learners (Short et al. 2013). While some researchers have already described the use of comics in interactive systems (McCloud, Dylan et al. 2016), our goal is to investigate and nurture the potential of data comics as educational tool in tutorials on the ACDH-CH self-study platform HowTo which is a self-paced learning platform aimed at ACDH-CH researchers and the Austrian Digital Humanities community and has been created as part of the DiTAH - Digital Transformation of Austrian Humanities - project. In our project, we argue that a collaboration in merging data representation and didactics lays the groundwork for successful science communication.

Combining the flexibility of 2-dimensional spatial layout of infographics and the linearity of narration inherent to videos and live presentations (Bach et el. 2018), comics offer unique possibilities for the presentation of data but also for the representation of interaction within a system (Short et al. 2013). Recent studies on data comics have identified that the interplay of text, image, panel structure and sequence (Data Comics 2021) can enhance user-friendly science communication, depending on the comic design (Bach et al. 2018). Sequential art and using 'imagery as communicator' (Eisner 1985: 13) supports understanding of complex problems. However, learners need to be trained in visual literacy to be able to extract information successfully. Therefore, we argue that data comics make use of familiar visual cues to transport meaning.

In our case study, we will investigate how data comics can affect and improve learning outcomes and knowledge retention in the ACDH-CH's Learning Management System (LMS) HowTo (ACDH-CH 2022). We will use learning materials created for the ACDH-CH Dylen project which investigates diachronic dynamics of Austrian German lexicographic networks in authentic language data (Wissik et al. 2021). First, these materials are transformed into a number of learning paths and learning objectives to enable users to expand their knowledge on the project, following Bloom's revised taxonomy (from passively remembering and understanding to actively applying and evaluating learning objectives) (Armstrong 2010). These learning paths include a screencast, an article-style tutorial and data comics following different patterns. Second, we run a series of user tests to understand which and how much information is retained, depending on the chosen learning path. We generate a number of tutorials that allow users to choose their preferred mode of learning (text, video, data comic) and track how much of the presented knowledge is retained in each medium. We do so by conducting multiple user tests with focus groups and online questionnaires, using both quantitative and qualitative methods. Using the web analytics tool Matomo (Matomo 2022), statistical information such as the number of clicks and the amount of time users spend on each of the tutorials will be evaluated as well. Our findings will be used to improve the quality of the tutorials in the HowTo platform and, based on our findings, we aim to create guidelines for the HowTo community. Given that our poster presentation will at the same time be a data comic, we would like to conduct an in-person field test at the DH2023 where we will present our findings and invite visitors to participate in our research via their mobile devices.

Bibliography

Armstrong, Patricia (2010): "Bloom's Taxonomy", Vanderbilt University Center for Teaching https://cft.vanderbilt.edu/guides-sub-pages/blooms-taxonomy/ [02.11.2022].

Austrian Centre for Digital Humanties and Cultural Heritage (ACDH-CH) (2022): HowTo. https://howto.acdh.oeaw.ac.at/ [28.04.2023]

Bach, Benjamin / Henry Riche, Nathalie / Carpendale, Sheelagh / Pfister, Hanspeter (2017): "The Emerging Genre of Data Comics" in: *IEEE Computer Graphics and Applications* 38, 6-13. DOI: 10.1109/MCG.2017.33.

Bach, Benjamin / Wang, Zezhong / Farinella, Matteo / Murray-Rust, Dave / Henry Riche, Nathalie (2018): "Design Patterns for Data Comics", in: Association for Computing Machinery (CHI '18): *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems,* New York, NY, USA, 1–12. DOI: 10.1145/3173574.3173612.

Bouchard, Frédéric / Sansoulet, Julie / Fritz, Michael / Malenfant-Lepage, Julie / Nieuwendam, Alexandre / Paquette, Michel / Harbor, Jon (2018). "Frozen-Ground Cartoons: Permafrost comics as an innovative tool for polar outreach, education, and engagement", Polar Record, 54(5-6), 366-372. DOI: 10.1017/S0032247418000633

Data Comics (2021): "3 Essential Features of Good Comics" https://www.datacomics.net/2021/07/25/3-essential-features-of-good-comics/ [28.04.2023].

Dylan, Thomas / Wallace, Jayne / Blythe, Mark / Thomas, James (2016): "Paper Street View: A Guided Tour of Design and Making Using Comics", in: Association for Computing Machinery Systems (DIS '16): *Proceedings of the 2016 ACM Conference on Designing Interactive*, New York, NY, USA, 334–346. DOI: 10.1145/2901790.2901904.

Eisner, Will (1985): *Comics and Sequential Art.* Poorhouse Press.

Matomo (2022) https://matomo.org/ [28.04.2023].

McCloud, Scott: *Google Chrome: Behind the Open Source Browser Project.* https://www.google.com/googlebooks/chrome/ [28.04.2023].

Short, Jeremy / Randolph-Seng, Brandon / McKenny, Aaron (2013). "Graphic Presentation: An Empirical Examination of the Graphic Novel Approach to Communicate Business Concepts", in: *Business Communication Quarterly* 76, 273-303. DOI: 10.1177/1080569913482574.

Wang, Zezhong / Ritchie, Jacob / Zhou, Jingtao / Chevalier, Fanny / Bach, Benjamin (2021): "Data Comics for Reporting Controlled User Studies in Human-Computer Interaction", in: *IEEE Transactions on Visualization and Computer Graphics* 27, 2: 967-977. DOI: 10.1109/TVCG.2020.3030433.

Wissik, Tanja / Neidhardt, Julia / Baumann, Andreas / Wünsche, Katharina / Kirnbauer, Martin / Pirker, Hannes / Marakasova, Anna / Sertkan, Mete / Hofmann, Klaus / Cetin, Asil (2021): Dylen. https://dylen.acdh.oeaw.ac.at/about/ [28.04.2023]

Zhao, Zhenpeng / Marr, Rachael / Elmqvist, Niklas (2015): *Data comics: Sequential art for data-driven storytelling.* University of Maryland: HCIL Technical Report HCIL-2015-15.

Collaboration within a shared digital paradigm: opportunities and outcomes

De Bastiani, Chiara
chiara.debastiani@unive.it
Università Ca'Foscari Venezia, Italy

Fabbris, Giulia
giulia.fabbris@unive.it
Università Ca'Foscari Venezia, Italy

With this poster, we aim to present a case study on the benefits of mutual collaboration in ongoing research about digitization of Cultural Heritage in Italy, which is being carried out through two distinct, but intertwined, PON (Piano Operativo Nazionale) projects (Ministero dell'Università e della Ricerca 2014-2020). The aim of this case study is twofold: on the one hand, we will argue that both projects add value to the Italian Cultural Heritage resources by inserting them in a digital paradigm, ensuring dissemination and reuse of the data within the panorama of Cultural Heritage Digitization. On the other hand, we aim to connect a SPARQL Endpoint with an annotation and visualization app, to enable further annotation of the SPARQL data retrieved and to complement the application with a functionality designed for annotating and enriching graphs, cf. Figure 1.

More specifically, we are retrieving and analyzing digital Italian resources and Germanic Cultural Heritage resources in the Veneto region to link and systematize them in a digital environment, and make them available to both a scientific and a non-scientific audience, e.g. cultural tourism. The users can adapt their research according to their needs thanks to a user-friendly interface, (cf. http://murucaracconta.muruca.cloud/), which will be adapted to our platform with its developers at Net7.

In more detail, the first project aims at creating a knowledge graph of Germanic Cultural Heritage artifacts in the Veneto region, reusing top-level and specific ontologies such as CIDOC-CRM (Bekiari et al. 2021, V7.2.1) , ArCo (Carriero et al., 2019) CIDOC-FRBRoo (Bekiari et al., 2015, V2.4) and Bibframe Vocabulary 2.0 (Library of congress) and using cataloging metadata complying with the EDM standards (Europeana Foundation, 2017, V2.4), to make them searchable through a SPARQL Endpoint (Gearon et al., 2013). The aim of the second research project is to create a platform with two applications integrated, one for the visualization and the other for the annotation of the resources. The data available will be retrieved from existing repositories through APIs, ensuring the reuse of information and, thanks to the addition of uniform tags, their interoperability.

The resources are designed both for academic and non-academic users. Scholars can have access to the APIs, XML files and IIIFs, when the original source provides them. The tourist can find information on a specific topic related to the Italian CH. Moreover, the model can be further reused by researchers and applied to existing ontologies and derived SPARQL queries. The retrieved data will be available from two different visualization interfaces, which have as a first common goal the valorization of Italian Cultural Heritage by inserting it in a digital paradigm. In fact, the projects are in line with the Italian National Digitalization Plan (Ministero della Cultura 2022-2023), whose purpose is to create a cultural ecosystem based on digital methods in order to strengthen existing digitalization projects and offer public policies and rules to operate in a common vision.

In fact, the online availability of resources is not enough, they need to be integrated in formally coherent, supplementable and reusable representations (Gagliardi/Guarino 2021). As a second goal of our mutual collaboration case study, we will present the results of experimental work, with the aim of enriching the visualization and annotation application with a specific functionality to further annotate and enrich data retrieved from SPARQL queries operated in the SPARQL endpoint created within the first project. For this, we are collaborating with Net7, a non-academic partner institution that operates in the DH field.

Summarizing, the shared outcomes from mutual collaboration are (1) scientific, since the projects allow to add value to existing cultural data and improve access to the resources because items divided in different heterogeneous platforms will be available in one single place and become interoperable; (2) social, since different types of users will benefit from the resources developed; (3) economical, thanks to the collaboration with partner non-academic institutions (Net7), which enhances the creation of professional profiles with interdisciplinary competences.

Finally, the development of a common digitization paradigm is a shared desideratum within the wider DH community, and efforts to harmonize metadata standards and enhance collaboration have proven fruitful in sustaining and expanding the functionality of information retrieval and modeling, and in supporting the management of complex digital data aggregates (cf. Doerr/Stear 2011, Wickett et al. 2013). Therefore, we argue that the shared outcomes stemming from the collaboration within the two projects are consistent with both national and international requirements in the digitization of Cultural Heritage.

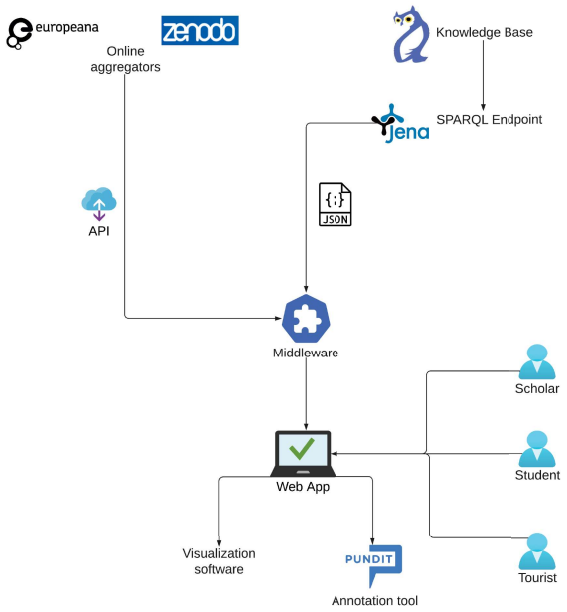

Figure 1- The Workflow

Bibliography

Bekiari, Chrysoula / Bruseker, George / Canning, Erin / Doerr, Martin / Michon, Philippe / Ore, Christian-Emil / Stead, Stephen / Velios, Athanasios (2022): Definition of the CIDOC Conceptual Reference Model, Version 7.1.2.

Bekiari, Chryssoula / Doerr, Martin / Le Bœuf, Patrick / Riva, Pat (2015): FRBR, object-oriented definition and mapping from FRBRER, FRAD and FRSAD, approved by IFLA, Version 2.4.

Carriero, Valentina Anita / Gangemi, Aldo / Mancinelli, Maria Letizia / Marinucci, Ludovica / Nuzzolese, Andrea Giovanni / Presutti, Valentina / Veninata, Chiara (2019): "ArCo: The Italian Cultural Heritage Knowledge Graph", In: The Semantic Web – ISWC 2019. ISWC 2019. Lecture Notes

in Computer Science, vol 11779. Springer, Cham. https://doi.org/10.1007/978-3-030-30796-7_311779:36–52

Doerr, Martin / Stead, Stephen (2011): "Harmonized models for the Digital World: CIDOC CRM, FRBROO, CRMDig and Europeana EDM". In Tutorial. 15th International Conference on Theory and Practice of Digital Libraries –TPDL. Abstract. https://cidoc-crm.org/Resources/harmonized-models-for-the-digital-world-cidoc-crm-frbroo-crmdig-edm.

Europeana Foundation: Europeana Mapping Guidelines, Version 2.4, https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Mapping_Guidelines_v2.4_102017.pdf

Gagliardi, Gloria / Guarino, Massimo (2021): "Risorse e applicazioni computazionali per l'accesso ai beni culturali: il Corpus CHerIDesCo". CHIMERA: Revista De Corpus De Lenguas Romances Y Estudios Lingüísticos, 8: 25-43. https://revistas.uam.es/chimera/article/view/13625/13639

Gearon, Paula / Passant, Alexandre / Polleres, Axel (2013): SPARQL 1.1 Update, W3C Recommendation, 21 March 2013, http://www.w3.org/TR/2013/REC-sparql11-update-20130321. Latest version available at http://www.w3.org/TR/sparql11-update.

Wickett, Karen M. / Isaac, Antoine / Fenlon, Katrina S. / Doerr, Martin / Meghini, Carlo / Palmer, Carole L. / Jett, Jacob (2013): Modeling Cultural Collections for Digital Aggregation and Exchange Environments, CIRSS TECHNICAL REPORT 201310-1 OCTOBER, 2013

Library of congress: Bibliographic Framework Initiative, Bibframe Vocabulary 2.0, https://www.loc.gov/bibframe/implementation/

Ministero della Cultura (2022-2023): Linee guida per l'acquisizione, la circolazione e il riuso delle riproduzioni dei beni culturali in ambiente digitale, https://docs.italia.it/italia/icdp/icdp-pnd-circolazione-riuso-docs/it/consultazione/index.html

Ministero dell'Università e della Ricerca (2014-2020): Piano Operativo Nazionale PON, Asse IV Innovazione, http://www.ponricerca.gov.it/opportunita/react-eu-contratti-di-ricerca-innovazione-e-green/

Collecting Strike Data from Historical Newspapers (19th Century): A Digital Workflow

Aurich, Jens

jens.aurich@hotmail.de
International Institute of Social History (IISH), Amsterdam

The epistemological revolution spearheaded by Digital Humanists emerges at the tail end of a global wave of workers' mobilization that originated in the second half of the 19th century. This historical junction begs the question, if and how the new toolbox of a Digital Historian can contribute to the incipient new global history of collective labor action (Linden 2012).

To that end, I create and implement a workflow for the retrieval and annotation of text relevant for the study of strikes from three major collections of digitized German historical newspapers. To the growing Digital History literature concerned with the exploitation of large collections of digitized newspapers, this project contributes a proof of concept for the suggested workflow, an in-depth quality review of a selection of newspapers from three major German digital collections, a re-OCR pipeline and a segmentation approach to separate news reports. To the historiography of strikes in the first two decades of the German Empire, it contributes a novel data resource and a first assessment of the coverage of domestic labor conflicts in major newspapers.

Data

The core corpus consists of 42,000 daily issues (1870-1890) of six newspapers selected from the three largest collections of retrodigitized historical German newspapers for the late nineteenth century (*DigiPress*, *ZEFYS* and *zeitpunkt.NRW*). The selection of newspaper titles follows existing newspaper-based studies of social unrest in the 19 th century (Tilly 1970) and includes two major interregional newspapers (*Kölnische Zeitung* and *Augsburger Allgemeine Zeitung)* and four regional newspapers published in Berlin and Bavaria. Secondary data sources are existing microdata on labor conflicts in the German Empire.

Workflow

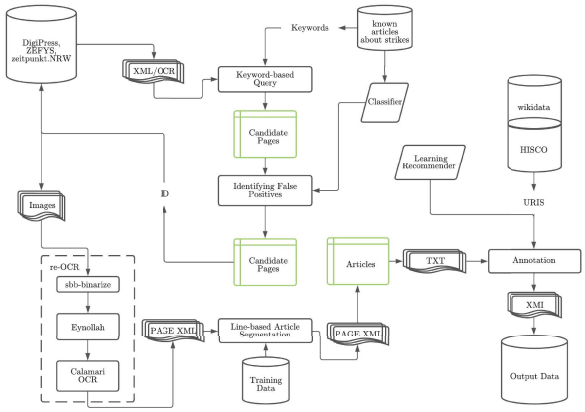

Figure 1: Workflow

The workflow consists of three steps: First, I create a subcorpus of candidate newspaper pages by identifying relevant historical keywords (Oberbichler / Pfanzelter 2021, Pfanzelter et al. 2021) in a bottom-up manner using a statistical approach (TF/IDF) (Zervanou et al. 2014) and a novel ground truth dataset of newspaper articles collected based on source references in existing quantitative sources. To increase the accuracy of the corpus, I identify false positives with a text classifier trained on a tagged set of 2000 random sentences embedded using a *FastText* model (Ehrmann et al. 2020) with subword features (F-score = 0.91).