Tacitus in his historical and ethnographic work *Germania*.

²³⁶ Rhotacism is the development of Germ. */z/ into /r/ and is absent in the Eastern branch (Fulk 2018, 3:108).

²³⁷ Francovich Onesti (1999). Moreover, Italian is the second romance language with the highest number of loan words from Germanic languages, and Lombard loans are the majority of these (Leonardi and Morlicchio 2009).

²³⁸ As far as the Lombards are concerned, the sources that provide linguistic material are laws, historiographies, glossaries, historical-literary works (e.g., the lives of Saints), notary deeds and inscriptions on different materials, epigraphs and graffiti (Francovich Onesti 1999). For the Goths, the most relevant source is the translation of the Bible by Wulfila, as mentioned above. Besides that, there are three religious texts of which only few pages have survived (a Gospel, the Old Testament and a collection of Latin homilies with Gothic glosses), a fragment of a church calendar, two deeds, alphabets, a possible toast, personal names found on Latin and Greek texts and runic inscriptions of doubtful origin (Robinson 1992; Zironi 2022). Moreover, it has to be pointed out that all the Gothic sources are found in the places and in the period of the Ostrogothic dominion in Italy .

The second group of Germanic people in Italy is the one of the Alemanni and Bavarians who migrated from their original territories in the period that goes from the 10th to the 13th century. The facts in this case are extremely different in comparison to the aforementioned Gothic and Lombard migrations and occupation of Italy, since these Southern Germanic people came to the peninsula peacefully, and peacefully they dwelled and are still dwelling.²³⁹ Their languages and cultures are now integral part of the communities they belong to, even if the social openness of the last decade (tourism, alphabetization and consequent increase in the use of the Italian language, relationships with external companies, etc.) and, not last and previous to that, Fascist regime, has put them into real danger of extinction.²⁴⁰ A good step forward for their recognition and preservation from a national point of view was the emanation of the law 482 in 1999,²⁴¹ which safeguards the languages and cultures of minority populations in Italy. Of course, this legislation alone would have done quite little for the preservation of these cultures and languages under a practical point of view, and it is here that local organizations, institutions and associations have made and are making a difference. In fact, social engagement has grown in the last decades and new bodies have been established that help go towards the direction of revival and safeguard of these cultures.²⁴²

Geographically speaking, the Germanic linguistic islands are spread all over the Alpine arc. They occupy five out of the eight northern regions of Italy: Aosta Valley, Piedmont, Veneto, Friuli-Venezia Giulia and Trentino-Alto Adige. There are six linguistic groups, namely Walser, Cimbrian, Mòcheno, Sappadino, Saurano and Timavese. It has to be pointed out that the first two are composed of different varieties, though, while the

²³⁹ Angster, Benedetti, and Geyer (2014, sec. Introduzione/Einleitung)

²⁴⁰ In fact, the Fascist regime aimed at reaching linguistic autarchy. The regime supported Fichte's theory according to which unitary language equals to a unitary nation. For this reason, xenophilia, together with toponyms and anthroponyms not of Italian origin were suppressed, bilingual schools were closed, print and publishing were invited to deliver Italian or at least Italianized versions of foreign words, and mass communication means were largely used to spread this approach (Rametta 2003; Tommasini 2021).

²⁴¹ Italian Government (1999)

²⁴² Comitato Unitario (2022, 15)

others are more homogeneous since they are present in smaller areas. Figures 20 and 21 show their distribution in the different regions.²⁴³

²⁴³ The map is taken from the website of the United committee of the historical Germanic linguistic islands in Italy (Comitato unitario delle isole linguistiche storiche germaniche in Italia) <https://www.isolelinguistiche.it/it/cartina.html>



Figure 20 Germanic linguistic islands in North-Western Italy.

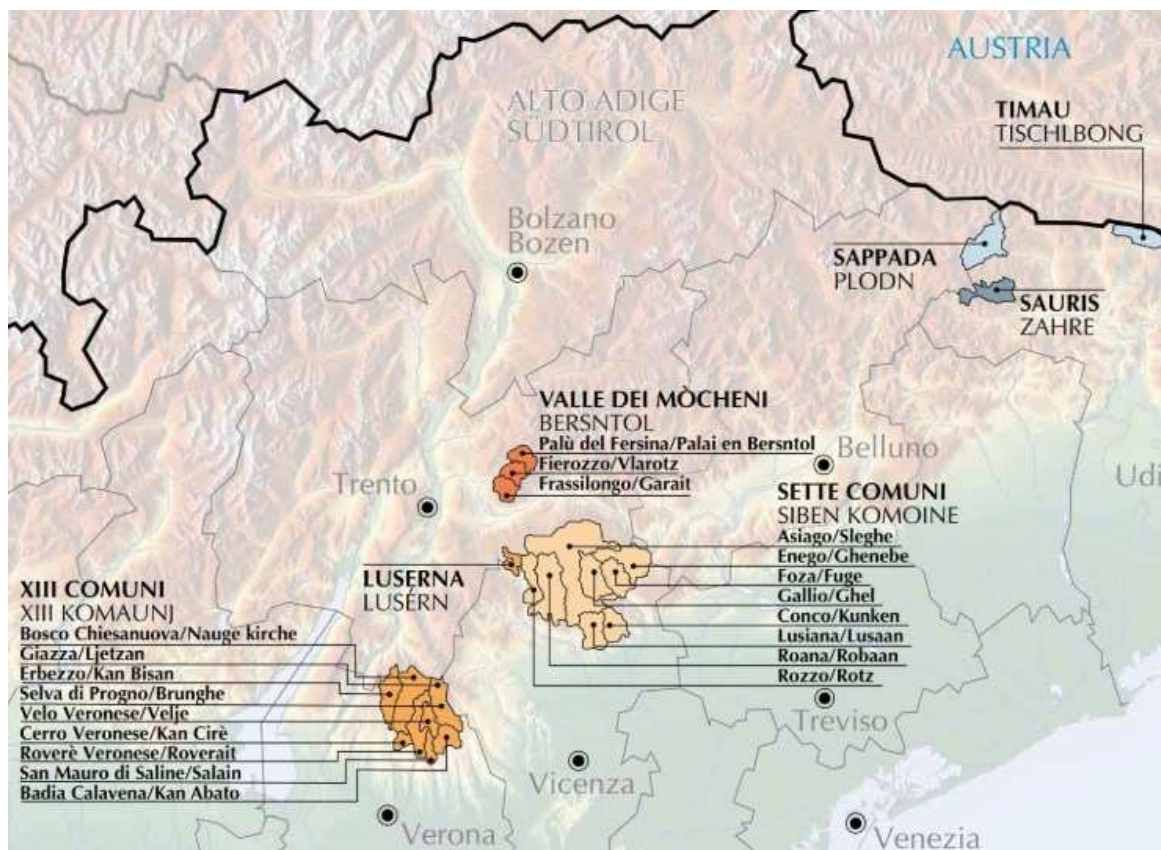


Figure 21 Germanic linguistic islands in North-Eastern Italy.

In table 13 below, there is a list of websites that preserve elements of the Germanic cultural heritage in Italy to some extent. This list is an excerpt of what was found during the three years of development of this project and is not meant to be exhaustive either referred to CHIt nor, even more so, for the web in general. The four providers described in section 4.2.1 are omitted because they are an integral part of the platform and offer single and specific resources, being miscellaneous repositories. The items of this table are all part of the CHIt backend and are therefore integrated in the platform, even if not in a FAIR modality.²⁴⁴

	Website	Short description	Link
<i>Medieval populations</i>			
Lombard people	Vatican Apostolic Library	The digital catalogue contains a huge number of digitized manuscripts, among which the most important one preserving the <i>Leges Langobardorum</i> .	https://digi.vatlib.it/
	Lombardische Lehnwoerter im Italienischen (Lombard-Italian Dictionary)	Dictionary from Lombard to Italian, 2015, Bigazzi Giovanni	https://www.koeblergerhard.de/BigazziGiovanniLombardischeLehnwoerterimItalienischen2015.htm
	Paesaggi longobardi (Lombard landscapes)	Project headed by the AReS association with the aim to offer a general view of the Lombard presence in the different Italian regions.	https://www.paesaggilongobardi.it/index.html

²⁴⁴ In fact, as already explained in section 4.2.2, the resources added manually to the back end are interoperable only within CHIt.

Gothic people

Longobardi in Italia (Lombards in Italy)	Website for the valorization of the Lombard culture in Italy, in particular of the seven sites that are part of the UNESCO heritage.	https://longobardinitalia.it/
e-codices	Website that contains thousands of digitized manuscripts from Switzerland, among which many law codes.	https://www.e-codices.unifr.ch/it
Titus Project	Database with texts annotated. Among the others, the Bible in a synoptic edition with Greek and Latin and the Gothic Palimpsest from Bologna.	https://titus.fkidg1.uni-frankfurt.de/index.exe.htm
Wulfila project	Small digital library dedicated to the study of the Gothic language and Old Germanic languages in general. It contains linguistically annotated texts.	http://www.wulfila.be/gothic/
Gotica	Project for the Gothic literature and writing, of particular interest is the Skeireins section.	https://www.gotica.de/

*Language islands***Walser**

Walser Cultura	Database that gathers, makes available and valorizes the Walser cultural heritage.	https://www.walser-cultura.it/
Centro Studi e Cultura Walser / Walser	Cultural center for the promotion and safeguard of the Walser language and culture of Gressoney and Issime (Titsch and Töitschu).	https://www.centroculturalewalser.com/

Cimbrian

Kulturzentrum di Gressoney e Issime		
Associazione Augusta Issime	Association that promotes the activities held to enhance the society and improve the life of Issime.	https://augustai ssime.it/
Gruppo Walser Campello Monti / Walsergemeinschaft Kampel	Website under construction.	https://www.walserkampel.it/
Centro Studi Walser Rimella	Website for the dissemination of the old culture of Rimella.	http://www.centrostudiwalserrimella.it/
Cimbrian institute of Luserna	Body of the Autonomous Province of Trento for the safeguard, promotion and valorization of the ethnographic and cultural heritage of the Cimbrian of Luserna.	https://www.istitutocimbri.it/
Institute of Cimbrian culture of Roana	Cultural institute for the research, documentation, preservation and valorization of the history and development of the Cimbrian culture in the Altopiano of the VII municipalities.	https://cimbri7comuni.it/
Cimbrian association of the Cansiglio	Cultural association of social promotion of the Cansiglio forest.	http://cimbridelcansiglio.it/

Mòcheno	Cimbrian association of Verona	Cultural association that protects and spreads the Cimbrian language of the XIII municipalities.	https://www.cimbri.it/
	Mòcheno Cultural institute / Bersntoler Kulturinstitut	Cultural institute for the protection and valorization of the Mòcheno ethnographic and cultural heritage.	https://www.bersntol.it/
	Mòcheno media library	Platform with linguistic and cultural resources for the spread of Mòcheno.	http://mediateca.kib.it/home.page
Sappadino	Plodar association	Association for the preservation and development of the local culture of Sappada.	https://www.plodn.info/
Saurano	Sauris cultural club	Cultural club for the preservation and promotion of the language and culture of Sauris.	https://www.sauris-zahre.org/
Timavese	Taic in Vriaul – Germanic-speaking culture in Friuli	Cultural organization for the preservation and promotion of the language and culture of Timau.	https://www.taicinvriaul.org/timau/ita/indice.html

Table 13 Excerpt of the websites that preserve Germanic digital heritage to some extent.

Even though there are some websites that make available digitized versions of this material, it is evident that all of them are scattered throughout the web and the simultaneous use of these resources or a systematic search are extremely difficult, if not impossible. Therefore, this specific subset of Germanic heritage seemed a good candidate to test the functionalities of CHI. In fact, thanks to this application, these materials are aggregated in one single platform and consulting or working on them is no

longer a problem. In this way, it is possible to exploit the potential of already existing resources and to valorize the work done by other institutions or individuals, efforts that risk to go unnoticed without the correct interoperability among the data. Section 6.2 shows the practical application of these principles and methodology guiding the construction of CHIt, hopefully proving that it is difficult to work on these resources without a tool such CHIt and that, as a consequence, this aggregator is a valid help for scholarly and non-scholarly purposes.

6.2. Some use cases with CHIt

This chapter concludes with three case studies which aim at testing the actual usability of CHIt as a scholarly and non-scholarly tool and valuable help for doing research performed by different types of users. The three case studies are presented separately in the three following subsections.

6.2.1. The student

The first case study deals with a student attending a foreign language high school in Italy who has to prepare a short paper to present to his/her German teacher. The aim of his/her work is to produce a lexical comparison of nouns in German and in the different Modern Germanic varieties of Italy belonging to a semantic field chosen by the student.²⁴⁵ After that, examples taken from real texts have to be provided in order to see the use of this lexicon in context. The student chooses as semantic field *nature*, hoping to find a good amount of resources given the fact that the Germanic populations in Italy that are still present dwell among the mountains.

²⁴⁵ The aim of this section is not a linguistic analysis of these varieties. Many books and articles have been written on the topic, but some bibliographic references are Bidese (2024); Saracco (2024); Sidraschi (2024); Caria (2024); Angster (2014); Zuin (2024), and Cognola (2024).

The student decides to start his/her research directly from the CHIt dedicated path at *Home > Percorsi > Le isole linguistiche* (Home > Paths > Language islands), where he/she can find general information on the topic and the first links that are common to all the language islands. In particular, there is the link to the United committee of the historical Germanic linguistic islands in Italy (*Comitato unitario delle isole linguistiche storiche germaniche in Italia*)²⁴⁶ and to two linguistic tools – AlpiLink²⁴⁷ and CLiMAIp²⁴⁸ – which both aim at preserving language minorities in Northern Italy.

From this sort of presentation page, the student has the possibility to investigate the resources available on various websites through further pages dedicated to the single languages, found at the bottom (cf. section 5.1, figure 15).

The resources are here presented with their most generic metadata; for deeper fruition, the user should leave CHIt and consult directly the external websites.

The first thing that he/she has to think about are the words to be analyzed. The examples provided here are with the words for tree (Modern German *Baum*, Middle High German *boum*²⁴⁹) and woods (Modern German *Wald*). For the operative phase, he/she goes through all the different sections of the Germanic islands as presented at the bottom of the aforementioned page of CHIt. In fact, each of them has a section dedicated to the linguistic resources, and the student is particularly interested in the dictionaries for the first stage of his/her project. From each dictionary, he/she can easily find the desired entries and at this point he/she starts to take advantage of Pundit. After he/she activates it on his/her browser, he/she creates a notebook dedicated to his/her project, where all the identified resources can be saved, organized and retrieved.

As successive step, he/she decides how to annotate the dictionary entries and opts for a comment with the German and Italian translations and a set of tags that will help him/her in visualizing the items as he/she needs them for his/her analysis. In particular, he/she creates tags for the Italian word, the type of entry (*entry*, It. *voce* in this case),

²⁴⁶ <https://www.isolelinguistiche.it/it/>

²⁴⁷ <https://alpilink.it/>, Kruijt and Rabanus (2025)

²⁴⁸ <https://www.climalp.org/>

²⁴⁹ <https://woerterbuchnetz.de/?sigle=Lexer&lemid=B03589>

the language variety and, in the case of Walser and Cimbrian, which are spoken in different areas, the specific place. Figure 22 shows an example of this kind of annotation with an entry from the Saurano dictionary.

The screenshot displays a web interface with a top navigation bar containing letters G, g, H, I, J, K, KH, L, M, N, O, P, PF, and a profile icon. The profile icon is labeled 'gfabbris8181@gmail.com'. Below the navigation bar, the main content area is divided into two columns. The left column features a dictionary entry for 'bolt' [bɔlt] with its grammatical forms and meanings in German. The right column shows a user profile for 'gfabbris8181@gmail.com' with a post dated '19/06/2024' titled 'in German thesis'. The post content includes the word 'bolt' and a definition 'Wald, bosco'. Below the definition, there are three tags: 'voce', 'bosco', and 'saurano'. At the bottom of the post, there are icons for comments, likes, and shares, all showing zero counts.

bolt [bɔlt]
n.m., dat.sg. *bolde* [bɔlde],
pl. *belder* ['bɛldər] - Wald;
bosco, foresta.
†*beldeſaubar*,
†*ozongenbolt*, †*eibnbolt*

Suggerimenti per la
A causa della semplicità
lessico saurano, alcune
parole potrebbero non t
alcuna corrispondenza
tre altre potrebbero e
tradotte in modo indi
con la locuzione.
Per una ricerca più ap
dita si suggerisce per
utilizzare il carattere j
(asterisco).
Esempio.

gfabbris8181@gmail.com
19/06/2024 · in German thesis

bolt
Wald, bosco

voce bosco saurano

0 0 0 0 0

Figure 22 Example of entry annotation

After that is done, he/she starts to look for the words in context. he/she goes back to the list of websites available on the CHIt single pages and goes through them patiently. Figure 23 provides an example by showing an excerpt of the websites available for the Walser section. The student could now click on the section *Cultura* (culture) of the *Walser-alps* website in the first section dedicated to cultural tourism.

Turismo culturale

- [Walser Cultura \(*\)](#)
 - [Beni etnoantropologici d'epoca](#)
 - [Toponimi](#)
- [Sportelli Walser \(*\)](#)
- [Walser-alps \(*\)](#)
 - [Cultura](#)
 - [Le regioni Walser](#)
 - [Musei Walser](#)
- [Centro Studi e Cultura Walser / Walser Kulturzentrum di Gressoney e Issime \(*\)](#)
- [Associazione Augusta Issime \(*\)](#)
- [Gruppo Walser Campello Monti / Walsergemeinschaft Kampel \(*\)](#)
- [Centro Studi Walser Rimella \(*\)](#)
- [Due sentieri culturali a Rimella](#)
- [Il territorio di Alagna Valsesia](#)

Media

- [Video sui Walser della Valle d'Aosta](#)
- [Fotografie - Walser Cultura](#)
- [Galleria fotografica Walser del comune di Alagna Valsesia](#)
- [Museo virtuale](#)

Figure 23 Excerpt of the websites available for the Walser section

Once he/she finds the desired example(s), he/she activates Pundit on the (external) page and adds the tag for the type (*context*, It. *contesto* in this case) and again Italian translation, variety and location if needed, as shown in figure 24.

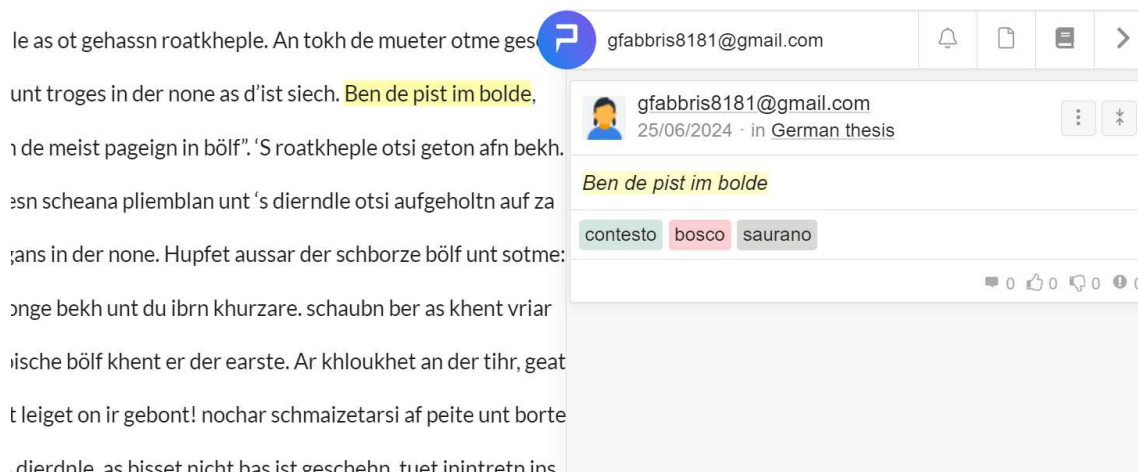


Figure 24 Example of entry-in-context annotation. Please consider that the web page visualized here is outside CHIt.

Now that all the data are retrieved, he/she can compare them in his/her Pundit notebook. His/her scope is to investigate the differences among these language varieties, and for this reason he/she selects the filter for a single word, say *tree* (It. *albero*) and starts analyzing the results going through the list. This visualization is represented in figure 25.

Home

Notebooks

Notifications

Home > Notebooks > German thesis

albero x

gfabbris8181@gmail.com

June 19, 2024 · in German thesis

palm

Baum, albero

voce albero timavese

gfabbris8181@gmail.com

June 19, 2024 · in German thesis

palm

Baum, albero

voce albero saurano

gfabbris8181@gmail.com

June 19, 2024 · in German thesis

palm

Baum, albero

voce albero sappadino

gfabbris8181@gmail.com

June 19, 2024 · in German thesis

palm

Baum, albero

voce albero mocheno

gfabbris8181@gmail.com

June 19, 2024 · in German thesis

palm

Figure 25 Excerpt of the Pundit notebook with the filter for tree

To ease the analysis, he/she summarizes the results in the following table.

<i>Baum, albero</i>	
<i>Walser – Gressoney</i>	baum
<i>Walser – Issime</i>	boum
<i>Cimbrian – Lusern</i>	puam
<i>Cimbrian – Seven Municipalities</i>	póom
<i>Cimbrian – XIII Municipalities</i>	póme
<i>Mòcheno</i>	pa'm
<i>Timavese</i>	paam
<i>Saurano</i>	pame
<i>Sappadino</i>	pame

Table 14 Entry summary for *Baum, tree*

From here, the student can immediately see a first difference: the Western varieties are somehow closer to the Modern and Middle High German lemma, while the Eastern ones present a substantial difference at the very beginning of the word. In fact, the Walser entries show the /b/ as in Modern and Middle High German, while the other varieties present the sound /p/. After a quick search on the web, the student finds out that between the 5th and the 9th century, a massive phonological change affected a wide area in which a variety of Western Germanic was spoken. This linguistic phenomenon, the already mentioned SCS (cf. section 6.1), sets apart what now is the German language from the other Germanic languages of the time. The specific change at hand is an isogloss that implies the shift of the Germanic voiced stops */b, d, g/ into Old High German voiceless stops /p, t, k/. If on the one hand, the realization of this isogloss is one of the most complete, on the other hand, it was also specified (cf. note 224) that the SCS faded as it spread northwards. In fact, Modern German is based on a variety of central

Germany.²⁵⁰ Therefore, the isoglosses that have remained in the modern language do not show a full realization of the shift. Since this particular change (Germ. */b/ > ModG /b/, OHG, Bavarian /p/) was not retained in Modern German, it ended up characterizing the Southern varieties, Bavarian in particular. It is therefore interesting to notice that the German islands in Italy of Bavarian origin still show this distinctive trait.

Another evident difference is the presence of the vowel /a/ or /o/ (long or short) instead of the diphthong /au/ of *Baum*. A literature review seems to suggest that diphthongs are actually a characteristic of Bavarian and that developed already starting from the 13th century, and in this sense it seems misleading. Moreover, the etymology of this word in the history of German does not seem to show a single vowel in this place, since it comes from Germanic **baumaz* > Old and Middle High German *boum*.²⁵¹ However, a check on the *Bayerisches Wörterbuch* (Bavarian dictionary) shows many examples of the entry for Baum with /a/ or /o/:

Baum M., Baum; Balken; Stange; Sarg. 1 Baum.- la Baum, Holzgewächs, 0 Gesamtgeb. vielf.: °dea Schturm heit Nacht hot an **Boom** neba den Schtadl wurza [ganz ab] krissa Peiting SOG; [...] °Nachba, e dein Wold san gfuchste **Bam** drinn „dürr mit roten Nadeln“ Wiefelsdf BUL; af den houchn **Bama** Suffershbm WUG; Wenn da Mond auf dy Nacht steht übern **Baamern** hint Bay.Wald P. Friedl, Die Gsangl des Baumstefenlenz, Grafenau 1975, 55; a Bär... beutlt dees **Baamei** gar g'walti her Kobell Ged. 42.²⁵²

Therefore, the student concludes that the presence of this monophthong is a characteristic of Bavarian and, for sure, of the Germanic languages in Italy.

²⁵⁰ In the 16th century, during the Protestant Reformation, Luther translated the Gospels and the Bible in a middle-Eastern German (the area of Meissen), and his translation became the most spread and read book of Protestant Germany. In this way, it became the basis for what is now Modern German. It was also a good candidate to unify the language, that was divided into Northern and Southern varieties. Nowadays, in the Southern areas of Germany the remains of the SCS can be found in the dialect called *Oberdeutsch*.

²⁵¹ Orel (2003); Wright (1917)

²⁵² Denz (2020, 1422), <https://publikationen.badw.de/de/bwb/index#4632>

Examples in other Germanic languages that follow this direction, even if in different geographic areas, can be found in Old Saxon and Middle Low German *bōm*, Middle and Modern Dutch *boom*, Old Frisian *bām* Old Norse *baðmr* and Gothic *bagms*.²⁵³

Analogously, the word for wood, *Wald*, shows other interesting features, as shown in the following table.

Wald, bosco

<i>Walser – Gressoney</i>	woald
<i>Walser – Issime</i>	woald/woalda/veeld
<i>Cimbrian – Lusern</i>	balt
<i>Cimbrian – Seven Municipalities</i>	balt
<i>Cimbrian – XIII Municipalities</i>	balt
<i>Mòcheno</i>	bòlt
<i>Timavese</i>	bolt
<i>Saurano</i>	bolt
<i>Sappadino</i>	bòlt

Table 15 Entry summary for Wald, bosco

Again, the difference between the Eastern and Western dialects is evident. The Walser varieties present a consonant pattern perfectly in line with German, while the others show recursive changes. In particular, the shift from German initial sound /v/ into /b/ is a change that characterizes Germanic varieties in Italy of Southern Bavarian origin.²⁵⁴ On the other hand, the different sound at the end of the word, e.g. /d/ into /t/, is the result of another phenomenon called *final devoicing* or *Auslautverhärtung*, according to which /b, d, g, z, v/ are pronounced /p, t, k, s, f/ at the end of words (this phenomenon

²⁵³ Pfeifer (1993)

²⁵⁴ Cognola (2024, 67)

is particularly evident in Middle High German).²⁵⁵ This characteristic is present in all the Eastern varieties.²⁵⁶

As for the vowels, the biggest differences are seen in the varieties at the extreme longitude (e.g., Walser, Sappadino, Saurano and Timavese) plus Mòcheno. As expected, the changes that affect Walser are not the same that affect the Eastern languages, which show cohesion once again. In fact, after some independent research, the student easily finds out that German /a/ – Bavarian /o/ is a common linguistic feature, especially in Tyrolese and Carinthian,²⁵⁷ and this is why it is not shown in the Cimbrian varieties.²⁵⁸ Last but not least, the difference German /a/ - Walser /oa/ is once again a Walser peculiarity of the Aosta Valley widely attested, as in German *alt*, *Schaf*, *nach*, *Arm* – Walser *oalt*, *schoaf*, *noa*, *oare* (old, sheep, after, arm).²⁵⁹

As already mentioned, the second type of analysis is an in-context one, and an example has already been provided in figure 24 above. This kind of entries can be found in different sections of the bullet lists available on CHIt for the Germanic islands, since they can be scattered among materials for kids, exercise books for tourists, journals and newspapers. The interesting fact of this analysis is that the student can observe the degree of inflection of these words, since German is a language that has cases. For example, figures 22 and 24 show that the Saurano word *bolt* has sonorization of /t/ into /d/ in its dative form, resulting in /d/, *bolde*.

As it is evident, the research for sources was performed outside CHIt, but the student could easily find his/her way into the various websites thanks to the presentation pages offered in the application and the aggregation of these websites in one single space. It would have taken incredibly much more time if he/she had had to find all the references herself. The result is that he/she can write a successful school assignment with the help

²⁵⁵ Salmons (2018, 204)

²⁵⁶ Alber (2014)

²⁵⁷ Pietrobon (2020, 17); Sidraschi and Costantini (2022, 48); Zuin (2024, 155)

²⁵⁸ It is massively present also in Mòcheno: German *Vater*, *Nacht*, *machen*, *was* – Mòcheno *voter*, *nòcht* *mòchen*, *bos* (father, night, to do, what).

²⁵⁹ Romano (2010, 62)

of CHIt and of Pundit, which helped him/her in the retrieval and organization of the information needed for his/her work.

In figure 26, the steps that the student takes within CHIt to develop his/her work are shown in a schematic representation. Please consider that once on the list with the available websites (cf. figure 23), the user is supposed to leave CHIt to better investigate the single resources.

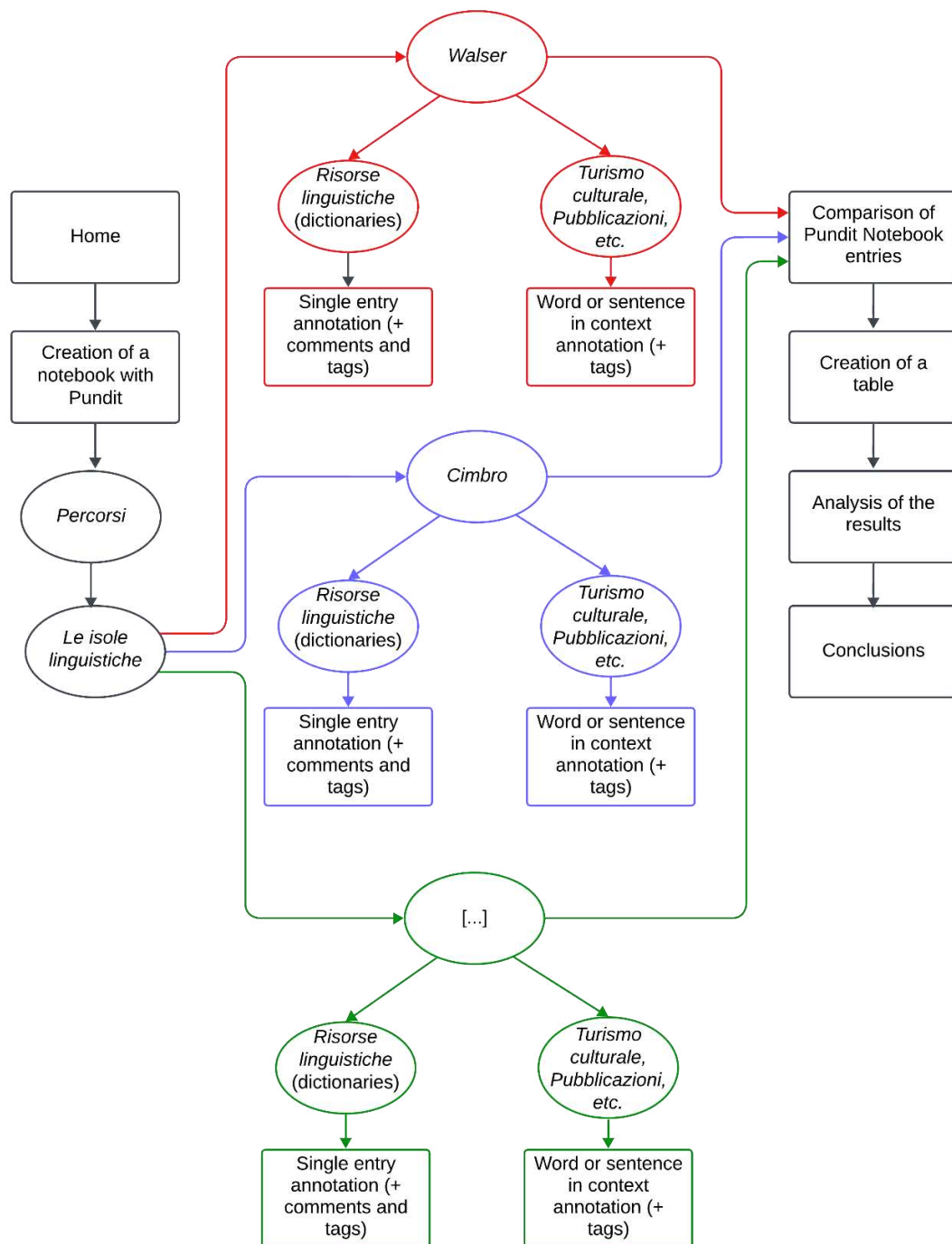


Figure 26 Student's steps

6.2.2. The tourist

The second user story has been thought of as a formative trip for a group of students from the faculty of Archaeology who are attending a course on medieval archaeology

and are asked to work on a project for the conservation and restoration of a specific heritage at the students' choice. This group decides, for convenience, to start from a place that is close to their University and agree on organizing their project as a trip of more days. Their University is based in Trieste (Friuli-Venezia Giulia) and, as university students of archaeology, they know that Cividale del Friuli played a major role during the Lombard period in Italy.²⁶⁰ It is 70 km away from Trieste, so it seems a good place where to start. This means that they are going to work on the Lombard archaeology in Northern Italy. Now, the itinerary has to be decided, and it is here that CHIt enters the scene. In this sense, the use of the application turns into the point of view of tourists with cultural interests, who need to understand what is the best tour that they can take to visit the places marked by the dwelling of these people in Italy.²⁶¹

By scrutinizing relevant scholarly resources and thanks to their knowledge of the Middle Ages, they find out that after the fall of the Western Roman Empire, four kingdoms were established in Italy: the Vandals in the major Italian islands (Sicily and Sardinia), the kingdom of the Germanic king Odoacer, which lasted only seventeen years, the Ostrogoths and the Lombards. These last two kingdoms were geographically more extended and their duration more consistent (altogether from 493 to 774 AD). In fact, much more evidence can be found on their account, and the capitals of these kingdoms were in the North of Italy: Ravenna (Emilia Romagna) and Pavia (Lombardy) respectively. The CHIt homepage helps already from the beginning, since there are both an itinerary and a collection dedicated to the Goths and Lombard people in Italy. The first one, in a similar way as the one about language islands, presents the topic from a general point of view and gives the possibility to go deeper into the two populations at the bottom of the page (figure 27).

²⁶⁰ Cividale del Friuli was the first capital of the Lombard duchy in 568 under the guidance of King Alboin and kept on being crucial throughout all the history of the reign (Zanella 1991).

²⁶¹ This case study has been carried out thanks to the collaboration of an archaeologist, L. S., who kindly made herself available in order to produce a realistic example.

I Goti e i Longobardi

I Goti e i Longobardi sono due popolazioni di origine germanica che hanno occupato vaste aree della penisola italiana (e non solo) durante il Medioevo. Appartengono al ramo dei Germani orientali i primi e occidentali i secondi. Sono popolazioni ormai estinte che hanno però lasciato testimonianze culturali che vanno preservate e studiate poiché contribuiscono a formare il tessuto culturale italiano.

Nelle due sottosezioni presenti in questa pagina, è possibile leggere un approfondimento per entrambe le popolazioni e fare riferimento alle indicazioni bibliografiche per avere ulteriori informazioni, nonché accedere alle risorse a loro collegate.

BIBLIOGRAFIA

Gasparri, Stefano e Cristina La Rocca. 2012. *Tempi barbarici. L'Europa occidentale tra antichità e medioevo (300-900)*. Roma: Carocci editore. (In particolare capitolo 5)

AA. VV. *Archeologia Barbarica*. Mantova: SAP Società Archeologica s.r.l. [Lista delle pubblicazioni](#).



Risorse
collegate



Figure 27 Presentation page of the Lombard and Gothic people

In these sections, bibliographic references can be found. Figure 28 shows an excerpt of what the user visualizes when clicking on the square in the figure above *I Longobardi*: bibliographic references and a list of the items available in CHIIt related to the Lombards.

Fortunati, Maria e Caterina Giostra (a cura di). 2019. *I Longobardi del ducato di Bergamo. Le necropoli di Fara Olivana e Caravaggio - Masano*. Mantova: SAP Società Archeologica s.r.l.

Gasparri, Stefano. 2012. *Italia longobarda. Il regno, i Franchi, il papato*. Editori Laterza.

Groppali, Riccardo (a cura di). 2014. *Sulle orme dei Longobardi a Cremona. Atti del convegno. 10 maggio 2014, Cremona*. Inner Wheel Club di Cremona PHF, Distretto Italia 206.
<https://www.innerwheel.it/documenti/libro-longobardi-inner-wheel-1145247826.pdf>

Lombardia Beni Culturali (Regione Lombardia). L'età longobarda (sec. VI - VII).
<https://www.lombardiabeniculturali.it/istituzioni/storia/?unita=01.01>

Paroli, Lidia (a cura di). 1997. *L'Italia centro-settentrionale in età longobarda. Atti del Convegno (Ascoli Piceno 1995)*. Borgo S. Lorenzo: All'Insegna del Giglio.
<https://www.insegnadelgiglio.it/prodotto/litalia-centro-settentrionale-in-et-longobarda/>

Possenti, Elisa. 2014. "Necropoli longobarde in Italia: lo stato della ricerca". In *Necropoli longobarde in Italia: indirizzi della ricerca e nuovi dati*, a cura di Possenti Elisa. 35-54. Trento: Castello del Buonconsiglio.

Risorse collegate

I Longobardi in Italia. I luoghi del potere (568 - 774 d.C.)

I Longobardi I...

Paesaggi Longobardi

Il progetto

Dizionario longobardo

Dizionario long...

Paulus Diaconus, Historia Longobardorum - Cod. Sang. 635

Codice Sang. 63...

I Longobardi in vetrina (I Longobardi in Italia)

Il progetto "Lo...

Castelseprio - Lapide Longobarda

Castelseprio-To...

Manoscritto di grande formato contenente la storia di S. Gallo, realizzato intorno al 1200 - Cod. Sang. 547

Codice Sang. 54...

Tempietto Longobardo

Il Monastero di...

Il lessico giuridico delle leggi longobarde

Tesi di dottora...

Codice diplomatico longobardo 1-2: the original charters and coeval copies written in Tuscany between AD 714 and 774

TEI XML edition...

Museo Archeologico Nazionale - Cividale del Friuli

Il Museo, che a...

Edictum Rothari (Veterum Fragmentorum Tomus III) - Cod. Sang. 730

Codice Sang. 73...

Figure 28 Excerpt of the path dedicated to the Lombard people

On the other hand, the collection is a pre-selection, as already mentioned above (section 3.2.3.1). Here, the students can find some general information on the resources like *I Longobardi in Italia* and *Paesaggi Longobardi*. The first one is the official website that presents the UNESCO sites in Italy, while the second is a review of the Lombard archaeological sites in Italy, organized on a map that shows toponyms, isolated findings, necropolises, episcopal and ducal seats. The official website reports information only on the seven sites enlisted in the UNESCO heritage, but the second one shows to the students that there are also other interesting places to be visited (figure 29).

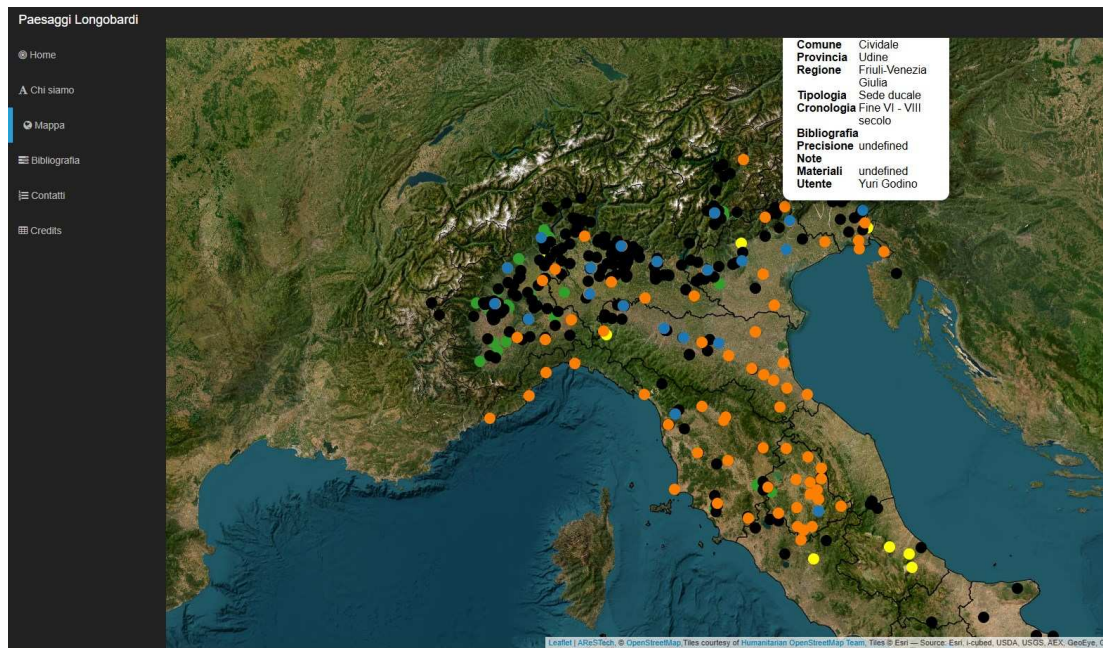


Figure 29 Map and sites available on *Paesaggi Longobardi*

Thanks to these two websites, they can identify what sites to visit and what is the best route they should take. The UNESCO website gives them the confirmation of the importance of Cividale del Friuli, which is one of the seven sites. There, they choose to visit the Lombard Temple and Monastery and the National Museum, which contains an exhibition of the Lombard culture. From the dedicated CHIt resource, they can find preliminary information, go to the external website of the museums and successively use Pundit to briefly annotate opening hours and the price of the ticket. Moreover, the students have decided to add tags indicating the day in which the various visits will take place, so to be able to organize their trip quickly. An example is provided in figure 30.



Figure 30 Example of annotation for the students' trip

After that, they decide to reach the Southernmost site of their trip. In fact, in the CHIt collection they explored at the beginning, they realized that in Ravenna (Emilia Romagna) there is an important mausoleum dedicated to Theoderic.²⁶² This city is not mentioned in the official website of the UNESCO because it is not part of the Lombard heritage, but of the Ostrogoth one instead. Thanks to CHIt they can also retrieve an interesting study related to archaeoastronomy that takes into consideration the orientation of the buildings and other architectural elements. 3D models (here in 2D format, though) too are present in the study, and further bibliographic references are available. This resource is aggregated from Zenodo. Under a technical point of view, the possibility to slightly edit the insertion of resources like in this case allows for a greater customization of the entries. In fact, it was possible to add bibliographic references to

²⁶² Theoderic brought the Ostrogoth people in Italy for the first time at the end of the 5th century. After he defeated Odoacer in 493, he established the first Ostrogoth reign in Italy with Ravenna as capital (Saibene and Buzzoni 2006, 46–47; Azzara 2013).

the set of metadata from which users can find further information and verify its accuracy or for their specific needs. For example, the students of this case study can use these entries to create their final project report and added further interesting links for their work, as the reference to a 3D model of the mausoleum (see figure 31).

SURVEY, ARCHAEOASTRONOMY AND CO THEODORIC IN RAVENNA (ITALY)

Metadati	Descrizione
	BIBLIOGRAFIA Marchi, Silvia. 2012. <i>Il Mausoleo di Teodorico</i>. Tesi di dottorato in Archeologia. Udine. https://air.uniu... Gelichi, Sauro. 2020. "Ravenna. L'Academia de Buenas Letras de Buenas Letras de Barcelona". https://racc... Guardigli Luca, Bragadin Marco Alvise e Chiara Bezzi. 2023. "Work packaging model per le ipotesi costruttive del Mausoleo di Teodorico". In <i>Transizione: sfide e opportunità per l'ambiente costruito</i>, a cura di Fabio Fatiguso, Francesco Fiorito, Mariella De Fino ed Elena Cantatore. 117-132. Monfalcone : Edicom.
Parole chiave	archaeoastronomy, architectural survey, architettura, digital museum, edutainment, Longobardi, Theodoric, Teoderico, Emilia-Romagna, medioevo, early middle ages, Ravenna
Istituzione o autore	Incerti M., Iurilli S., Lavoratti G.

gfabbris8181@gmail.com
29/07/2024 - in University trip

Work packaging model per le ipotesi costruttive del Mausoleo di Teodorico

<https://cris.unibo.it/handle/11585/964309>

Modello 3D su Sketchfab <https://sketchfab.com/3d-models/mausoleo-di-teodorico-b475d6e770104c55befaa8d80c65cea6>

finalReport

Figure 31 Bibliographic references and annotation

They also know that Theoderic had a cathedral built, the Cathedral of Saint Apollinare Nuovo, which is very close to the mausoleum. That is the last stop for the first day, and they use Pundit to annotate information on their accommodation too (see figure 32).

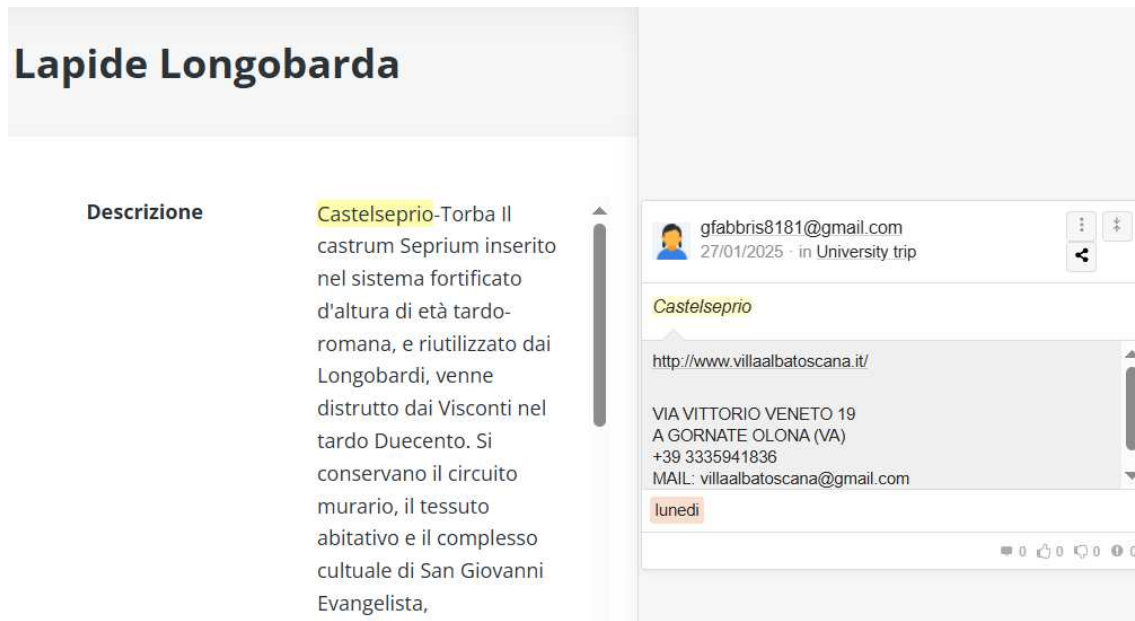


Figure 32 Annotation of accommodation

The rest of the trip is organized as shown in the following table.

University trip's schedule

<i>Monday</i>	Cividale del Friuli	Lombard Temple
		Monastery
		National Museum
	Ravenna	Mausoleum of Theoderic
		Cathedral of Saint Apollinare Nuovo
<i>Tuesday</i>	Monza	Museum and Treasure of the Dome
		Dome (Lombard tower and Teodolinda's Chapel)
	Bergamo	City Archaeological Museum

Wednesday

Brescia	Monumental complex (Museum and Monastery of Saint Giulia, Church of Saint Salvatore)
Verona	Museum of Castelvecchio
Povegliano Veronese	Archaeological Museum
Belluno	City Museum

Table 16 Students' trip

All the resources listed in table 16 have been gathered thanks to CHIt. However, further details like, for example, opening hours, ticket prices and contact should be consulted by visiting the original website (always available among the metadata of each resource provided by CHIt). As shown in this case study, this kind of information can be inserted in CHIt personally by the user in a comment (cf. figures 30 and 32).

Figure 33 shows the schema of the students' steps.

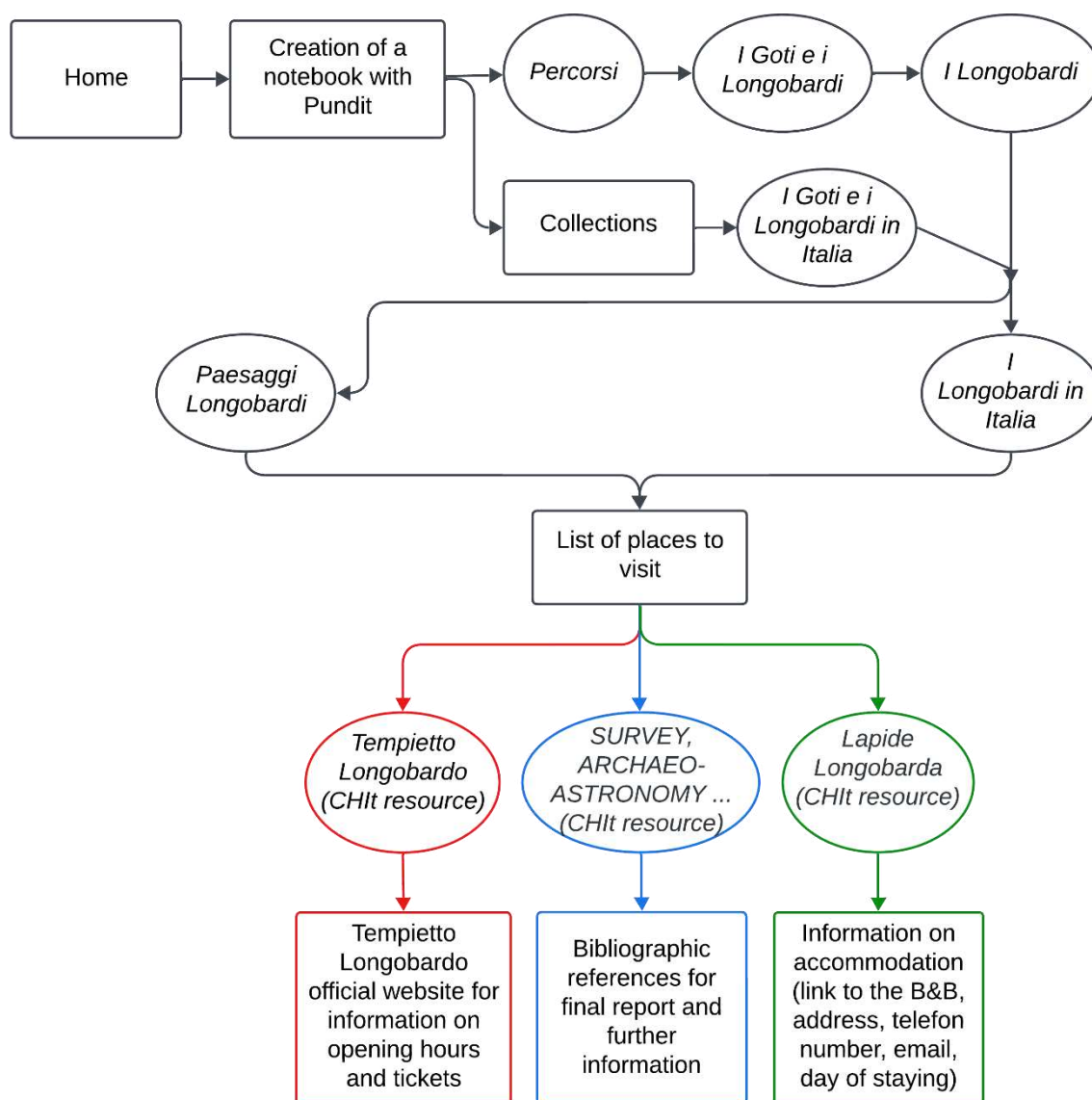


Figure 33 Students' steps

6.2.3. The philologist

The third and last case study involves a Germanic philologist as a potential user. Therefore, the research is in this case a scholarly one. This type of use will be rather different in comparison to the previous case studies, since probably an expert already knows all the information that could be found on a non-domain-specific platform such as CHIt is. However, this does not mean that the introductory articles and the resources with their metadata are of less value and scientific rigor. Nonetheless, the scholar can surely make suggestions and spot inaccuracies. The supposed use that an expert can

make of this app is to support his/her research and, thanks to Pundit, to organize the material. The case study is developed as follows.

Let's assume that the philologist is interested in medieval laws in Italy, related to the Germanic people. As said in the introductory part of this chapter, the Lombard and Gothic people inhabited Italy from the fifth to the eighth century and their dwelling left much to the society of the peninsula.²⁶³ The scholar is particularly interested in the Lombards, since their juridical activity was prominent and their law codes are almost all that is left as written source. More specifically, he is working on the Edict of Rothari, as briefly hinted above.

As already mentioned, there are both a collection and a path dedicated to the Lombard people in Italy, and the scholar might want to start directly from these. Even if the topic would be probably already well-known to the scholar, the general introduction found on the path can give him/her an idea of how the material is organized in the platform and what he/she can be expected to find. In particular, the ultimate target of this scholar is a study on the manuscript production of the Edict of Rothari.

This juridical text is the first of a series that make up the corpus of the *Leges Langobardorum*, which transmit through different witnesses the norms promulgated by the Lombard kings in the period that goes from 643 (Edict of Rothari) and 755 (laws of Astolfo). It is dated 22 November 643. It was written in Pavia and is 388 chapters long. The laws contained in it were previously transmitted orally, and fixed in written form for the first time on this occasion.²⁶⁴

The user decides to browse the section *I Longobardi* within the path *I Goti e i Longobardi*, and finds out that some manuscripts are listed. In particular, he/she sees the oldest manuscript preserving the Edict, the *Codex Sangallensis* 730, kept at the *Stiftsbibliothek* in St. Gallen, Switzerland.²⁶⁵ Moreover, the philologist can find the manuscript with the

²⁶³ Rotili (2011)

²⁶⁴ Buzzoni and Rosselli Del Turco (2015)

²⁶⁵ <https://www.stiftsbezirk.ch/it/stiftsbibliothek>. Digital facsimile of the manuscript available at <https://www.e-codices.ch/en/searchresult/list/one/csg/0730>.

signature Vat. Lat. 5359, held by the Vatican Apostolic Library.²⁶⁶ Surprisingly, he/she also retrieves the fragment of a manuscript that he had been missing during his research: Codex AA 90.28 (A), which can be found in Bern (Switzerland) at the *Burgerbibliothek* as a loose sheet in a French manuscript.²⁶⁷ In fact, the fragments of this tradition have not been studied completely, yet their importance should not be underestimated. The philologist sees the potential of CHIt, but also its limits.²⁶⁸ For this reason, he/she decides to contact the email indicated for suggestions to offer his/her help to improve the section. In particular, among the resources of the section there is the link to the GitHub repository of the digital edition of the *Leges Langobardorum*. However, the edition has been recently published as a complete actual publication in the form of a diplomatic-interpretative digital edition,²⁶⁹ and, as it still misses in CHIt, it is worth mentioning. However, had the edition not been published, users could have seen a sketched version of it thanks to EVT (figure 34).

²⁶⁶ https://digi.vatlib.it/view/MSS_Vat.lat.5359

²⁶⁷ <https://katalog.burgerbib.ch/suchinfo.aspx>. Digital facsimile of the manuscript available at <https://www.e-codices.unifr.ch/it/searchresult/list/one/bbb/AA0090-28>.

²⁶⁸ CHIt is thought of as a tool that provides generic information on the items that it gathers. Therefore, it does not allow to perform deep analysis. Moreover, being handled by a single scholar, users might find some relevant works missing. This is the reason why users are encouraged to contact the email provided in the *Contatti* section to help improve the platform.

²⁶⁹ Buzzoni and Rosselli Del Turco (2024)

Quartusdecimus agilulf, turingus, ex genere anawas. Quintusdecimus adalwald, filius agilulf.
Sextusdecimus hariwald, ex genere caupus. Septimusdecimus ego in dei nomine qui supra rotari rex,
filius nandinig, ex genere harodos.

NANDINIG filius noctzoni, noctzo filius adhamund, adhamund filius alaman, alaman filius
hiltzoni, hiltzo filius wehiloni, wehilo filius weoni, weo filius fronchononi, fronchono filius fachoni,
faccho filius mammoni, mammo filius ustbora.

EXPLICIT PROLOGUS INCIPIUNT CAPITULA CAUSAS

1. 1. Si quis **hominum**

hominum] homo B5 omit. B3

Info XML

Metadata for lemma *hominum*
No data available

MORE INFO ABOUT THE APPARATUS ENTRY

Metadata for *homo*
WIT: #B5

Metadata for *omit.*

Search Select items

mainText

Figure 34 Visualization with EVT from CHIt of the *Leges Langobardorum*

Moreover, the user believes that a wider bibliography could be implemented. For this reason, he/she suggests the addition of a couple of bibliographic references. In particular, given the fact that there are some interesting resources related to this set of laws and their importance for Italy, an academic article dealing specifically with this topic should not be missing. Buzzoni and Rosselli Del Turco (2015) is what should be added according to the expert, since in this article the reader can find a list of all the manuscripts handing down this text, together with the fragments.²⁷⁰ Besides that, the work cited comments upon different editions, and then focuses on two specific

²⁷⁰ However, the authors specify that the list they present is taken from a book dated 1868, and the fragments 1995. Therefore, the most recent manuscripts are missing.

manuscripts, those from Piedmont. They are the basis of the abovementioned digital edition. Last but not least, he/she notices that some basic references are missing. In fact, among the bibliographic entries, there is not any link to any comprehensive edition of the *Leges*. The philologist recommends adding Bluhme (1868), Beyerle (1947), and Azzara and Gasparri (2005), not much for the text, but to have an Italian translation of the work.

Thanks to these suggestions, the section of CHIt dedicated to the Lombard people changes its content from figure 35 to figure 36. It is evident how the contribution of the scholar has improved the quality of the thematic path.

BIBLIOGRAFIA

Azzara, Claudio. 2010. "I longobardi in Italia e i longobardi nella storia d'Italia". In *I longobardi del sud*, a cura di Giuseppe Roma. 79-83. Roma: Giorgio Bretschneider editore.

<https://www.bretschneider.it/libro/9788876892523>

Brogiolo, Gian Pietro. 2005 "Capitali e residenze regie nell'Italia longobarda". In *Alto medioevo mediterraneo*, a cura di Stefano Gasparri. Firenze: Firenze University Press.

<https://library.oapen.org/bitstream/handle/20.500.12657/34898/1/345120.pdf#page=233>

Brogiolo, Gian Pietro e Paola Marina De Marchi (a cura di). 2020. *I Longobardi a nord di Milano. Centri di potere tra Adda e Ticino*. IV Incontro per l'Archeologia barbarica (Cairate, 21 settembre 2019). Archeologia Barbarica, 4.

Fortunati, Maria e Caterina Giostra (a cura di). 2019. *I Longobardi del ducato di Bergamo. Le necropoli di Fara Olivana e Caravaggio - Masano*. Mantova: SAP Società Archeologica s.r.l.

Gasparri, Stefano. 2012. *Italia longobarda. Il regno, i Franchi, il papato*. Editori Laterza.

Groppali, Riccardo (a cura di). 2014. *Sulle orme dei Longobardi a Cremona. Atti del convegno. 10 maggio 2014, Cremona*. Inner Wheel Club di Cremona PHF, Distretto Italia 206.

<https://www.innerwheel.it/documenti/libro-longobardi-inner-wheel-1145247826.pdf>

Lombardia Beni Culturali (Regione Lombardia). L'età longobarda (sec. VI - VII).

<https://www.lombardiabeniculturali.it/istituzioni/storia/?unita=01.01>

Paroli, Lidia (a cura di). 1997. *L'Italia centro-settentrionale in età longobarda. Atti del Convegno (Ascoli Piceno 1995)*. Borgo S. Lorenzo: All'Insegna del Giglio.

<https://www.insegnadelgiglio.it/prodotto/litalia-centro-settentrionale-in-et-longobarda/>

Possenti, Elisa. 2014. "Necropoli longobarde in Italia: lo stato della ricerca". In *Necropoli longobarde in Italia: indirizzi della ricerca e nuovi dati*, a cura di Possenti Elisa. 35-54. Trento: Castello del Buonconsiglio.

Risorse collegate

I Longobardi in Italia. I luoghi del potere (568 - 774 d.C.) "I Longobardi i...	I Longobardi in vetrina (I Longobardi in Italia) Il progetto "Lo...	Il lessico giuridico delle leggi longobarde Tesi di dottora...
Paesaggi Longobardi Il progetto	Castelseprio - Lapide Longobarda Castelseprio-To...	Codice diplomatico longobardo 1-2: the original charters and coeval copies written in Tuscany between AD 714 and 774 TEI XML edition...
Dizionario longobardo Dizionario long...	Manoscritto di grande formato contenente la storia di S. Gallo, realizzato intorno al 1200 - Cod. Sang. 547 Codice Sang. 54...	Museo Archeologico Nazionale - Cividale del Friuli Il Museo, che a...
Paulus Diaconus, Historia Longobardorum - Cod. Sang. 635 Codice Sang. 63...	Tempietto Longobardo Il Monastero di...	Edictum Rothari (Veterum Fragmentorum Tomus III) - Cod. Sang. 730 Codice Sang. 73...
SURVEY, ARCHAEOASTRONOMY AND COMMUNICATION: THE MAUSOLEUM OF THEODORIC IN RAVENNA (ITALY) The Mausoleum o...	Cividale: gli studi ed i restauri del tempietto longobardo Foramitti, Vitt...	Cod. AA 90.28 (A) Leges Langobardorum, lat. (frammento) Foglio sciolto ...
Cod. Sang. 620 Codice Sang. 62...	Manoscritto vat. lat. 5359 Manoscritto pre...	Edizione digitale delle <i>Leges Langobardorum</i> XML...

Figure 35 Lombard path before...

BIBLIOGRAFIA

Azzara, Claudio. 2010. "I longobardi in Italia e i longobardi nella storia d'Italia". In *I longobardi del sud*, a cura di Giuseppe Roma. 79-83. Roma: Giorgio Bretschneider editore.

<https://www.bretschneider.it/libro/9788876892523>

Azzara, Claudio, e Stefano Gasparri. 2005. *Le leggi dei Longobardi. Storia, memoria e diritto di un popolo germanico*. Roma: Viella.

Beyerle, Franz. 1947. *Leges langobardorum: 643-866*. Weimar.

Bluhme, Friedrich. 1868. *Leges Langobardorum*. Hannover: Hahn.

Brogiolo, Gian Pietro. 2005 "Capitali e residenze regie nell'Italia longobarda". In *Alto medioevo mediterraneo*, a cura di Stefano Gasparri. Firenze: Firenze University Press.

<https://library.oapen.org/bitstream/handle/20.500.12657/34898/1/345120.pdf#page=233>

Brogiolo, Gian Pietro e Paola Marina De Marchi (a cura di). 2020. *I Longobardi a nord di Milano. Centri di potere tra Adda e Ticino*. IV Incontro per l'Archeologia barbarica (Cairate, 21 settembre 2019). Archeologia Barbarica, 4.

Buzzoni, Marina e Roberto Rosselli Del Turco. 2024. "Edizione diplomatico-interpretativa con facsimile digitale dell'Edictum Rothari, Vercelli, Biblioteca e Archivio Capitolare, CLXXXVIII". Edizioni Ca' Foscari. <http://doi.org/10.30687/978-88-6969-888-0>

Buzzoni, Marina e Roberto Rosselli Del Turco. 2015. "Verso un'edizione digitale dell'Editto di Rotari". In *I Longobardi in Italia: lingua e cultura. XV Seminario avanzato in Filologia germanica*, a cura di Carla Falluomini. Alessandria: Edizioni dell'Orso.

<https://iris.unive.it/handle/10278/3662705>

Fortunati, Maria e Caterina Giostra (a cura di). 2019. *I Longobardi del ducato di Bergamo. Le necropoli di Fara Olivana e Caravaggio - Masano*. Mantova: SAP Società Archeologica s.r.l.

Gasparri, Stefano. 2012. *Italia longobarda. Il regno, i Franchi, il papato*. Editori Laterza.

Groppali, Riccardo (a cura di). 2014. *Sulle orme dei Longobardi a Cremona. Atti del convegno. 10 maggio 2014, Cremona*. Inner Wheel Club di Cremona PHF, Distretto Italia 206.

<https://www.innerwheel.it/documenti/libro-longobardi-inner-wheel-1145247826.pdf>

Lombardia Beni Culturali (Regione Lombardia). L'età longobarda (sec. VI - VII).

<https://www.lombardiabeniculturali.it/istituzioni/storia/?unita=01.01>

Paroli, Lidia (a cura di). 1997. *L'Italia centro-settentrionale in età longobarda. Atti del Convegno (Ascoli Piceno 1995)*. Borgo S. Lorenzo: All'Insegna del Giglio.

<https://www.insegnadelgiglio.it/prodotto/litalia-centro-settentrionale-in-et-longobarda/>

Possenti, Elisa. 2014. "Necropoli longobarde in Italia: lo stato della ricerca". In *Necropoli longobarde in Italia: indirizzi della ricerca e nuovi dati*, a cura di Possenti Flica. 35-54. Trento: Castello del

Risorse collegate	I Longobardi in Italia. I luoghi del potere (568 - 774 d.C.) I Longobardi L...	I Longobardi in vetrina (I Longobardi in Italia) Il progetto "Lo...	Il lessico giuridico delle leggi longobarde Tesi di dottora...
	Paesaggi Longobardi Il progetto	Castelseprio - Lapide Longobarda Castelseprio-To...	Codice diplomatico longobardo 1-2: the original charters and coeval copies written in Tuscany between AD 714 and 774 TEI XML edition...
	Dizionario longobardo Dizionario long...	Manoscritto di grande formato contenente la storia di S. Gallo, realizzato intorno al 1200 - Cod. Sang. 547 Codice Sang. 54...	Museo Archeologico Nazionale - Cividale del Friuli Il Museo, che a...
	Paulus Diaconus, Historia Longobardorum - Cod. Sang. 635 Codice Sang. 63...	Tempietto Longobardo Il Monastero di...	Edictum Rothari (Veterum Fragmentorum Tomus III) - Cod. Sang. 730 Codice Sang. 73...
	SURVEY, ARCHAEOASTRONOMY AND COMMUNICATION: THE MAUSOLEUM OF THEODORIC IN RAVENNA (ITALY) The Mausoleum o...	Cividale: gli studi ed i restauri del tempietto longobardo Foramitti, Vitt...	Cod. AA 90.28 (A) Leges Langobardorum, lat. (frammento) Foglio sciolto ...
	Cod. Sang. 620 Codice Sang. 62...	Manoscritto vat. lat. 5359 Manoscritto pre...	Edizione digitale delle <i>Leges Langobardorum</i> XML...
	Edizione diplomatico-interpretativa dell'<i>Edictum Rothari</i> Edizione diplom...		

Figure 36 ... and after

The support of one single user helped increment the path with five more bibliographic references and one resource (visible in figure 34 in the boxes). With a wide use, the platform could therefore gain great benefits.

Once given his/her contribution, the philologist keeps on exploring the functionalities of the platform. In particular, he/she is attracted by the use of Pundit. He/she creates a notebook to organize his/her work. He/she wants to create a list collecting all the manuscripts together with their place of preservation, in order to have the information quickly at hand. Moreover, thanks to the metadata that CHIt retrieves for each resource, the user can collect all the links to the digitized manuscripts. By accessing the external websites, more information can be found and gathered and possibly added to the

dedicated Pundit notebook, for example the year of the manuscript or bibliographic references, as shown in figure 37.

gfabbri8181@gmail.com
February 28, 2025 · in Editto

vat. lat. 5359

Roma, https://digi.vatlib.it/view/MSS_Vat.lat.5359
Bibliografia: <https://digi.vatlib.it/mss/detail/Vat.lat.5359>

IXsec

gfabbri8181@gmail.com
February 28, 2025 · in Editto

Cod. Sang. 730

San Gallo, <https://www.e-codices.unifr.ch/it/searchresult/list/one/csg/0730>

VIIsec

gfabbri8181@gmail.com
February 28, 2025 · in Editto

Cod. AA 90.28 (A)

Berna, <https://katalog.burgerbib.ch/detail.aspx?ID=195192>

XIIsec

Figure 37 Example of annotation with information gathered also outside CHIt

Moreover, in case the Germanic philologist is also interested in LOD (cf. section 3.1), he/she could use the platform to enrich the resources available with semantic annotations (see section 3.2.4 for a definition of *semantic annotation*).

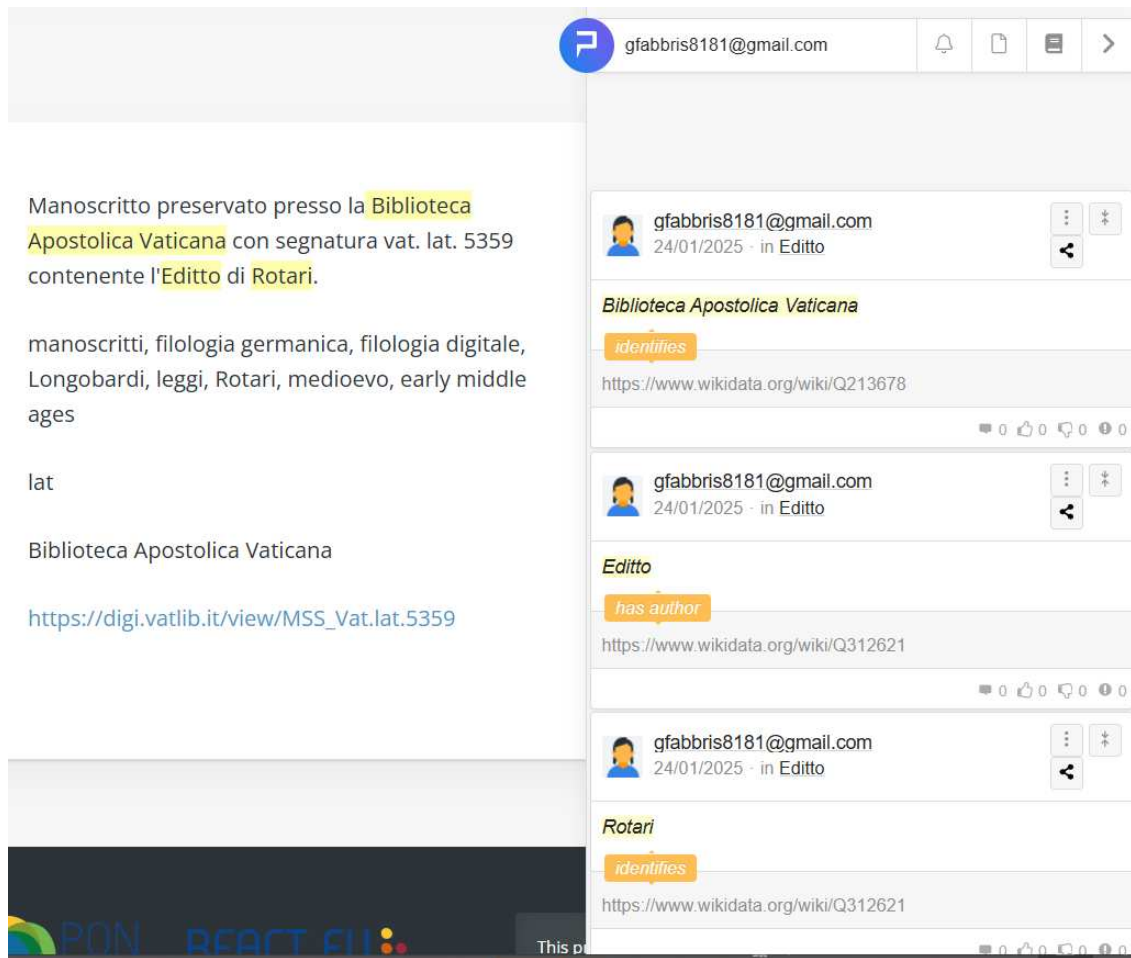


Figure 38 Examples of semantic annotations

As seen in figure 38, the user has linked the unique identifiers provided by Wikidata,²⁷¹ a free and open knowledge base that collects data in a format readable both by humans and machines. This concept is in line with what has been written above about FAIR data, since Wikidata is an open, free and collaborative database that gathers a huge amount of open data – meaning that they are accessible, modifiable and usable by everyone – which are linked to each other. These data can be read not only by humans, but also by machines, and for this reason information can be extracted and complex elaborations

²⁷¹ https://www.wikidata.org/wiki/Wikidata:Main_Page. Wikidata is just one of the many knowledge bases available to feed the Semantic Web.

can be performed.²⁷² However, such a high-level annotation can be expected only by a user who knows what semantic annotations deal with.

Last but not least, provided that the philologist has also teaching duties, he/she could decide to exploit CHIt for a work for his/her Germanic Philology class. For example, students could be asked to locate the relevant historical artefacts left by the Lombards that can be found on CHIt and mark the areas where these are found. After that, they should read the *Historia Langobardorum* by Paolo Diacono²⁷³ and annotate on a shared Pundit notebook whether the areas where these artefacts are located correspond to the areas mentioned in the *Historia*. Figure 39 shows an example of annotation shared between two students.²⁷⁴

²⁷² Casagrande (2023), but cf. also Lemus-Rojas and Pintscher (2018).

²⁷³ Zanella (1991)

²⁷⁴ This last case study is part of an abstract presented for the conference AIUCD 2025. The original proposal continues with a further “exercise” for the students in which they are asked to leave the CHIt app and access the OntoVE-Sampo-UI (<https://ontove-ui.chitontove.it/en/>) to consult the *Events* perspective in order to individuate the relevant acts issued by Charlemagne after conquering the Lombard reign, and then filter the locations in Veneto mentioned in the documents through the *Subject* filter provided in the *Texts* perspective, comparing them again to the places mentioned in the *Historia*. I thank my colleague Chiara De Bastiani for her collaboration and her idea for this case study.

del Duomo di Monza

Descrizione

Sito ufficiale del Museo e Tesoro del Duomo di Monza. Al suo interno, si possono consultare approfondimenti sulle seguenti categorie:

- Duomo di Monza
- Museo e Tesoro
- Corona Ferrea
- **Regina Teodolinda**
- Le altre chiese
- La Monaca di Monza
- Biblioteca e Archivio
- Longobard Ways

Parole chiave

Longobardi, monza

Collegamento alla fonte originale

<https://www.museoduomomonza.it/>

gfabbris8181@gmail.com

01/03/2025 · in Progetto univ

Regina Teodolinda

Cf. Historia pp.363-365

1 0 0 0

Chiara DE BASTIANI Importanza di Teodolinda per la conversione al cattolicesimo: aggiungere info slides lezione

0 0 0 0

Figure 39 Students' shared annotations

Figure 40 shows in a schematic way all the actions described above.

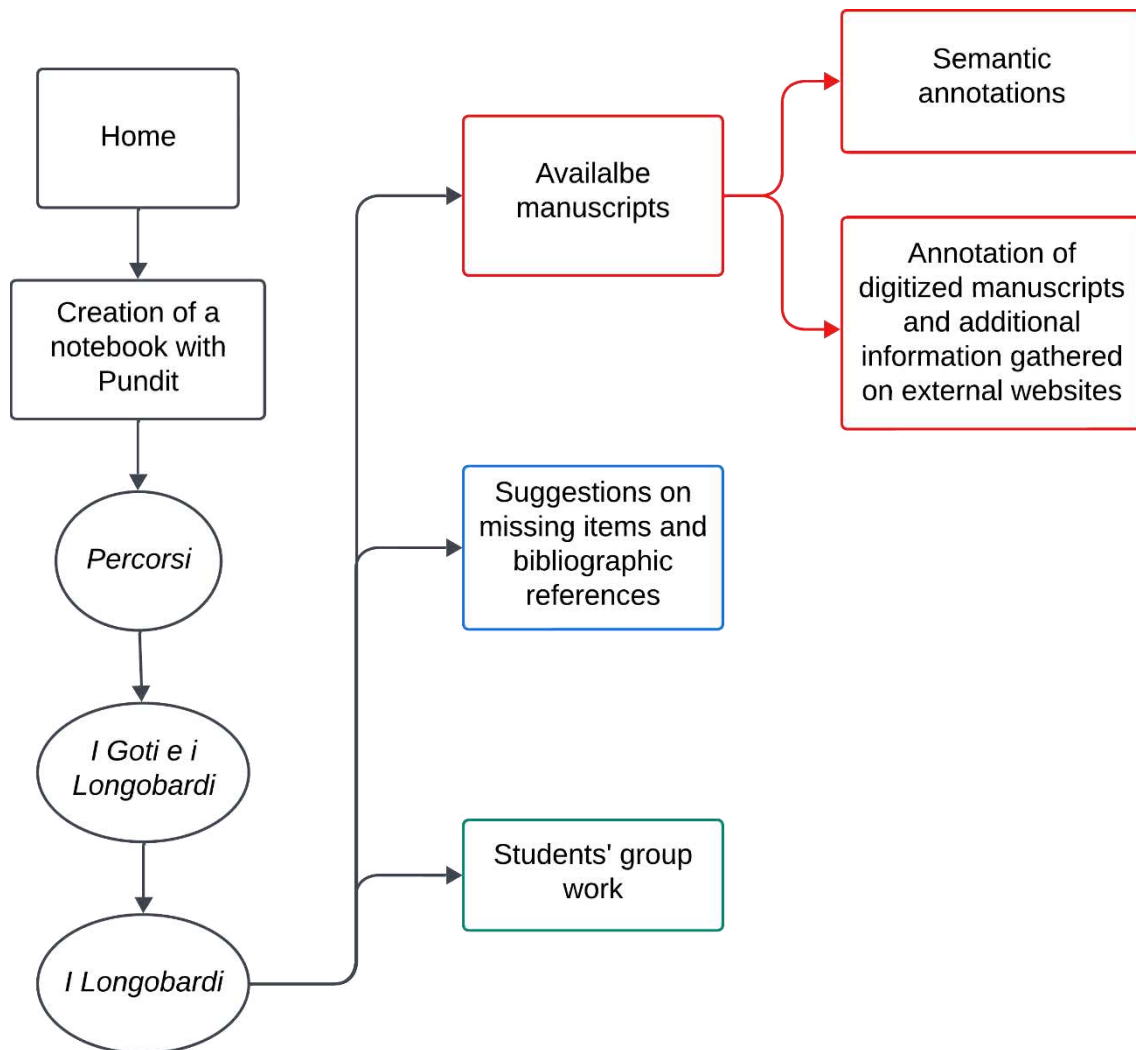


Figure 40 Philologist's steps

In conclusion, this final case study – the most articulate – demonstrates that CHIt can be a valuable tool for a Germanic philologist, adaptable to different levels of IT proficiency. It can be utilized both by users familiar with and interested in LOD, and by expert Germanic philologists who may not be accustomed to using complex digital tools. Lastly, the app can be used also in a didactic environment.

6.3. Some preliminary conclusions

The three case studies presented have been conceived to prove the usefulness of CHIt and its adaptability to different types of users. It seems fair to assert that all the users

have found in the app a good support for their research, even if to different extent. CHIt can be used as a basis for various kinds of investigations, and should the metadata gathered for the single resources not be enough, there is always the possibility to refer to the external websites for more precise information.

Thanks to these user stories, it seems evident that research that requires a considerable number of details and precision should be supported by the original sites or other sites to collect what CHIt misses. If, on the one hand, this could be considered a downside, on the other it is intrinsic of its nature. In fact, CHIt collects so many different types of resources that it would be impossible to go too in detail for each of them. In particular, those retrieved via API or SPARQL query cannot even be modified, and the metadata displayed are those offered by the original providers. As for the items added manually, more details could be actually inserted, but this brings along a risk. If, for example, opening hours or ticket fees of a museums were added to the metadata, what if they changed? The absence of automated retrieval implies that manual updates should be performed periodically. In the meantime, users could find incorrect information. For this reason, such kind of information is left to the original website.

These case studies have also proven the usefulness of itineraries, as they help users find their way into such a huge number of resources. For this reason, in the future more paths will be implemented.

Last but not least, all three case studies have shown how CHIt is suitable for the teaching field, from high school to university, from individual work to group projects. This is interesting since it could help younger users come closer to cultural heritage via an engaging tool, in line with the European documents cited in chapter 1.

7. Future developments

In the six chapters above, it has been shown in detail all the phases that have brought to the development of a web application for the valorization and use of the Italian cultural heritage that exploits existing tools and resources. The result, as chapter 6 has proven, is more than satisfactory and it can be asserted that CHIt is going to serve well a varied audience, and this was one of the *desiderata*.

However, what was conceived in the original idea and what has been produced diverge to some aspects. The main aspect regards the FAIR retrieval of resources. Unfortunately, some limits were imposed by the providers themselves, while others by the nature of the project, which did not allow for a huge number of collaborators and, therefore, the restricted team had to work with their forces.

Nonetheless, it is mandatory and inevitable to dream of the future of CHIt in order to envisage and structure its further potential, as at the present moment it could be defined as a Version 1 which could – and should – be improved under different aspects. Some of them have been postponed because of time issues, others for the lack of manpower, as already stated above.

- To enlarge the pool of potential users, the app will be translated in English.
- The testing phase (section 5.2) has highlighted a problem with the visualization of images. In fact, they are opened in different ways according to the various providers. This problem will have to be solved and all images will be likely to be opened in the image-viewer above the Metadata section, as it already happens for Europeana.
- The whole app will be revised to be as inclusive as possible, following accessibility guidelines.²⁷⁵
- More itineraries will be created, since the case studies have proved how helpful they are to guide users in their navigation. For example, a path on another Germanic

²⁷⁵ For example, those issued by the W3C (<https://www.w3.org/TR/WCAG21/>).

tribe could be added to enrich the dedicated section. The Vandals, for instance, reached Sicily.

- To improve the general quality of the app, specific targets of users could be asked to perform searches and provide suggestions concerning resources and bibliographic references to be added. For instance, the Italian Association of Germanic Philology (AIFG)²⁷⁶ has a dedicated mailing list through which this request could be sent to the members.
- As for the SPARQL Endpoint, a try with a different endpoint will be performed. The *Catalogo generale dei beni culturali*²⁷⁷ seems a good candidate.
- If on the one side EVT is totally integrated in the app and is used automatically when needed, Pundit is an external tool. For this reason, a pop-up message will be integrated to facilitate users in finding and installing it, while in the future it will be possibly integrated in turn.
- More formats will be added (the first tries will be performed with 3Ds and videos).
- A section dedicated to the documentation will be implemented.
- The app will be hosted in a different platform to ensure long-term availability.²⁷⁸
- The problem on the filtering of the collections will be hopefully solved.
- Thanks to the use of Zenodo and GitHub, it would be desirable to improve the number of independent scholarly projects that do not have an official website for the publication, as it is often the case for bachelor and master thesis.
- As mentioned earlier (section 2.3), many independent websites and projects that had been considered at the beginning, could not be inserted because their metadata are not open and/or exposed. It might be interesting to contact the supervisors of these projects and ask them if they could provide access in some way.

²⁷⁶ <https://aifg.it/>

²⁷⁷ General catalogue of cultural heritage, an Italian database created by the ICCU (Central Institute for Catalogue and Documentation). Homepage: <https://catalogo.beniculturali.it/>.

²⁷⁸ Please contact the email address giulia.fabbris@unive.it if a redirect will not be implemented.

- This web application is ever-growing. This means that resources can always be added.
- As mentioned in section 3.2.3.1, collections now work thanks to the addition of a single filter. However, in this way, that information is mixed together with all the other filters, and this method reduces the usefulness of collections. In fact, the user could simply click on the filter dedicated to the collection from the main search page and the result would be the same. To solve this problem, two paths could be followed. One solution is to add a new field in the backend which would be dedicated only to this function (e.g., *type*). In this way, the dedicated filters could be removed and the, say, *type* section would remain hidden from the UI. The other possibility is to change the logic of the app, which now follows the `and` logic, and change it into the `or` logic. In this way, more filters could be added simultaneously without excluding each other. The result in the research bar would be something like <http://chit.chitontove.it/risorse?subject=manoscritti,leggi,charters>. The first solution is quicker and less invasive than the second, which requires a massive change in the code. However, the second way would allow to add also OntoVE filters, while the first one not. In fact, OntoVE (and potentially also Europeana) do have their own filters, which could be added to the string. On the other hand, if the solution adopted would be the first one, the problem with these two providers would persist since they cannot be edited manually in the backend.
- It would be desirable – and challenging for sure – to improve the retrieval of metadata from Zenodo, but especially GitHub. In fact, at the present moment, GitHub resources are added completely manually, even though APIs are available. A deeper analysis of the documentation could be performed, and the same is true for Zenodo.

Conclusions

At the beginning of this work the importance of FAIR data was stressed, importance that is recognized at the European level. Numerous are the decrees, documents, meetings and regulations that have been issued in the new millennium by different bodies (UNESCO, EUCO, EC, MUR, etc.), and we have assisted to an increase as years go by.

However, these documents are directed to people that have to *act* in order to put into practice their content.

Findability, accessibility, interoperability and reusability are the key to ensure a digital development of societies at all levels, from jobs to healthcare, from education to culture. Therefore, it should be in the interest of the whole community to observe them.

In this view, this project has been carried out with a constant adherence to a series of guidelines and milestones firmly tied to the European recommendations that have hopefully ensured its success. After the initial discomfort regarding the impossibility to reuse the target resources, the new path undertaken proved adequate nonetheless. The different types of retrieval were not part of the initial idea, which encompassed only API. Therefore, it has been challenging to add the SPARQL Endpoint and the resources in the backend. For every provider, the most adequate manner of retrieval had to be found and this required a long phase of development.

One of the most interesting aspects of CHIt is the fact that it reuses all open-access existing tools, so the process of “reinventing the wheel”, which is a huge problem in Digital Humanities, was avoided. There are innumerable tools available on which scholars and programmers have spent skills, time and money, and that have already been tested. Exploiting these open-access instruments and adapting them to the needs of one’s project is a crucial step that should be considered all the time at the beginning of a new work.

Another fundamental step and strictly connected to the just cited question on reuse has been the collaboration of the PhD candidate with a company that works in the field of IT but specialized in Digital Humanities. In fact, the development of an application requires very technical skills that a humanist, even if skilled in technology, is unlikely to

have. At the same time, a programmer, even if skilled in Humanities, probably misses the knowledge necessary to complete a project of this kind (for example, it could be difficult for him/her to decide which resources to consider under a scientific point of view). Therefore, it can be concluded that this synergy strongly encouraged by the PON (cf. section 1.1) proved fundamental for the conclusion of this project.

However, it is tempting to say that this project is not concluded, because there is always a chance to improve such a lively application. If, on the one hand, the UI can be considered finished at some point, on the other hand, the content can always be updated, edited, and enlarged. This aspect was highlighted during the third case study presented in chapter 6.

The case studies proved fundamental to test the usability of CHIt and showed that the application actually fits different kinds of users. This was one of the main goals, also in the light of the Faro Convention (cf. *Introduction*).

Chapter 6 has shown that CHIt is suitable for amusement (e.g., tourists that need to find information on museums), education (school or university works) and scholarly research. This range of uses seems to be quite wide.

Thanks to this innovative application, a piece of the cultural heritage of Italy is delivered to the widest audience possible in a digital format on a digital platform. Now, it is their turn to find innovative ways to exploit it.

Bibliography

Addink, Wouter, Niki Kyriakopoulou, Lyubomir Penev, David Fichtmueller, Ben Norton, and David Shorthouse. 2023. "Deliverable D1.3 Best Practice Manual for Findability, Re-Use and Accessibility of Infrastructures." *ARPHA Preprints* 4 (May). <https://doi.org/10.3897/arphapreprints.e107169>.

Aiola, Chiara, Giulio Andreini, Francesca Di Donato, and Tiziana Lombardo. 2021. "Sharing Knowledge Digitally, the Muruca Case Study." In , 428–31. Pisa. <https://aiucd2021.labcd.unipi.it/book-of-abstracts/>.

Aiola, Chiara, and Tiziana Lombardo. 2022. "Storytelling through Digital Scholarly Editions." In *DARIAH Annual Event 2022: Storytelling*. Athens. <https://doi.org/10.5281/zenodo.6594346>.

Aiola, Chiara, Tiziana Lombardo, and Carlo Teo Pedretti. 2023. "Towards a Cultural User Experience approach in designing Scholarly Digital Edition. *Kierunek – kulturowe doświadczenie użytkownika w projektowaniu naukowych edycji cyfrowych*." *Sztuka Edycji* 23 (1). <https://apcz.umk.pl/sztukaedycji/issue/view/2634>.

Alber, Birgit. 2014. "Obstruent Systems of Northern Italy." *L'Italia Dialettale. Rivista Di Dialettologia Italiana* 75:13–36.

Angster, Marco. 2014. "Lingue Di Minoranza e Di Maggioranza. 200 Anni Di Lingue Straniere a Gressoney (AO)." In *Minoranze in Mutamento. Etnicità, Lingue e Processi Etnografici Nelle Valli Alpine Italiane*, edited by Valentina Porcellana and Federica Diémoz, 105–21. Alessandria: Edizioni dell'Orso.

Angster, Marco, Marcella Benedetti, and Ingeborg Geyer. 2014. *Il Tesoro Linguistico Delle Isole Germaniche in Italia*. Luserna: Comitato unitario delle isole linguistiche storiche germaniche in Italia. <https://www.isolelinguistiche.it/it/il-tesoro-linguistico.html>.

Arroyo Poleo, Fabiana. 2021. "What Are the Effects of the So-Called API Economy on the Financial Industry." *Dana Connect* (blog). November 12, 2021. <https://www.danaconnect.com/what-is-the-api-economy/>.

AWS. n.d. "What Is Unit Testing? - Unit Testing Explained." Amazon Web Services, Inc. Accessed May 22, 2024. <https://aws.amazon.com/what-is/unit-testing/>.

Azzara, Claudio. 2013. *Teoderico*. Bologna: Il Mulino.

Azzara, Claudio, and Stefano Gasparri. 2005. *Le Leggi Dei Longobardi. Storia, Memoria e Diritto Di Un Popolo Germanico*. Roma: Viella. <http://www.viella.it/libro/9788883340994>.

Baker, Mark. 2013. "Findability Is a Content Problem, Not a Search Problem." *Every Page Is Page One* (blog). May 28, 2013. <https://everypageispageone.com/2013/05/28/findability-is-a-content-problem-not-a-search-problem/>.

Barbuti, Nicola. 2021. "Thinking Digital Libraries for Preservation as Digital Cultural Heritage: By R to R4 Facet of FAIR Principles." *International Journal on Digital Libraries* 22 (3): 309–18.

Bekiari, Chrysoula, George Bruseker, Erin Canning, Martin Doerr, Philippe Michon, Christian-Emil Ore, Stephen Stead, and Athanasios Velios. 2022. "Definition of the CIDOC Conceptual Reference Model, Version 7.1.2." <https://cidoc-crm.org/Version/version-7.1.2>.

Bekiari, Chrysoula, Martin Doerr, Patrick Le Boeuf, and Pat Riva. 2023. "LRMOO Object-Oriented Definition and Mapping from IFLA LRM, Version 0.9.6."

Beyerle, Franz. 1947. *Leges Langobardorum: 643-866*. Weimar.

Bidese, Ermenegildo. 2024. "Il cimbro." *Linguistik Online* 130 (6): 83–107. <https://doi.org/10.13092/lo.129.11151>.

Bioglio, Livio, Giacomo Cerretini, Giulia D'Agostino, Elisabetta Magnanti, and Roberto Rosselli Del Turco. 2023. "Three Is the Charm: A New Architecture, New Features and New Projects in EVT 3." In *Digital Humanities 2023: Book of Abstract*. Graz: University of Graz. <https://doi.org/10.5281/zenodo.8107481>.

Biswas, Goutam, and Dibyendu Paul. 2019. "An Evaluative Study on the Open Source Digital Library Softwares for Institutional Repository: Special Reference to Dspace and Greenstone Digital Library." *African Journal of Library and Information Science* 5 (6): 1–10.

Bluhme, Friedrich. 1868. *Leges Langobardorum*. Hannover: Hahn.

Boffo, Vanna. 2022. "Transizioni per Il Lavoro in Higher Education: Il Ruolo Del Dottorato Di Ricerca in Italia." In *Esercizi Di Ricerca: Dottorato e Politiche per La Formazione*, edited by Vanna Boffo and Fabio Togni, 17–35. Studies on Adult Learning and Education 15. Firenze: Firenze University Press. <https://dx.doi.org/10.36253/979-12-215-0081-3>.

Braake, Serge Ter, Antske Fokkens, Niels Ockeloën, and Chantal Van Son. 2016. "Digital History: Towards New Methodologies." In , AICT-482:23. https://doi.org/10.1007/978-3-319-46224-0_3.

Brunetti, Dimitri. 2018. "Mèmore. La nuova piattaforma digitale per i beni culturali piemontesi." *DigItalia* 13 (1): 109–24.

Buzzoni, Marina, and Roberto Rosselli Del Turco. 2015. "Verso un'edizione digitale dell'Editto di Rotari." In *I Longobardi in Italia: lingua e cultura. XV Seminario avanzato in Filologia germanica*, edited by Carla Falluomini, Edizioni dell'Orso. Alessandria: ITA. <https://iris.unive.it/handle/10278/3662705>.

———. 2024. "Edizione diplomatico-interpretativa con facsimile digitale dell'Edictum Rothari, Vercelli, Biblioteca e Archivio Capitolare, CLXXXVIII," *Filologie medievali e moderne*, 30. <https://doi.org/10.30687/978-88-6969-888-0>.

Caffo, Rossella. 2008. "CulturaItalia: il Portale della cultura italiana." *DigItalia* 3 (1): 71–75.

Cahill, Kay. 2009. *User-Generated Content and Its Impact On Web-Based Library Services: Questioning Authority*. Elsevier.

Callaway, Elizabeth, Jeffrey Turner, Heather Stone, and Adam Halstrom. 2020. "The Push and Pull of Digital Humanities: Topic Modeling the 'What Is Digital Humanities?' Genre." *Digital Humanist Quarterly* 14 (1). <http://digitalhumanities.org/dhq/vol/14/1/000450/000450.pdf>.

Candela, Gustavo, María Dolores Sáez, MPilar Escobar Esteban, and Manuel Marco-Such. 2022. "Reusing Digital Collections from GLAM Institutions." *Journal of Information Science* 48 (2): 251–67. <https://doi.org/10.1177/0165551520950246>.

Caria, Marco. 2024. "Il sappadino." *Linguistik Online* 130 (6): 109–30. <https://doi.org/10.13092/lo.129.11216>.

Carriero, Valentina Anita, Aldo Gangemi, Maria Letizia Mancinelli, Ludovica Marinucci, Andrea Giovanni Nuzzolese, Valentina Presutti, and Chiara Veninata. 2019. "ArCo: The Italian Cultural Heritage Knowledge Graph." In *The Semantic Web – ISWC 2019: 18th International Semantic Web Conference, Auckland, New Zealand, October 26–30, 2019, Proceedings, Part II*, 36–52. Berlin, Heidelberg: Springer-Verlag. https://doi.org/10.1007/978-3-030-30796-7_3.

Casagrande, Paolo. 2023. "Che cos'è Wikidata." *Wikimedia Italia* (blog). 2023. <https://www.wikimedia.it/news/che-cose-wikidata/>.

Chiarelli, Andrea, Neil Beagrie, Lotte Boon, Ruth Mallalieu, Rob Johnson, Amy Warner May, and Rowan Wilson. 2022. "To Protect and to Serve: Developing a Road Map for Research Data Management Services." *Insights: The UKSG Journal* 35 (0): 4. <https://doi.org/10.1629/uksg.566>.

Chikobava, Gogi. 2023. "The Crucial Role of APIs in Web Development." *TCLoc Master's* (blog). October 5, 2023. <https://mastertcloc.unistra.fr/2023/10/05/the-crucial-role-of-apis-in-web-development/>.

Christie, Neil. 1999. *The Lombards – The Ancient Longobards*. Wiley-Blackwell. <https://www.wiley.com/en-us/The+Lombards%3A+The+Ancient+Longobards-p-9780631211976>.

Cimiano, Philipp, Christian Chiarcos, John P. McCrae, and Jorge Gracia. 2020. "Linguistic Linked Data in Digital Humanities." In *Linguistic Linked Data: Representation, Generation and Applications*, edited by Philipp Cimiano, Christian Chiarcos, John P. McCrae, and Jorge Gracia, 229–62. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-30225-2_13.

Ciotti, Fabio. 2018. "Le Digital humanities in Italia: la tradizione del nuovo." *AIB studi* 58 (2). <https://doi.org/10.2426/aibstudi-11828>.

Codecademy. n.d. "What Is REST?" Codecademy. Accessed November 28, 2023. <https://www.codecademy.com/article/what-is-rest>.

Cognola, Federica. 2024. "Il Mòcheno." *Linguistik Online* 130 (6): 59–82. <https://doi.org/10.13092/lo.129.11150>.

Comitato Unitario. 2022. *Le Isole Linguistiche Germaniche Fioriscono*. Luserna: Comitato unitario delle isole linguistiche storiche germaniche in Italia. <https://www.isolelinguistiche.it/it/le-isole-linguistiche-germaniche-fioriscono.html>.

De Bastiani, Chiara. 2023. "Harmonizing and Modelling a Bibliographic Ontology of Imperial Diplomas within the OntoVE Knowledge Base." In *La Memoria Digitale: Forme Del Testo e Organizzazione Della Conoscenza. Atti Del XII Convegno Annuale AIUCD*, 51–56. Siena: Università degli Studi di Siena.

———. 2024. "Connecting and Visualizing Data: A Knowledge Base of Germanic Cultural Heritage Items in Venice." Presented at the DH Benelux 2024, Leuven, June 5. <https://doi.org/10.5281/zenodo.11485003>.

De Bastiani, Chiara, and Giulia Fabbris. 2023a. "Collaboration within a Shared Digital Paradigm: Opportunities and Outcomes." Presented at the ADHO2023, Graz, July 31. <https://doi.org/10.5281/zenodo.8198555>.

———. 2023b. "Collaboration within a Shared Digital Paradigm: Opportunities and Outcomes." In *Digital Humanities 2023: Book of Abstracts*, edited by Anne Baillot, Toma

Tasovac, Walter Scholger, and Georg Vogeler, 489–91. Graz: University of Graz.
<https://doi.org/10.5281/zenodo.8210808>.

De Bastiani, Chiara, Giulia Fabbris, and Alessandro Bertozzi. n.d. “Retrieving, Sharing, Reusing: A Collaborative Case Study.”

De Santis, Luca, Duccio Breschi, Niccolò Cardelli, Edgar Gomez, Massimiliano Pardini, Marco Zizi, Maxime Bouillard, Simone Kopeinik, and Laurent Capelli. 2021. “TRIPLE Deliverable: D5.5 Report on the Open Annotation Tool.” Zenodo.
<https://doi.org/10.5281/zenodo.5653219>.

Denz, Josef. 2020. “Bayerisches Wörterbuch.” 2020.
<https://publikationen.badw.de/de/bwb>.

Dewi, Risqi Puspa, Mira Kania Sabariah, and Veronikha Effendy. 2019. “Design Recommendation Information Architecture of Hospital Website Using Bottom-Up Approach on Card Sorting Method.” *International Journal on Information and Communication Technology (IJOICT)* 5 (2): 20–31.
<https://doi.org/10.21108/IJOICT.2019.52.190>.

Di Giorgio, Sara. 2014. “Culturaitalia, the Italian National Content Aggregator in Europeana.” *Procedia Computer Science* 38 (November).
<https://doi.org/10.1016/j.procs.2014.10.008>.

Di Noia, Tommaso, Azzurra Ragone, Andrea Maurino, Marina Mongiello, Maria P. Marzocca, Giuseppe Cultrera, and Mauro P. Bruno. 2016. “Linking Data in Digital Libraries: The Case of Puglia Digital Library.” *Proceedings of the 1st Workshop on Humanities in the Semantic Web Co-Located with 13th ESWC Conference 2016* 1608:27–38.

Directorate-General for Research and Innovation. 2016. *Realising the European Open Science Cloud: First Report and Recommendations of the Commission High Level Expert Group on the European Open Science Cloud*. LU: Publications Office of the European Union. <https://data.europa.eu/doi/10.2777/940154>.

Doerr, Martin, and Steve Stead. 2012. "Harmonized Models for the Digital Worlds: CIDOC CRM, FRBROO, CRMDig Snd Europeana EDM." In *CIDOC 2012 - Enriching Cultural Heritage. Abstracts and Programme*. Helsinki.

Dolcetti Corazza, Vittoria, and Carla Falluomini. 2020. *I Goti. Percorsi Storici, Letterari e Linguistici*. Spoleto: Fondazione Centro italiano di studi sull'alto medioevo. https://shop.cisam.org/index.php?route=product/product&product_id=1128.

Dziuba, Anna. 2023. "Why and When to Use Node.js in 2021." Relevant Software. June 8, 2023. <https://relevant.software/blog/why-and-when-to-use-node-js/>.

EC. 2020a. *A European Strategy for Data*. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0066>.

———. 2020b. *Data Governance Act - Regulation of the European Parliament and of the Council on European Data Governance*. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020PC0767>.

———. 2020c. "Digital Economy and Society Index (DESI). Italy."

———. 2020d. *Shaping Europe's Digital Future*. <https://eur-lex.europa.eu/legal-content/en/TXT/PDF/?uri=CELEX:52020DC0067&from=EN>.

———. 2021. *COMMISSION RECOMMENDATION (EU) 2021/1970 of 10 November 2021 on a Common European Data Space for Cultural Heritage*. <https://eur-lex.europa.eu/eli/reco/2021/1970/oj>.

———. 2022a. *Data Act - Regulation of the European Parliament and of the Council on Harmonised Rules on Fair Access to an Use of Data*. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020PC0767>.

———. 2022b. “Digital Economy and Society Index (DESI). Italy.”

———. 2022c. “Linking Data: What Does It Mean?” Data.Europa.Eu - The Official Portal for European Data. August 1, 2022. <https://data.europa.eu/en/publications/datastories/linking-data-what-does-it-mean>.

———. 2023a. “Report on the State of the Digital Decade 2030.”

———. 2023b. “Report on the State of the Digital Decade 2030 - Annex Italy.”

ERC. 2022. “Open Research Data and Data Management Plans. Information for ERC Grantees.” European Research Council. https://erc.europa.eu/sites/default/files/document/file/ERC_info_document-Open_Research_Data_and_Data_Management_Plans.pdf.

EUCO. 2014. *Cultural Heritage as a Strategic Resource for a Sustainable Europe*. [https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52014XG0614\(08\)](https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52014XG0614(08)).

———. 2021a. *Conclusions of the European Council Meeting (21-22 October 2021)*. <https://www.consilium.europa.eu/media/52622/20211022-euco-conclusions-en.pdf>.

———. 2021b. *Statement of the Members of the European Council Meeting (25 March 2021)*.

EUCO, and EP. 2020. *REACT-EU — Recovery Assistance for Cohesion and the Territories of Europe*. OJ L. Vol. 437. <http://data.europa.eu/eli/reg/2020/2221/oj/eng>.

———. 2021. *REGULATION (EU) 2021/694 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 29 April 2021 Establishing the Digital Europe Programme and Repealing Decision (EU) 2015/2240*. Vol. 694. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32021R0694&from=EN>.

Fabbris, Giulia. 2023. “Use, Reuse and Valorisation: A Web App for the Italian Cultural Heritage.” In . Torino. <https://zenodo.org/records/15022004>.

———. 2024. “Germanic Minorities in Italy: A Survey of What Is Available.” In *Small Languages, Big Ideas: The Smaller Germanic Languages from a Theoretical, General and Comparative Perspective*, 21–24. Uppsala. <https://www.uu.se/download/18.60263fe818f775e7806130e2/1716370041037/SLBI%20Booklet.pdf>.

Fabbris, Giulia, and Alessandro Bertozzi. 2024. “Use, Reuse and Valorisation: A Web App for the Italian Cultural Heritage.” *Mimesis Journal* 13 (2): 135–49. <https://doi.org/10.13135/2389-6086/9998>.

Francovich Onesti, Nicoletta. 1999. *Vestigia Longobarde in Italia (568-774). Lessico e antroponimia*. Roma: Artemide Edizioni. <https://www.artemide-edizioni.it/prodotto/vestigia-longobarde-in-italia-568-774/>.

———. 2007. *I Nomi Degli Ostrogoti*. Firenze: Firenze University Press.

———. 2011. *Filologia Germanica. Lingue e Culture Dei Germani Antichi*. Roma: Carocci.

———. n.d. “La lingua dei Longobardi, caratteristiche e problemi.” Accessed February 14, 2025. https://www.academia.edu/7008968/La_lingua_dei_Longobardi.

Frye, Ma-Keba. n.d. “What Is an API? (Application Programming Interface).” MuleSoft. Accessed November 24, 2023. <https://www.mulesoft.com/resources/api/what-is-an-api>.

Fulk, Robert Dennis. 2018. *A Comparative Grammar of the Early Germanic Languages. Sigl.3*. Vol. 3. Studies in Germanic Linguistics (SiGL). John Benjamins Publishing Company. <https://benjamins.com/catalog/sigl.3>.

G7 Italy. 2017. “The Protection of Cultural Heritage: Rules, Practices, and Education.” Firenze. http://www.g7italy.it/sites/default/files/documents/Protection_of_cultural_heritage_italian_contribution/index.pdf.

Gasparri, Stefano. 2014. “I nodi fondamentali della storia longobarda.” In *L’eredità longobarda*. Viella. <https://iris.unive.it/handle/10278/3693542>.

———. 2021. *Italia longobarda. Il regno, i Franchi, il papato*. Bari: Laterza.

GeeksforGeeks. 2022. “Role of Middleware in Distributed System.” *GeeksforGeeks* (blog). July 10, 2022. <https://www.geeksforgeeks.org/role-of-middleware-in-distributed-system/>.

———. 2023a. “Functional Testing - Software Testing.” *GeeksforGeeks*. December 6, 2023. <https://www.geeksforgeeks.org/software-testing-functional-testing/>.

———. 2023b. “Manual Testing vs Automated Testing.” GeeksforGeeks. December 7, 2023. <https://www.geeksforgeeks.org/software-engineering-differences-between-manual-and-automation-testing/>.

Gkoumas, Georgios, and Fotis Lazarinis. 2015. “Evaluation and Usage Scenarios of Open Source Digital Library and Collection Management Tools.” *Program* 49 (3): 226–41. <https://doi.org/10.1108/PROG-09-2014-0070>.

Grassi, Marco, Christian Morbidoni, Michele Nucci, Simone Fonda, and Giovanni Ledda. 2012. “Pundit: Semantically Structured Annotations for Web Contents and Digital Libraries.” In *Proceedings of the 2nd International Workshop on Semantic Digital Archives*, 49–60. Paphos, Cyprus. <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=5111580dca14e316837ddfe80495001459af3415#page=49>.

Guiliano, Jennifer, and Laura Estill. 2023. “What Gets Categorized Counts: Controlled Vocabularies, Digital Affordances, and the International Digital Humanities Conference.” *Digital Scholarship in the Humanities* 38 (3): 1088–1100. <https://doi.org/10.1093/lhc/fqac091>.

Haubrichs, Wolfgang. 2010. “Sprache und Schriftlichkeit im langobardischen Italien — Das Zeugnis von Namen, Wörtern und Entlehnungen. Ein Kommentar zu Nicholas Everett, *Literacy in Lombard Italy*.” *I Germani in Italia* 2:133–201. <https://doi.org/10.14672/fg.v2i.2419>.

Hodgson, Justin, Jeremiah Kalir, and Christopher D. Andrews. 2023. “Social Annotation: Promising Technologies and Practices in Writing.” In *Digital Writing Technologies in Higher Education : Theory, Research, and Practice*, edited by Otto Kruse, Christian Rapp, Chris M. Anson, Kalliopi Benetos, Elena Cotos, Ann Devitt, and Antonette Shibani, 141–

55. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-36033-6_9.

Hogan, Aidan. 2020. "The Semantic Web: Two Decades On." Edited by Pascal Hitzler and Krzysztof Janowicz. *Semantic Web* 11 (1): 169–85. <https://doi.org/10.3233/SW-190387>.

Iannotta, Iolanda Sara, and Roberta Scarano. 2022. "Formazione in azienda in chiave Industria 4.0: CRM e Cloud Computing." *Lifelong Lifewide Learning* 18 (40): 52–61. <https://doi.org/10.19241/lll.v18i40.661>.

IBM. n.d. "What Is Middleware?" Accessed November 15, 2023. <https://www.ibm.com/topics/middleware>.

ICOM. 2020. "Digital Library." In *Glossario Museologia Digitale e Nuove Tecnologie*. <https://www.icom-italia.org/wp-content/uploads/2020/10/ICOMItalia.CommissioneTecnologie.2020.GlossarioTecnologie.pdf>.

Ikkala, Esko, Eero Hyvönen, Heikki Rantala, and Mikko Koho. 2022. "Sampo-UI: A Full Stack JavaScript Framework for Developing Semantic Portal User Interfaces." *Semantic Web Journal* 13 (1). <https://www.semantic-web-journal.net/system/files/swj2631.pdf>.

Italian Government. 1999. *Norme in materia di tutela delle minoranze linguistiche storiche*. <https://www.parlamento.it/parlam/leggi/99482l.htm>.

Jacob, Elin K., and Aaron Loehrlein. 2009. "Information Architecture." *Annual Review of Information Science and Technology* 43 (1): 1–64. <https://doi.org/10.1002/aris.2009.1440430110>.

Jacobson, Daniel, Greg Brail, and Dan Woods. 2012. *APIs: A Strategy Guide*. O'Reilly Media, Inc.

Jin, Brenda, Saurabh Sahni, and Amir Shevat. 2018. *Designing Web APIs: Building APIs That Developers Love*. O'Reilly Media, Inc.

Kalir, Jeremiah, and Antero Garcia. 2021. "Joining the 'Great Conversation' – The Fundamental Role of Annotation in Academic Society." *Impact of Social Sciences* (blog). 2021. <https://blogs.lse.ac.uk/impactofsocialsciences/2021/05/28/joining-the-great-conversation-the-fundamental-role-of-annotation-in-academic-society/>.

Khan, Shazia. 2019. "Dspace or Fedora: Which Is a Better Solution?" *Journal of Information Management* 56 (1): 45–50. <https://doi.org/10.17821/srels/2019/v56i1/139265>.

Kruijt, Anne, and Stefan Rabanus. 2025. "From VinKo to AlpiLink: Web-Based Long-Term Storage and Accessibility of Information | Korpus Im Text." *Korpus Im Text*. <https://www.kit.gwi.uni-muenchen.de/?p=106395>.

Kush, R. D., D. Warzel, M. A. Kush, A. Sherman, E. A. Navarro, R. Fitzmartin, F. Pétavy, et al. 2020. "FAIR Data Sharing: The Roles of Common Data Elements and Harmonization." *Journal of Biomedical Informatics* 107 (July):103421. <https://doi.org/10.1016/j.jbi.2020.103421>.

Lampert, Cory, and Su Kim Chung. 2011. "Strategic Planning for Sustaining User-Generated Content in Digital Collections." *Journal of Library Innovation* 2 (2): 74–93.

Lemus-Rojas, Mairelys, and Lydia Pintscher. 2018. "Wikidata and Libraries: Facilitating Open Knowledge." In *Leveraging Wikipedia: Connecting Communities of Knowledge*,

edited by Merrilee Proffitt, 143–58. Chicago: ALA Editions.
<https://hdl.handle.net/1805/16690>.

Leonardi, Simona, and Elda Morlicchio. 2009. *La Filologia Germanica e Le Lingue Moderne*. Bologna: Il Mulino.

Lister, Allyson, and Susanna-Assunta Sansone. 2023. “FAIRsharing in a Nutshell,” July.
<https://doi.org/10.5281/zenodo.8191958>.

Lo Duca, Angelica. 2021. “Is the Semantic Web Really Dead?” *Geek Culture* (blog). 2021.
<https://medium.com/geekculture/is-the-semantic-web-really-dead-7113cfd1f573>.

Marinelli, Luca. 2020. *Ecosistemi per La Trasformazione Digitale Delle PMI*. Torino: Giappichelli editore.

Marras, Anna Maria. 2022. “Le Associazioni MAB e Il Piano Nazionale Di Digitalizzazione.” *Biblioteche Oggi* 40 (8). <http://dx.doi.org/10.3302/0392-8586-202208-004-1>.

Matina, Shrestha. 2023. “What Is the 3-Tier Architecture?” *Medium* (blog). May 1, 2023.
<https://medium.com/@shrestha.matina.20/what-is-the-3-tier-architecture-4520522e0720>.

McMillan, Taryn. 2023. “Manual Testing vs Automated Testing: Key Differences.” TestRail. September 9, 2023. <https://www.testrail.com/blog/manual-vs-automated-testing/>.

MDN. 2023. "Introducing Asynchronous JavaScript." Mozilla Developer Network. July 7, 2023. <https://developer.mozilla.org/en-US/docs/Learn/JavaScript/Asynchronous/Introducing>.

———. 2024. "Express/Node Introduction." Mozilla Developer Network. 2024. https://developer.mozilla.org/en-US/docs/Learn/Server-side/Express_Nodejs/Introduction.

Medium. 2023. "Unleashing the Power of API Integration: How Businesses Collaborate Seamlessly with Other Companies." *Medium* (blog). 2023. <https://medium.com/@codekvnnn/unleashing-the-power-of-api-integration-how-businesses-collaborate-seamlessly-with-other-companies-b259630289cc>.

Menghin, Wilfried. 1985. *Die Langobarden. Archäologie Und Geschichte*. Stuttgart: Konrad Theiss.

MiC. 2023. "Piano Nazionale Di Digitalizzazione Del Patrimonio Culturale." <https://digitallibrary.cultura.gov.it/il-piano/>.

Mierow, Charles Christopher. 1908. "Jordanes. The Origin and Deeds of the Goths." Princeton: Princeton University Press.

Moilanen, Jarkko, Marjukka Niinioja, Marko Seppänen, and Mika Honkanen. 2019. *API Economy 101: Changes Your Business*. BoD - Books on Demand.

Mons, Barend, Erik Schultes, Fenghong Liu, and Annika Jacobsen. 2020. "The FAIR Principles: First Generation Implementation Choices and Challenges." *Data Intelligence* 2 (1–2): 1–9. https://doi.org/10.1162/dint_e_00023.

Morville, Peter. 2005. *Ambient Findability: What We Find Changes Who We Become*. O'Reilly Media, Inc.

Morville, Peter, and Louis Rosenfeld. 2006. *Information Architecture for the World Wide Web: Designing Large-Scale Web Sites*. O'Reilly Media, Inc.

MUR. 2021. *Regolamento Recante Modalità Di Accredimento Delle Sedi e Dei Corsi Di Dottorato e Criteri per La Istituzione Dei Corsi Di Dottorato Da Parte Degli Enti Accreditati*. <https://www.mur.gov.it/it/atti-e-normativa/decreto-ministeriale-n226-del-14-12-2021>.

Nielsen, Jakob. 2011. "Incompetent Research Skills Curb Users' Problem Solving." Nielsen Norman Group. April 10, 2011. <https://www.nngroup.com/articles/incompetent-search-skills/>.

Novak, Elena, Rim Razzouk, and Tristan E. Johnson. 2012. "The Educational Use of Social Annotation Tools in Higher Education: A Literature Review." *The Internet and Higher Education, Social Media in Higher Education*, 15 (1): 39–49. <https://doi.org/10.1016/j.iheduc.2011.09.002>.

OED. n.d. "Findability." In . Accessed November 20, 2023. https://www.oed.com/dictionary/findability_n?tab=factsheet.

Ontotext. n.d.-a. "What Are Linked Data and Linked Open Data?" Ontotext. Accessed March 12, 2025. <https://www.ontotext.com/knowledgehub/fundamentals/linked-data-linked-open-data/>.

———. n.d.-b. "What Is Semantic Annotation?" Ontotext. Accessed February 25, 2025. <https://www.ontotext.com/knowledgehub/fundamentals/semantic-annotation/>.

Orel, Vladimir. 2003. *A Handbook of Germanic Etymology*. Leiden: Brill.

Pech, Fernando, Alicia Martinez, Hugo Estrada, and Yasmin Hernandez. 2017. "Semantic Annotation of Unstructured Documents Using Concepts Similarity." *Scientific Programming* 2017. <https://doi.org/10.1155/2017/7831897>.

Pfeifer, Wolfgang. 1993. "Baum." In *Etymologisches Wörterbuch Des Deutschen*. <https://www.dwds.de/wb/etymwb/Baum>.

Pietrobon, Alessia. 2020. "Il Dialecto Bavarese: Tra Storia e Vitalità." Padova: Università degli Studi di Padova.

Polycarpou, Christiana. 2018. "The ViMM Definition of a Virtual Museum." Virtual Multimodal Museum. January 10, 2018. [https://www.vi-mm.eu/2018/01/10/the-vimm-definition-of-a-virtual-museum/#:~:text=A%20virtual%20museum%20\(VM\)%20is,experience%20and%20richness%20of%20content](https://www.vi-mm.eu/2018/01/10/the-vimm-definition-of-a-virtual-museum/#:~:text=A%20virtual%20museum%20(VM)%20is,experience%20and%20richness%20of%20content).

Pyrounakis, George, and Mara Nikolaidou. 2009. "Comparing Open Source Digital Library Software." In *Handbook of Research on Digital Libraries: Design, Development, and Impact, Information Science Reference*, edited by Yin-Leng Theng, Schubert Foo, Dion Goh, and Jin-Cheon Na. Hershey and New York: Information Science Reference. <https://www.igi-global.com/gateway/chapter/www.igi-global.com/gateway/chapter/19869>.

Radice, Sara. 2014. "DESIGNING FOR PARTICIPATION WITHIN CULTURAL HERITAGE Participatory Practices and Audience Engagement in Heritage Experience Processes." Tesi di dottorato, Milano: Politecnico di Milano. https://www.politesi.polimi.it/bitstream/10589/89798/1/2014_03_PhD_Radice.pdf.

Rametta, Gaetano. 2003. *Discorsi Alla Nazione Tedesca. Johann Gottlieb Fichte*. Biblioteca Universale Laterza. Roma-Bari: Editori Laterza.
<https://www.laterza.it/scheda-libro/?isbn=9788858118283>.

Robinson, Orrin W. 1992. *Old English and Its Closest Relatives. A Survey of the Earliest Germanic Languages*. Stanford: Stanford University Press.
<https://www.sup.org/books/literary-studies-and-literature/old-english-and-its-closest-relatives>.

Romano, Antonio. 2010. "Inventario Fonetico e Fonologico Delle Varietà Walser Italiane." In *Scrivere Tra i Walser. Per Un'ortografia Delle Parlate Alemanni-Che in Italia*, edited by Federica Antonietti, 55–72. Borgomanero: Litopress.

Rosselli Del Turco, Roberto. 2019. "Designing an Advanced Software Tool for Digital Scholarly Editions: The Inception and Development of EVT (Edition Visualization Technology)." *Textual Cultures* 12 (2): 91–111.
<https://doi.org/10.14434/textual.v12i2.27690>.

Rotili, Marcello. 2011. "Aspetti Dell'integrazione Delle Popolazioni Germaniche in Italia." In *Archeologia e Storia Delle Migrazioni. Europa, Italia, Mediterraneo Fra Tarda Età Romana e Alto Medioevo. Atti Del Convegno Internazionale Di Studi. Cimitile-Santa Maria Capua Vetere, 17-18 Giugno 2010. Giornate Sulla Tarda Antichità e Il Medioevo* (3), 97–115. Cimitile: Tavolario.

Saibene, Maria Grazia, and Marina Buzzoni. 2006. *Manuale Di Linguistica Germanica*. Milano: Cisalpino.

Salmons, Joseph. 2018. *A History of German. What the Past Reveals about Today's Language*. Oxford: Oxford University Press.

Saracco, Caterina. 2024. "Le varietà walser dell'Italia settentrionale: Gressoney e il titsch." *Linguistik Online* 130 (6): 235–55. <https://doi.org/10.13092/lo.129.11157>.

SBB. 2022. "API Principles - v. 2.3.0." Schweizerische Bundesbahnen. 2022. <https://schweizerischebundesbahnen.github.io/api-principles/>.

Scardigli, Piergiuseppe. 1964. *Lingua e Storia Dei Goti*. Firenze: Sansoni Editore.

Sidraschi, Diego. 2024. "Il saurano." *Linguistik Online* 130 (6): 131–47. <https://doi.org/10.13092/lo.129.11152>.

Sidraschi, Diego, and Francesco Costantini. 2022. "Un Manoscritto Ottocentesco Della Dottrina Cristiana Nella Parlata Tedesca Di Sauris/Zahre in Carnia: Origini Del Testo e Analisi Linguistica." *Linguistica e Filologia* 42. <https://air.uniud.it/retrieve/9737d800-8ca9-4d45-9e53-7ff7a1fb5551/Sidraschi-Costantini%2b-%2bLinguistica%2b%26filologia.pdf>.

Simou, Nikolaos, Alexandros Chortaras, Giorgos Stamou, and Stefanos Kollias. 2017. "Enriching and Publishing Cultural Heritage as Linked Open Data." In *Mixed Reality and Gamification for Cultural Heritage*, edited by Marinos Ioannides, Nadia Magnenat-Thalmann, and George Papagiannakis, 201–23. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-49607-8_7.

Sotirova, Kalina, Juliana Peneva, Stanislav Ivanov, Rositza Doneva, and Milena Dobрева. 2012. "Digitization of Cultural Heritage – Standards, Institutions, Initiatives." In *Access to Digital Cultural Heritage: Innovative Applications of Automated Metadata Generation*, edited by Krassimira Ivanova, Milena Dobрева, Peter Stanchev, and George Totkov, 23–68. Plovdiv: Plovdiv University Publishing House "Paisii Hilendarski."

Strange, Damon, Megan Gooch, and Alwyn Collinson. 2023. "Equality, Findability, Sustainability: The Challenges and Rewards of Open Digital Humanities Data." *International Journal of Performance Arts and Digital Media* 0 (0): 1–21. <https://doi.org/10.1080/14794713.2023.2206286>.

Techmango. 2022. "What Are the Benefits of Cross-Platform Application Development?" *Techmango* (blog). May 24, 2022. <https://www.techmango.net/what-are-the-benefits-of-cross-platform-application-development>.

Terras, Melissa. 2012. "Digitization and Digital Resources in the Humanities." In *Digital Humanities in Practice*, edited by Claire Warwick, Melissa Terras, and Julianne Nyhan, 23–46. Facet Publishing.

TestSigma. 2023. "Functional Testing : What It Is, How to Do & Examples." Testsigma Blog. November 29, 2023. https://testsigma.com/guides/functional-testing/#How_To_Do_Functional_Testing_8211_Manual_and_Automated.

Tomašević, Aleksandra, Biljana Lazić, Dalibor Vorkapić, Mihailo Škorić, and Ljiljana Kolonja. 2017. "The Use of the Omeka Platform for Digital Libraries in the Field of Mining." *Infotheca* 17 (2): 26–49. <https://doi.org/10.18485/infotheca.2017.17.2.2>.

Tommasini, Antonio. 2021. "La repressione delle minoranze linguistiche durante il fascismo. La lingua al servizio del potere." Roma: Luiss Guido Carli. <https://tesi.luiss.it/33399/>.

UNESCO. 2001. "Resolution Adopted by the General Assembly 56/8. United Nations Year for Cultural Heritage, 2002." <https://documents-dds-ny.un.org/doc/UNDOC/GEN/N01/475/56/PDF/N0147556.pdf?OpenElement>.

———. 2002. “United Nations Year for Cultural Heritage.” <https://unesdoc.unesco.org/ark:/48223/pf0000127155>.

———. 2003. “Resolution Adopted by the General Assembly 57/249. Culture and Development.” <https://documents-dds-ny.un.org/doc/UNDOC/GEN/N02/555/82/PDF/N0255582.pdf?OpenElement>.

———. 2017. *Re/shaping Cultural Policies. Advancing Creativity for Development*. UNESCO. <https://uis.unesco.org/sites/default/files/documents/reshaping-cultural-policies-2018-en.pdf>.

———. 2020. “Cutting Edge | Protecting and Preserving Cultural Diversity in the Digital Era.” UNESCO. October 28, 2020. <https://www.unesco.org/en/articles/cutting-edge-protecting-and-preserving-cultural-diversity-digital-era>.

Wagner, Anne, and Marie-Sophie De Clippele. 2023. “Safeguarding Cultural Heritage in the Digital Era – A Critical Challenge.” *International Journal for the Semiotics of Law - Revue Internationale de Sémiotique Juridique* 36:1915–23. <https://doi.org/10.1007/s11196-023-10040-z>.

Wallace, Andrea, and Douglas McCarthy. 2023. “Open GLAM Survey.” <https://archive.org/search?query=creator%3A%22Douglas+McCarthy+and+Dr.+Andrea+Wallace%22>: Internet Archive. <https://douglasmccarthy.com/projects/open-glam-survey/>.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. “The FAIR Guiding Principles for Scientific Data Management and Stewardship.” *Scientific Data* 3 (1): 160018. <https://doi.org/10.1038/sdata.2016.18>.

Wright, Joseph. 1917. *A Middle High German Primer*. Oxford: Clarendon Press.

Zagato, Lauso. 2015. "The Notion of 'Heritage Community' in the Council of Europe's Faro Convention. Its Impact on the European Legal Framework." In *Between Imagined Communities of Practice : Participation, Territory and the Making of Heritage*, edited by Nicolas Adell, Regina F. Bendix, Chiara Bortolotto, and Markus Tauschek, 141–68. Göttingen Studies in Cultural Property. Göttingen: Göttingen University Press. <https://doi.org/10.4000/books.gup.220>.

Zaka, Bilal. 2022. "Emerging Technologies in Digital Libraries: Net Interactive Document (NID) Experiences and Prospects." *IPSI Transactions on Internet Research* 18 (02): 64–69. <https://doi.org/10.58245/ipsi.tir.2201.09>.

Zaka, Bilal, and Hermann Maurer. 2022. "Web Sites Need Much More User-Driven Interactivity." *IPSI Transactions on Internet Research* 18 (02): 64–69. <https://doi.org/10.58245/ipsi.tir.2201.08>.

Zanella, Antonio. 1991. *Storia Dei Longobardi*. Milano: Rizzoli.

Zironi, Alessandro. 2022. "La Letteratura Gotica." In *Le Civiltà Letterarie Del Medioevo Germanico*, edited by Marco Battaglia. Roma: Carocci. <https://www.carocci.it/prodotto/le-civilta-letterarie-del-medioevo-germanico-2>.

Zuin, Francesco. 2024. "Il Timavese." *Linguistik Online* 130 (6): 149–69. <https://doi.org/10.13092/lo.129.11153>.