A Case Study in Cultural Heritage: A System Linking Three Open Data Tools – Digital Philology for Dummies (DPhD), Edition Visualization Technology (EVT), and a Relational Database

Renato Caenaro[1], Chantal Pivetta[2]
[1] Silentwave SRLS, Italy – renato.caenaro@silentwave.eu
[2] Lund University, Sweden – chantal.pivetta@rom.lu.se

ABSTRACT (ENGLISH)

This contribution presents a practical case study focused on Cultural Heritage, specifically the study of Cassinese monasteries in Italy. The emphasis is on the methodologies and technologies used within an inclusive and open science framework. The project is inherently multidisciplinary, involving art history, architectural history, computer science, engineering, and philology, and aims to connect data across multiple platforms, including text visualizers, maps, relational databases, and engineering-architectural models.

The methodology is designed to accommodate the diverse expertise of the team, which includes members without technical backgrounds and from various academic disciplines. The presentation focuses on two key phases of the workflow: semantic annotation and encoding of documentary sources. The focus is on two platforms: Digital Philology for Dummies (DPhD) and Edition Visualization Tecnhology (EVT). Finally, complex analysis and interconnection of data.

The project establishes a robust data-sharing framework that promotes the early adoption of open data and open access principles. This ensures that data can be published immediately in alignment with open data and open access philosophies. By implementing structured open APIs, the data becomes accessible not only as complete database downloads but also as interconnected, linked data, thereby enhancing accessibility and application potential.

This methodology demonstrates how collaborative tools, interdisciplinary workflows, and linked data technologies can advance research in Cultural Heritage while ensuring inclusivity, accessibility, and open knowledge sharing.

Keywords: Open Data; XML/TEI; Interdisciplinary Research; Relational Database

ABSTRACT (ITALIANO)

Un caso studio per il patrimonio culturale: un sistema che collega tre strumenti Open Data - Digital Philology for Dummies (DPhD), Edition Visualization Technology (EVT), e un Database Relazionale

Questo contributo presenta un caso studio pratico, inerente al campo del Cultural Heritage, che ha come obbiettivo lo studio dei monasteri cassinesi in Italia. Si vuole illustrare il sistema e le tecnologie utilizzate, in un contesto inclusivo e orientato ai principi dell'open science. Il progetto è multidisciplinare e coinvolge esperti di storia dell'arte, storia dell'architettura, informatica, ingegneria e filologia, con lo scopo di connettere dati attraverso strumenti diversi, tra cui visualizzatori di testi, mappe, database relazionali e modelli ingegneristico-architettonici.

Il metodo è stato progettato considerando la provenienza multidisciplinare dei membri del team – costituito anche da risorse senza conoscenze informatiche – creando quindi un approccio strutturato che risponda alle diverse esigenze del gruppo. Il contributo si concentra su due fasi principali del flusso di lavoro: prima la codifica e annotazione delle fonti documentarie. Si evidenzierà il ruolo di Digital Philology for Dummies (DPhD) e Edition Visualization Technology (EVT). Segue l'analisi dei dati e loro connessione con altri sistemi.

Il progetto dispone di dati immediatamente pubblicabili e accessibili secondo i principi openData e openAccess. Grazie all'implementazione di openAPIs strutturate, i dati sono accessibili non solo come download completi, ma anche come dati collegati e interconnessi, ampliandone le possibilità di utilizzo e applicazione.

Questo approccio dimostra come l'uso di strumenti collaborativi, metodologie interdisciplinari e tecnologie di linked data possa migliorare la ricerca nei Beni Culturali, garantendo inclusività, accessibilità e una condivisione aperta della conoscenza.

Parole chiave: Open Data; XML/TEI; Ricerca Multidisciplinare; Database Relazionale

1. INTRODUCTION

The study of Cultural Heritage, particularly the historical and architectural significance of the Cassinese monasteries in Italy, has been increasingly enriched by the integration of digital methodologies and open science frameworks. This contribution presents a practical case study- a pivotal project- focusing on the methodologies and technologies employed to facilitate access to and the interlinking of Digital Humanities data resources, including text visualizations, geospatial mapping tools, relational databases, and engineering-architectural models. The outcomes enhance the inclusivity and efficiency of digitization, annotation, and data integration processes. Moreover, the study underscores how this approach supports the principles of open data, promoting long-term accessibility, sustainability, and interdisciplinary collaboration.

The significance of this pilot project and the resulting system lies in two key aspects: 1. The system is designed to be reusable in multidisciplinary projects involving large teams with diverse backgrounds, where contributions must be monitored and approved. 2. The system demonstrated how research interests may evolve after the initial data collection phase, offering a valuable perspective when confronted with a large volume of information to be processed. In the case study, the initial focus was on monasteries and the artworks they contained. However, following the preliminary results, the project's emphasis shifted toward documentary sources for the definition of the Entity-Relationship (ER) model, with a particular focus on Named Entities and their interrelations.

2. CASE STUDY

The PRIN project *"CoenoBIuM"* (www.digitalcoenobium.eu) – coordinated by Gianmario Guidarelli (University of Padua), in collaboration with Sonia Cavicchioli (University of Bologna) and Paolo Borin (University of Brescia) – investigates the architectural and artistic culture of the Cassinese Benedictine Congregation from its foundation through the late 18th century. It adopts a comparative perspective while experimenting with the innovative application of Building Information Modeling (BIM). The reform initiated by Ludovico Barbo in 1419 profoundly reshaped monastic life, redirecting its focus toward prayer and contemplation, and leaving a significant impact on architectural spaces and artistic expression within churches and monasteries.

Scholars have long hypothesized the existence of a shared architectural and visual culture across the Congregation, fostered by the common regulations issued by the Capitoli Generali. To test this hypothesis, the project aims to reconstruct the original relationships between decorative programs and built environments, integrating a wide range of heterogeneous sources (archival, iconographic, material, and spatial).

The pilot project presented here focuses on the development of a system that enables the interconnected digitization of information across multiple disciplines. The system includes the design of an Entity-Relationship (ER) schema, the implementation of a relational database, the development of a data entry portal, the publication of open APIs, the integration of authoring software and text visualizator.

REQUIREMENTS AND CHALLENGES

This paper addresses only a small portion of the complex system that was developed. The main requirements and challenges encountered during its development were:

1. The need to account for practical constraints, such as the cost and time availability of the participants involved in the project.
2. The varying levels of digital literacy among participants necessitated the use of user-friendly technological tools to ensure satisfactory task completion.
3. The imperative to guarantee interoperability and long-term sustainability of the data produced through the various technological tools adopted.

METHODOLOGIES EMPLOYED

The methodology is structured around two key phases: 1. Semantic Annotation and Encoding: The project use TEI guidelines to transcribe, annotate, and encode documentary sources collaboratively. A user-friendly, shared environment allows participants from diverse fields and technical skill levels to contribute effectively. The resulting structured data is stored in a relational database, enabling systematic organization and query-based retrieval. 2. Complex Data Analysis and Integration: Structured data is used to build tools that enable relational and geospatial analysis, linking encoded sources to external systems such as mapping platforms. This integration supports the creation of new insights, facilitating connections between textual, architectural, and geographical data. Throughout both phases, a semi-structured data

environment is employed, which supports the semi-automatic generation of relationships between datasets from various disciplines. OpenAPIs ensure that the data can be accessed as both complete database dumps and interconnected linked data.

3. FIRST STEP: ENCODING AND TRANSCRIBING IN XML/TEI

The workflow's first phase focuses on transcribing and encoding documentary sources using *Digital Philology for Dummies* (DPhD). This user-friendly and open-data software is designed to simplify and accelerate the transcription and encoding of texts following TEI guidelines. It is specifically created to assist users without technical expertise or knowledge of XML/TEI while offering advanced functionality for experienced users.

In the project's initial phase, *Digital Philology for Dummies* (DPhD) was chosen for two primary reasons. First, the group of transcribers was highly heterogeneous, composed of participants with no prior experience in XML/TEI encoding and representing a range of disciplinary backgrounds. As illustrated in the figure below, the interface is intuitive and accessible even for users unfamiliar with TEI. It allows the image and transcription to be displayed side by side, with the corresponding XML/TEI encoding automatically generated in real time. Semantic annotation is performed by selecting the relevant portion of text and applying predefined tags. These tags are standardized across all users and curated by more experienced contributors.

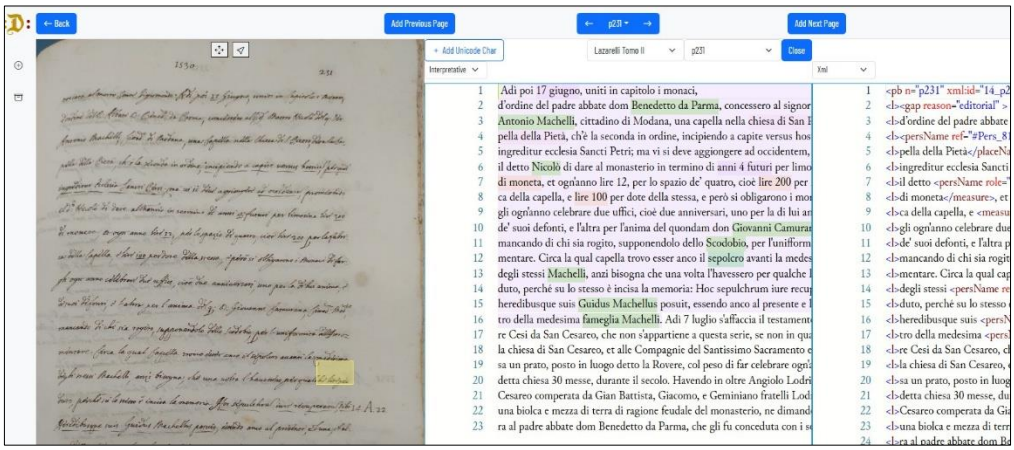

Figure 1 DPhD Interface

Second, a shared workspace was needed, in which all transcribers could collaborate following a unified set of rules to ensure consistency and standardization. Within the workspace, each transcriber is assigned a specific role by the project coordinator. The activities carried out by the transcribers are summarized in the following figure:

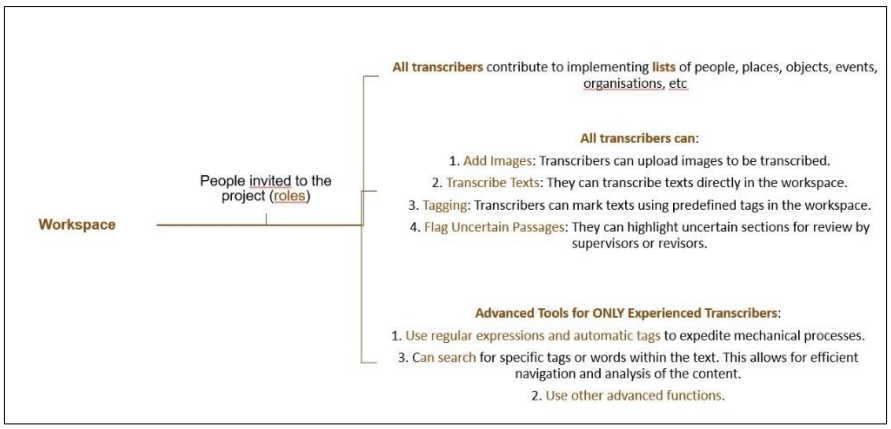

Figure 2 Transcription workspace

A critical step in the process, successfully managed, is the creation of unique identifiers (@xml:id), which is managed by the relational database. The database generates these identifiers through the dynamic creation of APIs, which continuously update the system. This workflow is highly interactive and iterative: transcribers input textual data into the platform and, in return, receive @xml:id values, which are

automatically integrated into the annotated texts. One example: among the main research questions, one focused on the study of artworks preserved within the monasteries—specifically, identifying where within the monastery these works were conserved, whether they were relocated to other monasteries over time, and how their context evolved. For this purpose, the artworks were identified and annotated using XML/TEI encoding. When a transcriber encountered a reference to a work of art in the text, they would input its metadata into a dedicated data entry form linked to the relational database (see figure 3).

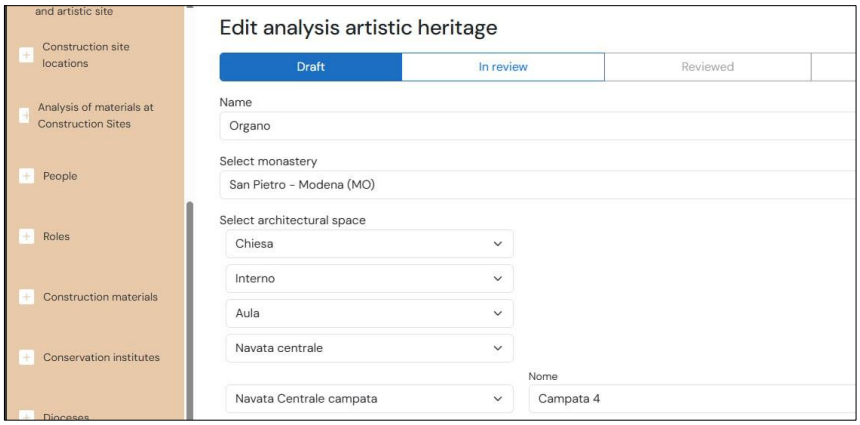

Figure 3 Coenobium Data entry for an organ

The database would then generate a unique @xml:id, which the transcriber would insert into the corresponding markup within the annotated text (see figure 4).

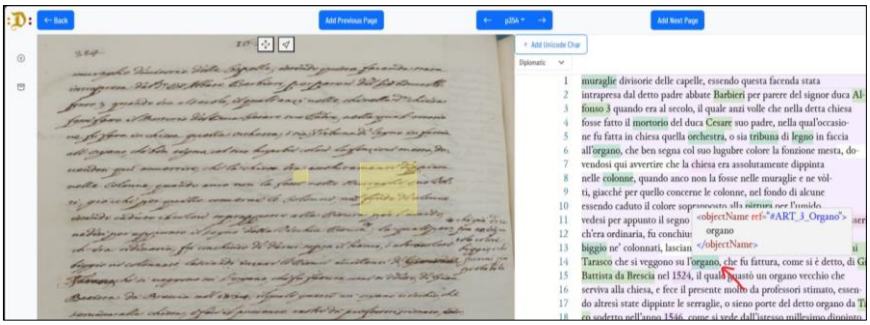

Figure 4 DPhD transcription of an organ

The database, in turn, links each artwork to additional elements, such as the location where it is preserved, the dates on which it is mentioned in documentary sources, its commissioner, cost, materials used, and other relevant attributes. This information is accessible also from the 3D model selecting the image:

Figure 5 3D model of an organ

Finally, regarding the visualization of these texts (in progress), another advantage of using DPhD is that the software is directly integrated with Edition Visualization Technology (EVT), enabling a smooth workflow that combines transcription, encoding, and visualization in a seamless process. DPhD uses a new feature

called 'EVT as a service,' enabling the real-time display of XML/TEI-encoded documents directly within EVT, without requiring installation or configuration. This feature simplifies its use for viewing general textual data, even beyond the specific context of digital scholarly editions Edition Visualization Technology (EVT) is an open data and open source platform specifically designed for visualizing and navigating texts encoded in XML/TEI. The software has reached its third version, which focuses on synoptic editions—a feature that has proven to be particularly effective for this project. The synoptic view allows the visualization of a diverse collection of texts in multiple languages, enabling users to see how they relate to one another. This capability is essential for addressing the various queries posed by the study of Cassinese monasteries, where the ability to navigate and analyze relationships between texts is critical to uncovering insights. This makes EVT an ideal tool for handling the complexity and heterogeneity of the project's textual data.

4. DATA ENTRY: WHY NOT ONLY RDF?

Concerning data entry, a hybrid system was adopted. While RDF excels at representing data with strong semantic structure, its use can be excessive in application domains that require simpler modeling or where semantic complexity is only marginal. In such cases, more traditional approaches—such as relational databases—may prove more appropriate and efficient.

Although RDF is designed to promote interoperability and data sharing, integrating data entry tools developed across different disciplinary domains into an RDF-based framework would have required significant customization. We also considered the perspective of the end users of the platform—namely, the data entry operators—who might not feel comfortable with the input modalities imposed by an RDF system, particularly when interfaces must conform to complex metadata structures. For this reason, we prioritized offering a more guided and user-friendly experience.

Finally, from a cost perspective, developing and maintaining a system based purely on RDF would have been significantly more expensive due to the need for specialized tools, advanced technical expertise, greater oversight of data input, and highly customized integration processes.

5. SYSTEM INTEROPERABILITY

The *Coenobium* project, as mentioned before, involved a range of disciplinary fields, which can be broadly categorized into four main areas: architectural history, engineering, art history, and philology. Regardless of the number or type of disciplines involved, each field brings its specificities, distinct research objectives, and, within the context of the Digital Humanities, specialized digital tools. It was therefore essential to develop a system capable of interconnecting and integrating the information produced by each discipline. To this end, an Entity-Relationship (ER) schema was designed to define the core entities relevant to the project, along with their relationships and constraints. Based on this schema, a relational database (RDB) was implemented to enable the structured mapping of entities and their interrelations.

A data entry portal was also developed, accessible to all project participants, allowing for the guided registration of entities through a structured interface. The system includes attribution of entered data and a workflow comprising review and approval phases.

In parallel, a set of open APIs was created to deliver the data stored in the RDB in various formats (XML, XML-TEI, JSON). Each software tool used within the different disciplinary domains can query these APIs to retrieve the information needed for generating project outputs, which may take diverse forms such as XML-TEI files, PDFs, images, 3D models (IFC, OBJ-GLTF), point clouds, and others.

Finally, a shared data environment was established, where each output is collected and becomes an extension of the RDB. This infrastructure allows users, within their respective software environments, to access information produced by other disciplines—even without direct knowledge of their tools or datasets—since the content is made available through structured open APIs based on a common schema.

6. EXAMPLE OF ENTITY RELATIONSHIP SCHEMA

As an example, we present below a subset of the Entity-Relationship (ER) schema from the DBR database, which illustrates the connections between the Named Entities <people> and those identified as <object type="MonasteryComplex"> and <event type="ConstructionSite"> within the textual sources.

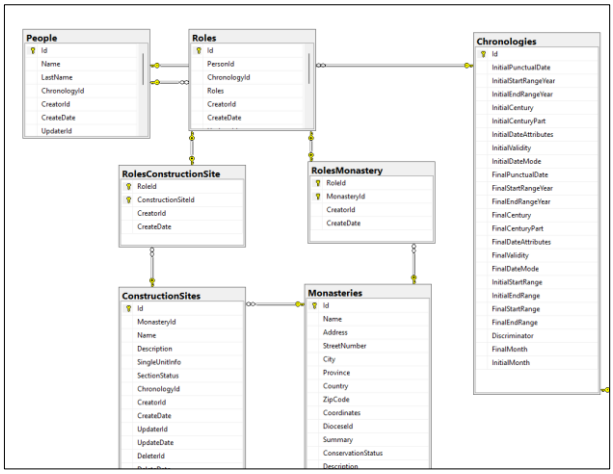

Figure 6 Portion of ER Schema

As can be observed, the relationships between people, construction sites, and monasteries are conceptually structured as RDF-like triples. However, in practice, the attribute corresponding to the "role" within each triple has been externalized into a dedicated table, as it is linked to a temporal dimension. This chronological component allows for the specification of start and end dates, which can be expressed as precise dates, intervals, or centuries, and includes the option to assign a degree of certainty to each temporal statement.

This schema plays a crucial role in the data entry phase, as it enables not only the normalization of data but also the validation of input. Furthermore, a public HTTP-based openAPI interface is provided, both for project participants and for the software systems that utilize the data. This API also serves external users, allowing for the mapping of these complex entities into a more canonical form that is compatible with SPARQL-based querying environments.

7. RESULTS

The outcome is a sustainable and interoperable system, capable of integrating large volumes of data from multiple disciplinary domains. Importantly, it also allows individuals without advanced digital expertise to contribute meaningfully to the enrichment of the project. Moreover, the system is replicable and adaptable to other initiatives involving architectural heritage and documentary sources, whether historical or contemporary.

More specifically, the methodology has already demonstrated significant potential in:

1. Supporting advanced textual and relational analysis through the use of structured data encoded in TEI and managed within relational databases.
2. Improving the accessibility and usability of historical data through visualization tools such as EVT and geospatial mapping platforms.
3. Facilitating collaborative workflows that include participants with varying levels of technical expertise.
4. Promoting the principles of open science and data interoperability by publishing data in alignment with open data and open access frameworks.

REFERENCES

Andrews, T. L. (2013). The Third Way: Philology and Critical Edition in the Digital Age. *The Journal of the European Society for Textual Scholarship*, *10*, 61–76. https://doi.org/10.1163/9789401209021_006

Baker C., Cheung K.-H. (eds) (2007). Semantic Web—Revolutionizing Knowledge Discovery in the Life Sciences. Berlin/Heidelberg: Springer.

Bleier, R., Bürgermeister, M., Klug, H. W., Neuber, F., & Schneider, G. (Eds.). (2018). *Digital Scholarly Editions as Interfaces* (Vol. 12). BoD. https://kups.ub.uni-koeln.de/9085/

Burghart, M. (2017). *Creating a Digital Scholarly Edition with the Text Encoding Initiative*. https://halshs.archives-ouvertes.fr/halshs-01984319

Cheng G., Shao F., Qu Y. (2017). An empirical evaluation of techniques for ranking semantic associations. IEEE Transactions on Knowledge and Data Engineering, 29(11): 1.

Ciotti, F., Corradini, E., Cugliana, E., D'Agostino, G., Ferroni, L., Fischer, F., Lana, M., Monella, P., Roeder, T., Turco, R. R. D., & Sahle, P. (2022). Manifesto per le edizioni scientifiche digitali. *Umanistica Digitale*, *12*, Article 12. https://doi.org/10.6092/issn.2532-8816/14814

Di Pietro, C., & Rosselli Del Turco, R. (2018). Between Innovation and Conservation: The Narrow Path of User Interface Design for Digital Scholarly Editions. In *Digital Scholarly Editions as Interfaces* (Vol. 12, pp. 133–163). BoD. https://kups.ub.uni-koeln.de/9115/

Eero, H. Heikki, R. (2021). Knowledge-based relational search in cultural heritage linked data, Digital Scholarship in the Humanities, Volume 36, Issue Supplement_2, Pages ii155–ii164, https://doi.org/10.1093/llc/fqab042

Favretto, A. (2022). Relational Database, GIS Layers, and Geodatabase for Cultural Heritage Management In: D'Amico, S., Venuti, V. (eds) Handbook of Cultural Heritage Analysis. Springer, Cham. https://doi.org/10.1007/978-3-030-60016-7_46.

Hyvönen, E. (2020). Using the Semantic Web in Digital Humanities: shift from data publishing to dataanalysis and serendipitous knowledge discovery. Semantic Web, 11(1):187–93.

Hyvönen E. (2012). Publishing and Using Cultural Heritage Linked Data on the Semantic Web. Palo Alto, CA: Morgan & Claypool.

Indrawan-Santiago, M. (2012). Database research: Are we at a crossroad? Reflection on NoSQL. In 2012 15th International Conference on Network-Based Information Systems (pp. 45-51). IEEE.

Monella, P., & Rosselli Del Turco, R. (2020). Extending the DSE: LOD Support and TEI/IIIF Integration in EVT. In C. Marras, M. Passarotti, G. Franzini, & E. Litta (Eds.), *Atti del IX Convegno Annuale dell'Associazione per l'Informatica Umanistica e la Cultura Digitale (AIUCD). La svolta inevitabile: Sfide e prospettive per l'Informatica Umanistica* (pp. 148–155). Available online as a supplement of Umanistica Digitale: https://umanisticadigitale.unibo.it. https://doi.org/10.6092/UNIBO/AMSACTA/6316

Robinson, P. M. W. (2022). An approach to complex texts in multiple documents. *Digital Scholarship in the Humanities*, fqab108. https://doi.org/10.1093/llc/fqab108

Rosselli Del Turco, R. (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

Rosselli Del Turco, R., Buomprisco, G., Di Pietro, C., Kenny, J., Masotti,R., Pugliese, J. (2015). Edition Visualization Technology: A Simple Tool to Visualize TEI-based Digital Editions. *Journal of the Text Encoding Initiative*, *Issue 8*, 1–21. https://doi.org/10.4000/jtei.1077

Schreibman, S., & Papadopoulos, C. (2019). Textuality in 3D: Three-dimensional (re)constructions as digital scholarly editions. *International Journal of Digital Humanities*, *1*(2), 221–233. https://doi.org/10.1007/s42803-019-00024-6

Tamper M., Leskinen P., Apajalahti K., Hyvönen E. (2018). Using biographical texts as linked data for prosopographical research and applications. In Digital Heritage. 7th International Conference on Progress in Cultural Heritage: Documentation, Preservation, and Protection, EuroMed 2018, Springer, Nicosia, Cyprus.