



# **BIOGRAPHICAL DATA IN A DIGITAL WORLD 2022**

*Proceedings of the Biographical Data in a Digital World 2022 (BD 2022) Workshop,  
co-located with the Digital Humanities 2022 (DH2022) conference in Tokyo*

Edited by EERO HYVÖNEN, MIKKO KOHO, ANGEL DAZA and GREGOR POBEŽIN

***Biographical Data in a Digital World 2022***

*Proceedings of the Biographical Data in a Digital World 2022 (BD 2022) Workshop, co-located with the Digital Humanities 2022 (DH2022) conference in Tokyo*

Edited by Eero Hyvönen, Mikko Koho, Angel Daza and Gregor Pobežin

**Authors** Anja Grebe, Eero Hyvönen, Stefan Jänicke, Nina Janz, Mikko Koho, Jakob Kusnick, Petri Leskinen, Johannes Liem, Paul Longley Arthur, Eva Mayr, Rachel Pierce, Gregor Pobežin, Isabel Smith, Neža Zajc, Florian Windhager

**Reviewers:** Matija Ogrin, David Movrin

**Design and layout:** Nina Semolič & CEUR

**Publisher:** Založba ZRC, ZRC SAZU

**Represented by:** Oto Luthar

**Issued by** ZRC SAZU Institute of Cultural History

**Represented by** Gregor Pobežin

**Založba ZRC editor-in-chief** Aleš Pogačnik

Ljubljana 2024

The first e-edition of the book is freely available under a Creative Commons licence CC BY-NC-ND 4.0: <https://doi.org/10.3986/9789610508120>

Cover photo: Einhard (775-840), medieval biographer of Charlemagne.

Katalogni zapisi o publikaciji (CIP) pripravili v Narodni in univerzitetni knjižnici v Ljubljani

COBISS.SI-ID=180599043

ISBN 978-961-05-0812-0 (PDF)



# Contents

## **Preface**

Eero Hyvönen, Mikko Koho, Angel Daza, Gregor Pobežin

## **The Promise of the Finding Aid: A Critical Approach to Finding Biographical (Linked) Data in the Archive**

Rachel Pierce

## **Dante Alighieri (1265–1321) and St Maximus the Greek (1470–1556): Two Authors on the Edge of Two Epochs (Pre-case Biographical studies for the Comparison of the Author’s Creative Vision – The Triangle Composition)**

Neza Zajc

## **Mapping biographies in a Relational Database. Biographies of Luxembourgish soldiers in the Second World War**

Nina Janz

## **Creating and Using Biographical Dictionaries for Digital Humanities Based on Linked Data: A Survey of Web Services in Use in Finland**

Eero Hyvönen

## **Biographical Research and Digital Mapping**

Paul Longley Arthur, Isabel Smith

## **Annotation of Named Entities in Medieval and Early Modern Epigraphic Texts**

Gregor Pobežin

## **Biographical and Prosopographical Analyses of Finnish Academic People 1640–1899 Based on Linked Open Data**

Petri Leskinen, Eero Hyvönen

## **Studying Occupations and Social Measures of Perished Soldiers in WarSampo Linked Open Data**

Mikko Koho, Eero Hyvönen

## **Traveling with Albrecht Dürer - A Case Study for Uncertainty-Aware Biography Visualization**

Florian Windhager, Eva Mayr, Johannes Liem, Jakob Kusnick, Stefan Jänicke, Anja Grebe



# Preface

Eero Hyvönen<sup>1,2</sup>, Mikko Koho<sup>1</sup>, Angel Daza<sup>3</sup>, and Gregor Pobežin<sup>4</sup>

*1 Aalto University*

*2 University of Helsinki, Helsinki Centre for Digital Humanities (HELDIG)*

*3 Vrije Universiteit Amsterdam*

*4 Institute of Cultural History, ZRC SAZU*

This volume contains the proceedings of the fourth edition of the Biographical Data in a Digital World Conference (BD2022), held in Tokyo, Japan, as a co-located workshop of the Digital Humanities 2022 conference. BD2022 started with three sessions for presenting papers on 1) Network Analysis and Semantic Web (five papers), 2) Finding and Preparing Biographical Data for Research (three papers), and 3) Use Cases and Advanced Ways of Working with Biographies (five papers). After this there was a separate interactive session for exchanging ideas, discussion, open challenges, and sharing resources. The proceedings of the event include the nine full papers that were accepted for publication after a peer review process.

## 1. Introduction

The first three conferences of Biographical Data in a Digital World in 2015 [1], 2017 [2], and 2019 [3] provided for many researchers the first opportunity to connect with other researchers working on digital biographical resources. These conferences brought together a wide variety of perspectives from the points of view of history, library and information science, literature studies, computer science, and computational linguistics. Despite the different perspectives and new angles, new interdisciplinary commonalities were found: a shared interest in the richness and variety of the biographical resources, approaches to natural language processing, data modelling, harmonization, visualization, analysis, and challenges with gaps in the data and its quality.

The BD conference series shows that projects in various countries are steadily making progress, digital resources are growing, and methods for publishing and using them are improving. There is a need for a community in the field and for organizing events on an international level in the future, too. As a community, we want not only to exchange ideas, learn from each other, and share solutions, but we are also interested in identifying cross-border connections between biographies and among the data resources. This volume manifests the motivation for continuing to organize the BD conference series.

## 2. Biographical Data in a Digital World 2022

The fourth edition of Biographical Data in a Digital World took place in Tokyo, Japan, on July 25, 2022. The conference was co-located with the international Digital Humanities 2023 conference (DH 2023), the leading conference series in the field of Digital Humanities. This time both conferences were virtual due to the Covid-19 pandemic. The workshop organizing committee included Angel Daza and Antske Fokkens from the Vrije Universiteit Amsterdam, Richard Hadden from the Austrian Academy of Sciences, Eero Hyvönen and Mikko Koho from the Aalto University and University of Helsinki, and Eveline Wandl-Vogt from the Ars Electronica Research Institute and Austrian Academy of Sciences.



A call for papers was opened for the workshop participants to submit abstracts of 500–1000 words on the topics of the workshop, including but not limited to:

- Digitizing and structuring biographical data
- Standards, vocabularies and best practices for processing biographical data
- Biographies and Linked Data
- Crowdsourcing biographical data
- Automatic biography generation
- Using biographical and prosopographical data for quantitative analyses
- Canonization of people and events in history
- Use of big data for biographical research
- Dealing with biographical data in heterogeneous datasets
- Creating and maintaining biographical dictionaries
- Enriching biographies from external sources
- Reconciling persons between biographical dictionaries
- Reconciling names against a biographical dictionary
- Visualizing biographical and prosopographical data
- Network analysis of biographical data
- Biographies and spatial analysis
- Biographies across countries and cultures

Altogether 20 submissions were received and reviewed by the members of the workshop Programme Committee: Mikko Koho, Aalto University, chair; Eero Hyvönen, Aalto University and University of Helsinki (HELDIG), chair; Eveline Wandl-Vogt, Ars Electronica Research Institute and Austrian Academy of Sciences, chair; Paul Arthur, Edith Cowan University; Angel Daza, Vrije Universiteit Amsterdam; Thierry Declerck, German Research Center for Artificial Intelligence (DFKI); Antske Fokkens, VU University Amsterdam and Eindhoven University of Technology; Richard Hadden, Austrian Academy of Sciences; Petri Leskinen, University of Helsinki; Rennie Mapp, University of Virginia; Johannes Scholz, Graz University of Technology; Lik Hang Tsui, City University of Hong Kong; Jouni Tuominen, University of Helsinki and Aalto University; Hongsu Wang, Harvard University; David Joseph Wrisley, New York University Abu Dhabi and Princeton. Thirteen abstracts were accepted and presented in the workshop.

A separate call for full papers was issued after the event for the workshop participants. Finally, 9 full papers were selected after a second peer review round of the Program Committee and are now included in this proceedings volume.

The BD conference proceedings have traditionally been published by the CEUR Workshop Proceedings [1–3]. In 2023, however, this publisher decided to publish only proceedings of computer science workshops, excluding the interdisciplinary field of Digital Humanities, and a new publisher for the BD2022 conference had to be found.

### **3. Overview of Papers**

The papers in the conference and the proceedings cover three themes: 1) Network Analysis and Semantic Web, 2) Finding and Preparing Biographical Data for Research, and 3) Use Cases and Advanced Ways of Working with Biographies.

#### **The Promise of the Finding Aid: A Critical Approach to Finding Biographical (Linked) Data in the Archive**

This article by Rachel Pierce addresses the digital finding aid as a potential source of biographical data grounded in the contextual information. A set of finding aids of a selection of personal and organizational collections from the Schlesinger Library at Harvard University and the Library of Congress are used as examples. It is argued that embedding digital technologies in finding aids may offer a more equitable and less hierarchical method of assembling biographical data, as well as new doors into physical archives.

#### **Dante Alighieri (1265–1321) and St Maximus the Greek (1470–1556): Two Authors on the Edge of Two Epochs (Pre-case Biographical studies for the Comparison of the Author’s Creative Vision – The Triangle Composition)**

Neža Zajc presents in this paper an in-depth study of two attempts at biographies of the poet Dante Alighieri and the humanist and theologian St Maximus the Greek. The complexity of the author’s creative process as well as of the historical moment of the period when (s)he lived is discussed as well as the problem of the valuation of the Medieval and Renaissance period. The presented authors are valid examples for the beginning of the new era of thinking in the history and culture of mankind. The article also concerns visualizing the relationship between the creative process and the biographical data. A triangle structure is proposed within which the basic idea of the authors’ will, and their artistic vision could be expressed.

#### **Mapping biographies in a Relational Database. Biographies of Luxembourgish soldiers in the Second World War**

This paper by Nina Janz presents project WARLUX pertaining to the Nazi occupation in Luxembourg. The focus is on the history of Luxembourgers who were recruited and conscripted into German services under the Nazi rule during the Second World War. The lives of these men, women, and families are uncovered by representing different personal testimonies and individual war experiences. A relational database is used to represent these war experiences, and challenges encountered during the research project are discussed.

## **Creating and Using Biographical Dictionaries for Digital Humanities Based on Linked Data: A Survey of Web Services in Use in Finland**

In this paper, Eero Hyvönen overviews work on creating Linked Open Data (LOD) ontology and data services and applications for publishing and using biographical collections on the Semantic Web. A series of six LOD services and related biographical portals in use in Finland are presented, based on biographies and person registries of historical people. These systems aim to support both Digital Humanities researchers and the public interested in studying biographical data.

## **Biographical Research and Digital Mapping**

Paul Longley Arthur and Isabel Smith present a mid-project report on an Australian project that focuses on the development of digital tools to map biographical data relating to the history and legacies of British slavery in Australia from the 1830s onward. This data-intensive project is tracing the movement of capital, people, and culture from slave-owning Britain to Western Australia. The paper presents the project setting and tools for visualizing the project's data on maps, and then discusses questions and tensions that have emerged during the research.

## **Annotation of Named Entities in Medieval and Early Modern Epigraphic Texts**

This paper by Gregor Pobežin refines the annotation scheme proposed by Álvarez-Mellado et al. for the annotation of named entities and adapts it to the needs of medieval and early modern epigraphy. As case study, the MEMIS corpus is used, which brings together medieval and early modern inscriptions from the area of present-day Slovenia. Digital humanities tools and protocols provide access and process elements of historical evidence on epigraphic monuments as documents where Named Entity Recognition (NER) is of paramount importance for the extraction of biographical, prosopographical, and other data.

## **Biographical and Prosopographical Analyses of Finnish Academic People 1640–1899 Based on Linked Open Data**

Petri Leskinen Eero Hyvönen present in this article prosopographical data analyses using the Finnish AcademySampo linked open data service and portal. The primary data covers a significant part of the Finnish university history based on the student registries in 1640–1852 and 1853–1899 of the University of Helsinki. The article analyses networks connecting the students as well as their relatives mentioned in the biographical data. Correlation matrices based on vocations of the students and their parents are presented as well as quantitative analyses and visualizations of the family lines of students, based on automatically created family trees of the students and their parents.

## **Studying Occupations and Social Measures of Perished Soldiers in WarSampo Linked Open Data**

The WarSampo Knowledge Graph contains data about Finland in the Second World War, including metadata about some 100 000 soldiers. This register contains occupational labels, which have been manually harmonized into an ontology and linked to occupational classifications such as



HISCO. This paper by Mikko Koho and Eero Hyvönen gives an overview of the harmonized occupation ontology and provides an outlook of how the ontology, the occupation classifications, and related social measures can be used for prosopographical studies to provide new insights into the events of the war or of the Finnish society.

### **Traveling with Albrecht Dürer – A Case Study for Uncertainty-Aware Biography Visualization**

This paper by Florian Windhager, Eva Mayr, Johannes Liem, Jakob Kusnick, Stefan Jänicke, and Anja Grebe provides a biographical visualization case study with a focus on Albrecht Dürer. The idea is to showcase novel strategies to communicate relevant stations of his life and works in an integrated fashion, including uncertainties and biographical gaps. Both the overall biographical trajectory of Dürer and the episode “Journey to the Netherlands” (1520–1521) are considered and visualized based on a space-time cube perspective. The paper discusses how the often-heard charge of an inherent “positivist bias” of data visualizations could be inverted and utilized for the explication of interpretive ambiguity and plurality.

### **Acknowledgements**

BD2022 conference was supported by the EU H2020 research and innovation action InTaVia: In/Tangible European Heritage (project ID 101004825) funded by the European Commission. We warmly thank the participants and authors of the BD2022 conference for their contributions and the members of the organizing and program committees for their work in running the conference and reviewing the submissions.

### **References**

- [1] Serge ter Braake, Antske Fokkens, Ronald Sluijter, Thierry Declerk, and Eveline Wandl-Vogt (eds). 2015. BD2015 Biographical Data in a Digital World 2015. Proceedings of the First Conference on Biographical Data in a Digital World 2015. Amsterdam, the Netherlands, April 9. CEUR Workshop Proceedings, Vol. 1399. <https://ceur-ws.org/Vol-1399/>
- [2] Antske Fokkens, Serge ter Braake, Ronald Sluijter, Paul Arthur, and Eveline Wandl-Vogt (eds). 2017. BD2017 Biographical Data in a Digital World 2017. Proceedings of the Second Conference on Biographical Data in a Digital World 2017. Linz, Austria, November 6–7. CEUR Workshop Proceedings, Vol. 2119. <https://ceur-ws.org/Vol-2119/>
- [3] Angel Daza, Antske Fokkens, Petya Osenova, Kiril Simov, Alexander Popov, Paul Arthur, Thierry Declerk, Ronald Sluijter, Serge ter Braake, and Eveline Wandl-Vogt (eds). 2019. BD2019 Biographical Data in a Digital World 2019. Proceedings of the Third Conference on Biographical Data in a Digital World 2019. Varna, Bulgaria, September 5–6. CEUR Workshop Proceedings, Vol. 3152. <https://ceur-ws.org/Vol-3152/>

# The Promise of the Finding Aid: A Critical Approach to Finding Biographical (Linked) Data in the Archive

Rachel Pierce<sup>1</sup>, PhD (Research coordinator)

<sup>1</sup>*KvinnSam, Humanities Library, University of Gothenburg, Renströmsgatan 4, 41255 Gothenburg, Sweden*

## Abstract

Projects focused on biographical archival data have traditionally placed community-wide institutional lists at the center – often church records for earlier historical periods and census information for years dominated by the modern state. Yet these lengthy lists of individuals lack contextual material, which is often necessary for definitively identifying individuals and building argument-driven historical research projects. This article will address the digital finding aid as a potential source of biographical data grounded in the contextual information so central to humanities research, using a set of finding aids of a selection of personal and organizational collections from the Schlesinger Library at Harvard University and the Library of Congress as examples. While resurrecting the evidentiary lives of everyone is impossible, embedding digital technologies in finding aids may offer a more equitable and less hierarchical method of assembling biographical data, as well as new doors into physical archives. However, a critical perspective is key; the creation of linked data will create new hierarchies and silences within the world of archives.

## Keywords

digitization, linked data, finding aids, representational equity, archives

## 1. Introduction

As feminist historian Joan Wallach-Scott has noted, “historians make death a minor episode, something that is transitory rather than final” 1, 144. And yet it is difficult – impossible even – to resurrect everyone, though historians have tried. Digitization is merely the latest method. Recently, projects have focused on the digitization and mining of censuses, church records, and judicial documents, which have long been considered the site of much bottom-up historical research. While implementation of linked data has received a fair amount of attention in the drive to save past actors from ignominy, scholars have discussed linked data almost in the context of digitized and born-digital primary source collections, although biographical dictionaries have also received substantial attention [2][3]. Strangely, the choice of text for linked data implementation has rarely been problematized.

Given the size of archival collections and tendencies within selection for digitization, there are considerable problems inherent in assuming that the material that has been digitized is

---


*DH2022 | Digital Humanities 2022; Responding to Asian Diversity. 25-29 July 2022, Toshi Center Hotel, Tokyo, Japan and Fully Online (Zoom)*

\*Corresponding author.

†These authors contributed equally.

✉ [rachel.pierce@ub.gu.se](mailto:rachel.pierce@ub.gu.se) (R. Pierce)

ORCID [0000-0003-3480-1474](https://orcid.org/0000-0003-3480-1474) (R. Pierce)

 © 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

[https://doi.org/10.3986/9789610508120\\_01](https://doi.org/10.3986/9789610508120_01)

representative of physical archival holdings [4]. And in reality, biographical dictionaries have many of the same representational problems, privileging a definition of individual importance grounded in nineteenth-century historiographical traditions that are out-of-tune with current historical work on grassroots publics, movement politics, and the everyday [5][6]. This article will argue that implementing linked data in finding aids, which have long provided routes into the archive for researchers, is potentially a more effective, efficient, and historically useful way of facilitating historical research. These documents are structured well for digital mining but also place individuals within correspondence networks, family structures, and work hierarchies while providing information on concrete context such as dates and places. Further, and most importantly, finding aids do not reflect history as it has been written, but rather open up linked data into the history that could be written. This article will address the structure of finding aids, which offer good contextualized biographical information while exploring the potential difficulties of differences between personal and organizational archives.

Once properly digitized, technical opportunities for analyzing the history embedded in these documents are numerous. But linked data in particular is a relatively unexploited and flexible tool that would make embedded biographical information computationally digestible without losing the contextual structure that is so central to humanities research. As noted by digital humanists, the field has struggled to create solutions that preserve context while producing machine-digestible humanities data [7]. But digital technologies are not simply well-suited to enhancing finding aids. They also offer an interesting set of solutions to problems of representation inherent within archival collections. With the addition of linked keywords and item-level description, potentially crowdsourced from researchers, this approach would also help to address historical imbalances in archival description. Scholars have pointed out the possibilities of linked data for depicting relationships between individuals, groups, and organizations without ordering these relationships hierarchically [8]. Linked data within finding aids is one potential practical solution that is worth exploration.

In order to analyze the possibilities integrating biographical linked data into finding aids, this article will look in depth at finding aids from both women and gender-centered archives and national archives. Materials from the Schlesinger, a library and archive devoted specifically to women's and gender collections, are finding aids for the collections of (1) the turn-of-the-century activist, economist, and author Charlotte Perkins Gilman and (2) the Boston Women's Health Book Collective, the group that authored the seminal feminist text *Our Bodies, Ourselves*. From the Library of Congress, finding aids for the papers of (3) former feminist congressperson Patsy T. Mink, of Hawaii, and (4) jazz drummer Max Roach. These materials span a century and a half and cover individuals and groups with diverse biographical and contextual backgrounds. The finding aids are also well-developed while displaying a structural diversity that will be useful in discussing the possibilities and limits of linked data within this genre.

## **2. Digitizing Finding Aids: Identifying Current Challenges**

What is a finding aid? Chung offers a loose description of the document as “any type of descriptive tool that provides intellectual access of the materials held in an archival or manuscript repository” 9, 147. This definition is necessarily broad, given that the descriptive practices



employed by archivists varied greatly from institution to institution and type of collection to type of collection for centuries before standardization began to take hold in the second half of the twentieth century [10]. Now, many countries have guides for archival description that prioritize and seek to preserve provenance and original order through the both the hierarchical organization and description of archival collections. Finding aids also often include information about the creation and historical use of collections. These are documents designed to both describe collections so that their contents are understood within a historical system of practices and provide a multiplicity of entry points into archival collections for the roving researcher. These two goals are not always synergetic, as multiple scholars have noted [11][12].

Archives began digitizing finding aids early in the history of the Internet, in order to make their collections more findable and provide information to researchers so that they could order sections of archival collections in advance of physical visits. This was a revolutionary shift in archival research practices that received little attention at the time [12]. The shift to digital (digitized) finding aids capped a broader shift towards standardized descriptive practices, making these now systematically structured documents open to the public. The implementation of encoded archival description (EAD) has dominated the transformation of finding aids to digital, computer-searchable documents. This description system is an international standard for encoding finding aids using XML, run by the Society of American Archivists and the Library of Congress. More recently, the RDF-based Encoded Archival Context for Corporate Bodies, Persons, and Families (EAC-CPF) was developed to encode archival entities and enable the implementation of linked data.

Since its introduction at the end of the 1990s, the implementation of EAD has proceeded without much debate. Yet concerns about the usability of the finding aid have only increased, as digitization has shifted the definition of archival users from a physical group of presumed historians with a sprinkle of genealogists to a broader public. Simultaneously, web access to finding aids was not accompanied by digital access to archivists themselves – traditional repositories of information about finding aid navigation in general and individual collection navigation in particular. As Cox has argued, “Perhaps the real business of postmodern archival enterprise ought to be re-evaluating just what finding aids represent over time, studying them as a documentary source reflecting attitudes and practices of the archival community at various times” 10, 9. What is clear is that while digitization of finding aids made whole archives more discoverable, it did not necessarily make the search for particular topics or individuals within the Archive – across collection and institutional boundaries – more findable or more usable<sup>1</sup>. As Garmendia and Retter note, this is a far more complicated process than it might seem at first [14].

This situation has particular repercussions for scholars seeking materials about and by historically underrepresented groups. Feminist scholars have long argued that the structure of both archives and finding aids obscure or erase the presence and importance of women, people of color, the lower and working classes, and other historically disempowered peoples [15][16]. In the context of digital finding aids, these scholars demand two things. The first demand is the

---

<sup>1</sup>Here, I follow Carolyn Steedman’s practice of differentiating between individual collections, archival institutions, archivist practices, and the Archive of the world’s current and future cultural historical matter [13]. The last of these is as much an idea as a(n ever-changing) physical reality.

preservation of historical hierarchies that not only lies at the heart of archival order but also provides meaningful contextual information about the hierarchies that governed the creation of these collections. The second demand is the creation of tools that allow archival researchers to circumvent these hierarchies in searching for historically and archivally marginal groups and individuals. It is becoming clear that current trends within the digitization of finding aids fulfill the first goal while failing on the second.

So why finding aids, given these barriers? Until now, most digital humanities scholars have focused on implementing linked data for digitized cultural heritage materials or national biographies. Issues with digitized cultural heritage materials are clear – these collections of documents represent a scattershot array of archival materials, whose representativeness is both highly doubtful and impossible to estimate. Biographical dictionaries seem, upon first glance, to go beyond the specificity of archival documents in some of the same ways as finding aids. However, as Warren (2018), points out, the biographical dictionary exists somewhere in between the categories of primary and secondary source material, as a time capsule of the history that has been written. The relatively young sub-fields of women's, gender, indigenous, Black, queer, and social movement history (among many others) take up very little space in these texts, even as materials from these histories have flooded into preexisting archival institutions and prompted the widespread founding of new, activist archives that prioritize [17]. Meanwhile, the genre of biography continues to suffer from a tendency to prioritize the historically powerful [18]. If archives move slowly, biographical research moves even slower. And some scholars argue that the genre is structured to prioritize in this way, that even future historiography will not undermine the individualism and elitism of biography [19].

The digitization of finding aids – including for these new, introduces new possibilities for integrating digital tools into these documents. Yet development here has been slow and ignores the criticisms of archival scholars whose work focuses on the importance of understanding and making transparent the power dynamics that the archive reflects and institutionalizes. As previously noted, digital humanists have investigated the possibilities of linked data and archives but have focused largely on digitized and born digital primary source material, that is, material from the archive rather than material about the archive. This emphasis has resulted in archivist work on “knitting” relationships between finding aids and digitized documents to “improve use and discoverability” [20], though some projects such as LOCAH translated biographical EAD description to linked data (Niu 2016). Similarly, the arguments and projects presented at the 2021 workshop on linked data in archives focused on using linked data to reproduce archival information in linked data form, with little attention to archival researcher practices and inbuilt archival hierarchies (Lopes et al. 2021). As a result, this work does not expose, question, or disrupt the hierarchical structure of the archive.

The application of linked data within finding aids has not been discussed much, and as this article, will show, linked data is far from a panacea for the inequities of the archive or the internet. Critical approaches within the digital humanities raise concerns that mere application of digital tools does not erase or rectify pre-existing hierarchies and can, in fact, deepen these problems (Risam 2019). Humanists must, as multiple scholars argue, be attuned to the infrastructures that linked data sets, visualizations, and interfaces build on – these infrastructures are also objects in need of (humanistic) analysis [21][22]. Yet digital humanists interested in queering the field have focused on metadata as a potential site of creativity for a reason [8]. There are

opportunities here, if we are willing to exploit them in an imaginative and critical way. This article will present the possibilities, built-in biases, and potential pitfalls of linked data for finding aids, as a starting point for scholarly discussion and potential future implementation.

### **3. Case Studies from the Schlesinger and the Library of Congress**

Four finding aids from two different institutions have been chosen for analysis. The Arthur and Elizabeth Schlesinger Library for Women's History is located at Harvard University, and two collections have been chosen for investigation, in order to look at the suitability of individual/familial papers versus organizational papers. The first case study is the finding aid for the papers of 19th century women's rights campaigner, economist, and author Charlotte Perkins Gilman [23]. The second finding aid covers the collections for the Boston Women's Health Book Collective, the 1970s feminist activist group that wrote the seminal reproductive rights text *Our Bodies, Ourselves* [24]. The last two finding aids are from the Library of Congress. The first is the finding aid of the papers of former Hawaii Rep. Patsy Mink, a legislator well-known for her work on women's and Asian American rights [25], and the second is the under development finding aid for bebop drummer Max Roach's papers [26].

These are well-developed finding aids that are far from representative. It takes extraordinary institutional resources to create this kind of document, underlining the resources that would be needed to create linked data while simultaneously highlighting the resources that would be saved if linked data was created as a part of finding aid creation. The need for crowdsourcing is acute here, and indeed crowdsourcing for linked data can provide not only a symbiotic relationship between archival scholars and archivists but a concrete introduction to digital tools for these scholars, who have long been seen as both ignorant of and resistant to digital tools [27].

The particular finding aids have also been chosen, somewhat ironically, to center a cross-section of underrepresented groups. But these finding aids should not be taken as representative. The scope of archival collections and the biographical histories that can be written based in these archives are neither representative of particular groups nor of the Archive in its voluminous, fractured, and fluid entirety. Instead, this article will use the four chosen case studies to illustrate more general possibilities and problems with using finding aids to create biographical linked data, with particular attention to issues of representation, classification, hierarchy, and binaries highlighted by feminist archival scholars.

#### **3.1. Case Study 1: Papers of Charlotte Perkins Gilman, 1846-1961**

A finding aid is a descriptive text and a map. These documents provide historical context for collections – who owned the materials over time, how and when they arrive at their final repositories – and the background information for the person(s) or organizations around whom/which the collections have been constructed. These introductory texts are minable for information that can become linked data. The Perkins Gilman finding aid from the Schlesinger Library at Harvard University is extensive and detailed, illustrating the potential for linked data implementation in well-developed finding aids.



This finding aid demonstrates that such documents – specifically those for individual and familial collections – are particularly well-suited for finding good, contextualized biographical linked data. The structure of the finding aid eases the process of finding this information. There are several places to look for such information, but in particular, the biographical note, the index, and the subject list all provide information on relationships between Perkins Gilman and a wide array of familial, work, and friendship relationships. The biographical note provides an abundance of potential well-contextualized linked data, including influences (such as Lester F. Ward and Sigmund Freud), friends (Grace Ellery Channing, for instance), and genealogical names (including her husband, artist Charles Walter Stetson) that can be linked to dates, places, and specific locations in the physical archival collection. The index and subject lists provide a total of 229 unique names that link these relationships to physical artifacts in the archive that provide evidentiary bases for these relationships. The finding aid structure also provides information on correspondence relationships, namely in Series II and III, where box-level description reveals familial, friend, and work relationships with substantial physical documentation. The Schlesinger has digitized much of this letter-writing, meaning that these relationships can also be linked to digitized material.

Perkins Gilman's biographical trajectory can also be traced via her position within a number of national and international women's rights and political organizations. These organizational names, which also appear in the index and subject list, provide important contextual information, placing her within a set of social movement networks that were the basis for mainstream women's rights politics of the 19th and early 20th centuries. Individuals are the basis of such organizations, as will be shown with the next section, and the implementation of linked data for charting organizational leadership and membership over time and space would provide important information on how social movements evolved and interacted with various people and institutions at specific points in time and in specific places.

Discussions of linked data in archives have until now not dealt with issues of hierarchy or representation, much less what is left out as these technologies are utilized. The assumption has been that linked data replaces hierarchies with graph-based relational representations that undermine archival hierarchical organization. But what this particular finding aid most reveals are the repercussions of selection in shaping the kinds of representations that are enabled with linked data implementation. This finding aid offers a fairly equal number of women and men, while class and race (amongst other kinds of) diversity is absent. The collection is populated with middle-class White figures. Those applying linked data in archival contexts should consider these qualities of each archival collection as well as what silences and new hierarchies are created as linked data is fashioned.

### **3.2. Case Study 2: Records of the Boston Women's Health Book Collective, 1905-2003 (inclusive), 1972-1997 (bulk)**

Looking at the finding aid for the Boston Women's Health Book Collective provides a glimpse of how the biographical make-up of social movements might be expressed, as well as illustrating the difficulties such a finding aid poses for linked data implementation. This finding aid has a subject list but no name index, and the subject list contains only two names. However, both the "Scope and Contents" note and the sub-series-level description provides a wealth of names

(11), particularly relating to the founders of the collective, both of whom have contributed a full series of materials (Series V for Judy Norsigian and Series VI for Esther Rome) board work, and the outward-facing members who engaged in communication with the general public (see Series XXIV). The “History” section of the finding aid includes a list of founding members (14 women in total) noting that most of these women “remained together for more than twenty years, sharing their personal and professional lives” [24].

Understanding that individuals and collectives are interactive and interdependent historical actors is an important consideration. A major question with biographical linked data is how the concept “biography” is defined. Recent historical biographical research has pushed back against the traditions of biography, resulting in very interesting work on couples, families, and groups that challenges the individualism that has long undergirded the genre [18]. The finding aid and collections for the Boston Women’s Health Book Collective challenge us to think about what we lose when we prioritize individual-oriented biographical data to the exclusion of group materials, particularly those related to social movements. It is here that less well-known individuals contributed to broad societal change through coordinated collective action. Being able to map the organizational relationships that sustained the activist-political careers of individual women over time would be an immense contribution to historical scholarship.

Actor-network theory treats people, events, objects, and sources as equal, providing a way of constructing biographical linked data that represents organizations and sub-units within organizations (for example, office staff or boards) alongside individuals [28]. This theory parallels SPARQL endpoints, which facilitate search for a multiplicity of objects, places, concepts, and other entities 29, 95. Such an approach to linked data would be highly appropriate, especially in the case of organizational archival collections. In the case of the Boston Women’s Health Book Collective, relationships between different sectors of the collective (the board, marketing staff, grant writers, etc.) could be articulated. Further, relationships between the Collective and other organizations and projects were so central that they warranted their own series (Series IV). The founders of the Collective also worked in and with numerous other local and national organizations, functioning as links between various kinds of women’s health work in and outside of government and across political, healthcare, and social movement environments. Representing these organizational synergies in linked data would be central to a good representation of the multifaceted networks that sustained the work of both the Collective and its individual members.

The finding aid for the Boston Women’s Health Book Collective also illustrates the need to exploit physical archival labor in order to develop and implement new digital tools within the finding aid. Historians increasingly take photographs, create spreadsheets, and do other digital work as part of their archival research [27]. These digital skills are developed to contend with the administration of large quantities of information and documents, and their development offers an opening for synergy between physical archival research and the digitization of finding aids. But other options exist, particularly for archives located at universities, where students, professors, archivists, and digital resources staff can collaborate on the construction of linked data. Experiments of this kind will help to equip the next generation of archivists and researchers with useful digital skills [30].

### 3.3. Case Study 3: Patsy T. Mink Papers

Patsy Mink's collection is a lesson in humility for archival linked data enthusiasts. Mink represented Hawaii in the United States House of Representatives for 26 years, as well as cultivating a career in local and state politics. Her collection demonstrates the problems with large archives. Resources and archival traditions mean that finding aids describing large collections can be somewhat more concise and often fail to describe materials at the item level. That 24 individuals worked on constructing the finding aid alone is a sign of the kind of work that implementation of linked data would require [25]. These collections also are more likely to have vague titling practices that, for instance, denote policy areas, correspondence for particular years or specific document formats like photographic collections. This is a set of problems that gets more dire the closer one gets to the state, an irony that illustrates the flipped dynamics at work with linked data creation. Here, the congressperson is an actor within a state apparatus, literally a representative of collective interests. This situation produces an elision of actors in order to highlight the representative's interaction with and within a governmental organ.

But while working on this kind of linked data is far from easy, the result would provide important information on the nature of lawmaking, which is built on networking and personal relationships. Mink commanded a formidable staff of secretaries, lawyers, writers, and researchers on her personal and committee staffs. These individuals had relationships with staffs in other offices. These relationships are the groundwork upon which lawmaking is built, and yet these names are not elaborated in the finding aids for the collections of congresspersons, who were supposed to own the work of their offices. This is information that would need to be crowdsourced, and it is voluminous. There are 535 members of Congress every session and thousands of staffers who have passed through the offices of Capitol Hill.

Equally technically challenging but ethically fraught as well are two components of the Mink finding aid: constituent materials and genealogical information. If we consider provenance from a community archives perspective, these individuals should be considered co-creators of the archives that have heretofore been classified as congressional. And yet this is an authorship that they did not foresee or intend. We cannot assume that constituents wrote to their congressional representatives thinking that they were, in doing so, contributing to the historical record, although some letter-writers may have thought this way. Nor can we assume that these individuals would want this information published for the broader public, given that the historical record of archival collections has until recently been understood to be relatively closed and inaccessible. The genealogical information in Mink's collection presents further issues with explicitly restricted material, a not-uncommon occurrence, especially with more recent collections.

Linked data is, as Terhi Nurmikko-Fuller notes, an "information publication paradigm," and it is important to understand that this technology requires the publication of the personal data of individuals who considered their communications somewhere between public and private [31]. While scholars often position linked open data as the best-case scenario, it is important to remember that openness is not always the choice of those whose information is in question, and publication in any form may be against the wishes of archival contributors. Recent scholarship on community archives provides an important corrective to traditional archivist practices that simplified and individualized ownership of archival collections, and this scholarship has

repercussions for the implementation of linked data [17].

It is important to take these concerns into account, especially because it is individuals from historically disempowered groups who tend to be wary of depositing their papers with archival institutions and often place substantial restrictions on their papers when they are donated [32]. The implementation of linked data should always take place in line with the CARE data principles: Collective benefit, Responsibility, Authority to control, and Ethics. Though these principles were developed to protect indigenous community rights amidst digitization drives, the ubiquity of rights and access restrictions in the context of archives means that this issue must be met head-on regardless of where one chooses to start with linked data creation. The resolving of publishing rights for linked data will not be an easy task, as the FAIR principles of making data Findable, Accessible, Interoperable, and Reusable can end up conflicting with CARE [33]. Much work with data rights education will need to be done to ensure that scholars create linked data that does not infringe on the rights of individuals or leave out large swaths of the historical record.

### 3.4. Case Study 4: Max Roach Papers and Linked Jazz

The Max Roach Papers are currently being processed by the LoC, providing a look at how a less well-developed finding aid might still provide good quality linked data that enhances current linked data projects based in primary documents [26]. Roach was a renowned jazz drummer whose work spanned much of the second half of the 20th century. The Linked Jazz project began by mining oral histories for information about relationships between jazz musicians who, as Roach's own career amply demonstrates, were highly collaborative, taking inspiration from multiple sources and playing and recording in numerous configurations - for instance, Martin Luther King, Jr. is now included. Roach does in fact appear in the project, though he is currently somewhat peripheral within the network.<sup>2</sup>

What we see when we look at the Roach finding aid is new influences and collaborators that do not conform to the themed bounds of projects like Linked Jazz. For instance, there is currently no Linked Jazz triple elaborating the relationship between Roach and Nina Simone, a relationship important enough to show up in the shorter contextual summary of this in-process biographical section. But we also see a much wider influence and collaboration net that reflects time- and place-specific Black Power cultural creativity. Amiri Baraka (who wrote about jazz but did not play) and Maya Angelou (best known as an author, but also a singer who released one studio album in 1957) feature prominently, largely because of their similar sociocultural-political positioning. Baraka is absent from Linked Jazz, while Angelou is linked, tellingly, only to other women, including Roach's wife and collaborator Abbey Lincoln.<sup>3</sup>

Thematic projects – which include biographical dictionaries – tend to create boundaries and engage in selection and classification processes that obscure other facets of or ways of understanding the individuals featured in these projects. What finding aids give us is a set of relationships reflected in the stuff and structure of archival collections, which bridge divides between the personal and the political, the private and the public in ways that themed projects often (accidentally or not) reify. As Carolyn Steedman asserts, “The Archive is made from

---

<sup>2</sup><https://linkedjazz.org/network/> (accessed 15 Aug 2022)

<sup>3</sup><https://linkedjazz.org/api/relationships/all/nt> (accessed 15 Aug 2022)

selected and consciously chosen documentation from the past and also from the mad fragmentations that no one intended to preserve and that just ended up there” 13, 65. This mishmash of intended and unintended, consciously kept and simply forgotten is not simply a goal of feminist research. It is also a key driver of historical research 1, 145.

At the same time, the Linked Jazz project illustrates how biographical relationships can be translated into linked triples via a combination of archival work, automation, and crowdsourcing. Linked Jazz began with the use of a transcript analyzer to identify personal names, which were then related to one another using the simple `rel:knowsOf` relationship, which crowdsourcing then refined [34]. Finding aids have, as noted, very regular structures, making them ideal documents for automated analysis. This approach allows for the identification of names and relationships that can (or must, as the case may be) later be nuanced as researchers or archivists use the collection. Given that finding aids are the door to archives, it is somewhat strange that co-creation of finding aids has not been proposed as a key to making the archive more accessible. Current archival science scholarship centers the archivist as non-neutral author [35], but future finding aid construction will need to account for multiple authors, including machines.

The project also demonstrates how linked data can allow users to center historically under-represented groups – if project creators have thought about and planned for this representation. The addition of gendered attributes came late in project development and was conceived because of the underrepresentation of women in the initial dataset. Attention to the smaller datasets and the less well-represented or seemingly invisible individuals and groups is important, not simply for producing linked data that is flexible and useful for lots of researchers but because such disparities and silences are historically meaningful and often can provide us with new perspectives rather than reproducing inequities that scholars already know exist. In particular, as the Linked Jazz team notes, available linked data tools are not particularly well-representative of the fluidity of sex and gender – an important area for future innovation 36, 922.

What an analysis of the Roach finding aid adds is a sense of how different the linked data possibilities are when one starts with an archival collection rather than an oral history transcript. Finding aids have the potential benefit of stretching across entire lives in a manner different from oral histories. They often center familial and work relationships quite differently, since finding aids are descriptive rather than narrative devices. Their subject indexes also – as Roach’s does – place individuals, groups (such as bands and recording constellations), and themes (like “percussion music” or “race relations in mass media”). The Roach finding aid also highlights the potential for cutting across documentary format, as materials belonging to these categories are arrayed across photographs, musical scores, interview transcripts, and other document types, which might help to rectify the historical privileging of text documents as the basis for historical analysis [37]. These categories overlap with contributor names and the entities that could be pulled from what is currently titled the “Summary” but will eventually end up in the “Scope and Content” note. It is this messy array that makes the collection “a dynamic portrayal of the personalities and cultural forces behind jazz in the United States and abroad in the latter half of the twentieth century” [26] It is also this mess that potentially makes linked data such a useful tool.



## 4. 8. The Challenges and Potential Gains of Finding Aid Linked Biographical Data

Christine Borgman has cogently argued that “data have no value or meaning in isolation. . . They exist within a knowledge infrastructure—an ecology of people, practices, technologies, institutions, material objects, and relationships” 38, 4 Any attempt to create biographical linked data should not forget that context is everything, and biographical data culled from archival collections has two important contexts: specific archives AND the historical context of the material itself. One of the potential promises of extraction of biographical linked data from finding aids is the preservation of both of these contexts. This approach to finding aids would provide a multitude of remixable routes into the physical and digital archives – an unfulfilled goal of digitizing archivists since the invention of the internet [12]. Crucially, linked data has the potential to break down the barriers between individual institutions that prevent researchers from seeing biographical networks that span archival collections and have the possibility to challenge rather than reify current and past historiography.

Perhaps most importantly, archival finding aids open up the array of primary materials in the Archive, facilitating future research in ways that bounded projects and biographical dictionaries cannot. There are vast oceans of material in archival collections that have not yet been researched. The limitations of biographical dictionaries for future historical research are clear; as historiographical, national documents, these dictionaries provide a window into what has been done and are not easily analyzed for what is not there but could be there. That much of history is simply inaccessible to us is a fact we must simply accept as the product of “archival silences,” as Michael Moss and David Thomas put it [39]. Current linked data implementations are simply not designed to increase the routes into the physical archives, across institutional and national boundaries, as well as across well-trafficked and under-used formats, persons, and groups. If the goal is to spur new research questions and topics within the humanities and particularly within the discipline of history, finding aids are a highly appropriate starting place.

The technical or administrative aspects of this kind of work are far from easy. The creation of finding aids is undergoing revision as these documents and archival material itself moves online. Teaching archivists simple markup languages such as Text Encoding Initiative (TEI) would provide the groundwork for implementation of initial linked data structures within finding aids themselves. This kind of work would necessarily span institutional boundaries and reject narrower project-based focuses. Here, archivists would need to think locally, in line with postcolonial digital humanities practices that foreground “the particular over the universal,” which via crowdsourcing and material selection can “offer [historically underrepresented groups] a more expansive humanities that takes advantage of the technical means of digital knowledge production to populate the digital cultural record with their own stories” 40, 9.

Defining what does and does not count as biographical data is key. Scholars have debated the relative merits of traditional biography, noting that the genre has certain conservative, “bourgeois” tendencies that deserve attention if not redress [19]. Given the Archive’s traditional status as a repository for public-minded materials and the raced, classed, and sexed history of our definitions of “the public,” it is difficult to reconstruct the biographies of individuals and groups who fell outside of this normative definition of the historically important [18]. An

analysis of the potential for linked data in finding aids demonstrates that there is material there that disrupts biography's tendency to use individuals with more substantial archives to represent groups. The structure of linked data embeds these individuals in sociopolitical networks, undermining the notion that great individuals – the ones with the archives – make history, while everyone else is at the mercy of historical forces. Treating organizations, spaces of interaction and invention, and other nonhuman entities as biographical would also do much to undermine this individualization of history, in no small part by providing representation to social movements.

However, the implementation of linked data cannot be seen as a panacea for the problems of representation and equitable engagement with archival materials. As the directors of the transnational digital African archive Aluka notes, "As it relates specifically to the domain of history as a discipline, the digitization process places contested archives in a cyberspace that is highly commodified" 41, 73. This commodification has numerous repercussions, primary amongst them a tendency to use digital tools to "sell" archival institutions, an imperative that results in the prioritization of flagship materials rather than the unexplored corners of collections. Good, transparent selection practices – which finding aids are chosen for processing and which sections are translated into linked (open) data – must be conducted together with affected groups and individuals. Processual openness is critical, if finding aid-based biographical linked data is to adhere to the CARE principles and serve as a means of making visible individuals and groups who relegated to the archival margins via traditional archival practices. But the fact of linked data as a salable tool inside and outside the academy will affect who and why this kind of work is considered valuable and useful.

Nor should we forget that archives have always served an exclusionary function. Death is only a "minor function" for the few. One of the most important issues for current digital humanities researchers is thinking about and finding ways to represent individuals, groups, and topics who/that are either underdocumented or evidence themselves as silences in the Archive. The goal is to research these hierarchies, but linked data has the potential to make such disjuncture and inequity less visible, as datasets tend to seem complete and neutral 42, 99. Linked data should be used to subvert but also make visible these exclusionary tendencies. New digital tools should be used to help us see some of the representational hierarchies built into the physical archives, while also allowing for a way of remixing archives to both center previously peripheral individuals and groups and make the silences and edges of archival collections more visible. Archival scholars and archivists themselves will then need to think about how to understand and interpret these silences and edges. Here, the humanities' emphasis on the need to account for doubt, multiplicity, and subjectivity is of critical value.

## **5. Conclusion**

Linked data has attracted a great deal of attention from scholars and archivists who are interested in opening up archival collections to the general public, making materials more findable for researchers, and challenging institutional silos. There are other interesting gains identified here, amongst them a questioning of distinctions between individuals and collectivities, as well as a contextualization of biographies in the materials that undergird our understanding of historical



actors, as well as the provenance of these materials. Selection is key, as finding aids provide biographical networks from a particular person or organization's perspective. Last, linked data offers archivists an opportunity to test some archival science theories about centers, peripheries, and local knowledge (see for example [43], since it is potentially more interesting and fruitful to begin with smaller archival collections from individuals and collectives that have until now been viewed as less important. The implementation of linked data multiplies the doors into physical and digital archival collections, potentially changing the process of archival research in profound ways.

However, as with many technological "solutions," scholarship has focused on problems of implementation rather than the opportunities and limitations of these tools, as well as what such tools will mean from a user perspective. Digitization of various aspects of the Archive does not free information or materials from the hierarchies of the archive, but instead rearranges and subverts some of the power dynamics of archival collections while bolstering other inequities. While using finding aids as a jumping off point for biographical linked data creation is a potentially more equitable solution than starting with the "flagship" materials that have been prioritized for digitization, this method is not an end in itself. But to users, digital representations of the Archive can seem encompassing, even if digitized collections are incomplete or contain OCR-related problems, precisely because digitization is a process of decontextualization [44]. A feminist approach to issues of ethics and representation can help archivists to structure the deployment of linked data technologies in ways that undermine archival hierarchies while also illustrating the power dynamics and silences that are always at play in cultural heritage, digitized or not.

## Acknowledgments

Thanks to the workshop organizers for creating the space for presentation and discussion of these ideas and to the reviewers for cogent and useful comments. Thanks also to Martin and little Oona, my buddies.

## References

- [1] J. Wallach-Scott, *The Fantasy of Feminist History*, Duke University Press, Durham, NC, 2011.
- [2] A. Hawkins, Archives, linked data and the digital humanities: Increasing access to digitised and born-digital archives via the semantic web, *Archival Science* 22 (2022) 319–344. doi:10.1007/s10502-021-09381-0.
- [3] M. Tamper, P. Leskinen, E. Hyvönen, R. Valjus, K. Keravuori, Analyzing biography collections historiographically as linked data: Case national biography of finland, *Semantic Web* 14 (2022). doi:10.3233/SW-222887.
- [4] R. Pierce, Documents in the dynarchive: Questioning the total revolution of the digital archive, *Proceedings from the Document Academy* 8 (2021). doi:10.35492/docam/8/2/4.
- [5] A. Fokkens, S. T. Braeke, N. Ockeloen, P. Vossen, S. Legêne, G. Schreiber, V. de Boer, *Europa*

- baut auf Biographien: Aspekte, Bausteine, Normen und Standards für eine europäische Biographik, new academic press, 2017, pp. 193–227. doi:10.1007/3-540-09237-4.
- [6] C. N. Warren, Historiography's two voices: Data infrastructure and history at scale in the oxford dictionary of national biography, *Journal of Cultural Analytics* 1 (2018) 1–31. doi:10.22148/16.028.
- [7] D. Oldman, *Information and Knowledge Organisation in Digital Humanities: Global Perspectives*, Routledge, New York, NY, 2016, pp. 131–153.
- [8] B. Ruberg, J. Boyd, J. Howe, *Bodies of Information: Intersectional Feminism and Digital Humanities*, University of Minnesota Press, Minneapolis, MN, 2018, pp. 108–127.
- [9] S. K. Chung, *Encyclopedia of library and information science*, 3 ed., Taylor and Francis, New York, NY, 2010, p. 147–153.
- [10] R. J. Cox, Revisiting the archival finding aid, *Journal of Archival Organization* 5 (2008) 5–23. doi:10.1080/15332740802153245.
- [11] L. Freund, E. G. Toms, Interacting with archival finding aids, *Journal of the Association for Information Science and Technology* 67 (2015) 994–1008. doi:10.1002/asi.23436.
- [12] C. B. Trace, A. Dillon, The evolution of the finding aid in the united states: From physical to digital document genre, *Archival Science* 12 (2012) 501–519. doi:10.1007/s10502-012-9190-5.
- [13] C. Steedman, *Dust: The Archive and Cultural History*, Rutgers University Press, New Brunswick, NJ, 2002.
- [14] J. Garmendia, A. Retter, Developing a pan-archival linked data catalogue, in: F. N. I. R. Carla Teixeira Lopes, Cristina Ribeiro, N. Freire (Eds.), *Proceedings of Linked Archives International Workshop 2021, 25th International Conference on Theory and Practice of Digital Libraries, 2021*, pp. 93–103.
- [15] A. Burton, *Dwelling in the Archive: Women Writing House, Home, and History in Late Colonial India*, Oxford University Press, Oxford, 2003.
- [16] A. Stoler, *Along the Archival Grain: Epistemic Anxieties and Colonial Common Sense*, Princeton University Press, Princeton, NJ, 2010.
- [17] J. Bastian, A. Flinn (Eds.), *Community Archives, Community Spaces: Heritage, Memory and Identity*, Facet Publishing, London, 2020.
- [18] N. Chaudhuri, S. Katz, M. E. Perry (Eds.), *Contesting Archives: Finding Women in the Sources*, University of Illinois Press, Urbana, IL, 2010.
- [19] S. Kracauer, *Biography in Theory: Key Texts with Commentaries*, de Gruyter, Berlin, 2017, pp. 107–111.
- [20] L. Woolcott, A. Payant, S. Skindeliën, Partnering for discoverability: Knitting archival finding aids to digitized material using a low tech digital content linking process, *code4Lib Journal* 34 (2016).
- [21] J. Drucker, *Graphesis: Visual Forms of Knowledge Production*, Harvard University Press, Cambridge, MA, 2014.
- [22] B. Mak, Archeology of a digitization, *Journal of the Association for Information Science and Technology* 65 (2014) 1515–1526.
- [23] E. Mosley, J. Gotwals, *Charlotte perkins gilman papers, 1846-1961, 1972*, updated 2010. URL: <https://hollisarchives.lib.harvard.edu/repositories/8/resources/4827>, finding aid.
- [24] K. Kraft, J. Carll, *Boston women's health book collective records, 1905-2003, 2005*. URL:

- <https://hollisarchives.lib.harvard.edu/repositories/8/resources/6841>, finding aid.
- [25] Patsy t. mink papers, 1883-2005, 2007. URL: <https://hdl.loc.gov/loc.mss/eadmss.ms010008>, finding aid.
- [26] Max roach papers, 1880-2012, 2016. URL: <https://hdl.loc.gov/loc.music/eadmus.mu016007>, finding aid.
- [27] I. Milligan, *History in the Age of Abundance?: How the Web Is Transforming Historical Research*, McGill-Queen's University Press, Montreal and Kingston, 2019.
- [28] G. Durepos, A. J. Mills, Actor-network theory, anti-history and critical organizational historiography, *Organization* 19 (2011). doi:10.1177/1350508411420196.
- [29] J. Niu, Linked data for archives, *Archivaria* 82 (2016) 83–110.
- [30] J. Needham, J. C. Haas, Collaboration adventures with primary sources: Exploring creative and digital outputs, *The Journal of Interactive Technology and Pedagogy* 14 (2019).
- [31] T. Nurmikko-Fuller, Teaching linked open data using bibliographic metadata, *Journal of Open Humanities Data* 8 (202). doi:10.5334/johd.60.
- [32] D. G. White, Mining the forgotten: Manuscript sources for black women's history, *The Journal of American History* 74 (1987) 237–242.
- [33] S. R. Carroll, I. Garba, O. L. Figueroa-Rodríguez, J. Holbrook, R. Lovett, S. Materechera, M. Parsons, K. Raseroka, D. Rodriguez-Lonebear, R. Rowe, R. Sara, J. D. Walker, J. Anderson, M. Hudson, The care principles for indigenous data governance, *Data science journal* 19 (2020) 1–12. doi:10.5334/DSJ-2020-043.
- [34] M. C. Pattuelli, M. Miller, L. Lange, S. Fitzell, C. Li-Madeo, Crafting linked open data for cultural heritage: Mapping and curation tools for the linked jazz project, *Code4Lib Journal* 21 (2013).
- [35] M. Caswell, Dusting for fingerprints: Introducing feminist standpoint appraisal', in radical empathy in archival practice, *Journal of Critical Library and Information Studies* 3 (2021). doi:10.24242/jclis.v3i2.113.
- [36] M. C. Pattuelli, K. Hwang, M. Miller, Accidental discovery, intentional inquiry: Levaraging linked data to uncover the women of jazz, *Digital Scholarship in the Humanities* 32 (2017) 918–924. doi:10.1093/llc/fqw047.
- [37] J. Bryant, Ephemera, archives, and another view of history, *Archivaria* 40 (1995) 189–198.
- [38] C. Borgman, *Big Data, Little Data, No Data: Scholarship in the Networked World*, MIT Press, Cambridge, MA, 2017.
- [39] M. Moss, D. Thomas, *Archival Silences: Missing, Lost, and Uncreated Archives*, Routledge, Abingdon, 2021.
- [40] R. Risam, *New Digital Worlds: Postcolonial Digital Humanities in Theory, Praxis, and Pedagogy*, Northwestern University Press, Evanston, IL, 2019.
- [41] A. Isaacson, P. Lalu, T. Nygren, Digitization, history, and the making of a postcolonial archive of southern african liberation struggles: The aluka project, *Africa Today* 52 (2005) 55–77.
- [42] E. Losh, J. Wernimont, L. Wexler, H.-A. Wu, *Debates in the Digital Humanities* 2016, University of Minnesota Press, Minneapolis, MN, 2016, pp. 107–111.
- [43] D. Agostinho, Archival encounters: Rethinking access and case in digital colonial archives, *Archival Science* 19 (2019) 141–165. doi:10.1007/s10502-019-09312-0.
- [44] N. Maurantonio, Archiving the visual, *Media History* 20 (2014) 88–102. doi:10.1080/

# Dante Alighieri (1265–1321) and St Maximus the Greek (1470–1556): Two Authors on the Edge of Two Epochs (Pre-case Biographical studies for the Comparison of the Author's Creative Vision – The Triangle Composition)

Neža Zajc<sup>1</sup>

<sup>1</sup> ZRC SAZU, 1000 Ljubljana, Slovenia

## Abstract

Dealing with two complex and decisive authors that have been important for humanistic encounters, the research in this paper presents an in-depth study of two attempts at biographies. With two different methodological approaches to the life and work of the poet Dante Alighieri and the humanist and theologian St Maximus the Greek, we introduce the complexity of the author's creative process as well as of the historical moment of the period when each lived. We also addressed the problem of the valuation of the Medieval and Renaissance period, because we considered the presented authors to be valid examples for the beginning of the new era of thinking in the history and culture of mankind. Further, the article seeks a precise graphic solution for visualizing the relationship between the creative process and the biographical data. Moreover, for deeper understanding of the theologically inspired and conducted poetic works of the authors, a triangle structure is proposed within which the basic idea of the authors' will and their artistic vision could be expressed.

## Keywords

St. Maximus the Greek, Dante Alighieri, Introduction  
Poetics, Biographies, Renaissance, Individualism, Creativity, Humanism, Artistic visionary.

## 1. Introduction

First The paper will present two examples of older biographies as preliminary pre-case studies. The first is a Dante biography that is an instance of a medieval biography that was among the first harbingers of professional biographies, which began to take shape in the early Renaissance. The second example will present the biography of man who lived at the height of the Renaissance, but whose own works determined the end of this period and announced a completely new era for comprehending an individual's work.

### 1.1. Authors


#### 1.1.1. Dante Alighieri


This example of a poet's biography [cf. later Dante biographies: 1, 2] raises key questions about the meaning and role of eyewitness in biographical testimony. This example will raise key questions about the meaning and role of the biographical testimony of eyewitnesses, as the first biography came from the pen of Giovanni Boccaccio. Though he never met the poet, he was acquainted with Dante's daughter Beatrice, his nephew, two of his close friends, and a near relative of Dante's great love, Beatrice "Bice" di Folco Portinari. As a result, he could gather much more personal information about the poet's life than anyone before.

The Life of Dante (*Trattatello in laude di Dante*) by Giovanni Boccaccio [3], who was Dante's contemporary, represents a typical biography of this prominent man of the high medieval period. In other words, it serves to establish a unity between Dante's personality and his life destiny. In

---

BD 2022-DH 2022: Workshop at the International Conference on Biographical Data in a Digital World 2022 (BD), 25 July 2022, online

 neza.zajc@zrc-sazu.si (N. Zajc)

 0000-0001-5220-3553 (N. Zajc)



© 2023 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

[https://doi.org/10.3986/9789610508120\\_02](https://doi.org/10.3986/9789610508120_02)

this context, it favorably builds the aura of the poet Dante, who was fully involved also in the most important events of his age, especially in Italy: he served in the military, in the cavalry at the Battle of Campaldino, near Arezzo, in 1289, in Florence was Dante at the very center of its political and civic life and served on various governing councils, and from June 15 to August 14, 1300, as he was one of the Six Priors—the highest political office in Florence [4].

Specifically, in Boccaccio's eyes Dante was the crucial personality of a very sensitive time, when the medieval world was already slightly on the decline. Moreover, Dante, although he had knowledge also of the Eastern (Byzantine) world, asserted himself in the center of the Italian lands that he considered to be the significant focal point of the Christian universe. The problem of national interests and universal subjects arose.

At the same time, the provided obligatory comparisons with classical, Roman, and Greek antiquity made in Boccaccio's text posit Dante, who wrote in the Italian vernacular, as being the first secular poet and most talented Italian poet since ancient times. The integrated nature of his character is shown to be an embodiment of the European past.

However, Boccaccio's biography, which preserves a uniquely personal portrait of the poet through a detailed account of his appearance, habits, and inner being, lacks the obligatory chronological pieces of evidence, periodization, investigation, introspective notes, and an objective biographical overview. In addition, it offers little possibility to function as a research tool as the work does not contain any systematic structure or tangible body. In this aspect Boccaccio's work could be properly compared to the hagiographical approaches of Byzantine historiography. From another aspect could be seen as "romanticised and mysticising approach" [5].

Somewhat later, the first Renaissance biographies appeared: by Leonardo Bruni, Giovanni Villani, and Fillippo Villani, whose *De Origine Civitatis Florentiae et ejusdem Famosis Civibus*, Florentine Chronicle, is a document that was preserved among Boccaccio's manuscripts. These were still early biographies of the poet, and characteristic for such texts was that they considered Dante's knowledge not only as being encyclopaedically brilliant, which resulted in his being understood as more than a theological poet but also as a philological poet who was crucial for the determination of the expression of human individuality. The latter assessment spurred, and worked in the context of, the myth of Florence as the "new Athens" or "New Rome." In the early Renaissance period, also the first editions of Dante's work were soon printed.

Leonardo Bruni, in his *Vita di Dante* (1436), included also a critical overview of Boccaccio's work [2]. Bruni took also a deeper look into Florentine matters, revealing some manuscript sources and expressing his doubts as to their being original, claiming that they were an act of forgery. Bruni also considered Dante as an emigrant/immigrant or refugee, pilgrimaging through Italy from Florence to Rome, and Siena, Bologna, Pistoia, and finally to Ravenna, which was indicative of a new author's compassionate sensibility of the Renaissance age. Bruni tried to introspectively gain a deeper view into Dante's theological intentions and his nobleness. Additionally, Bruni in his *Dialogues to Pier Paolo Vergerio* (1401), in which he criticized Dante's inferior Latinity [6], recognized the linguistic superiority of Petrarch and Dante's role in the birth of Florentine humanism.

But, like Boccaccio, also Bruni paid much attention to defining the poetic skill of the poet. For Bruni, Dante was a poet of a second type, since through the study of philosophy, theology, astrology, arithmetic, and geometry, as well as by reading histories, Dante acquired the knowledge he used to adorn and expound upon in his verses.

Giovanni Villani included Dante's social, political, and public involvements in Florentine Chronicle. He mentioned Dante's scrupulous "Embassy to Venice" in his service of the Lords of Polenta. He connected Dante's exile with the entrance of Charles of Valois into Florence in 1301.

Villani interpreted also three noble letters of Dante penned in Latin. Written in a lofty style, one was addressed to the Government of Florence, complaining of his undeserved exile; the second, to Emperor Henry; the third to Italian cardinals, after the death of Pope Clement, praying for them to be united in electing an Italian Pope. Similarly, Filippo Villani, in his *Life of Dante* in *De Origine Civitatis Florentiae et ejusdem Famosis Civibus*, mentioned Venetian matters, their lack of eloquence and that fact that the Venetians lacked a poetic tradition (that is, history had seen no good Venetian poets).

Also worth mentioning is the first literary “exegetical” work of Cristophoro Landino (*La Divina Commedia*, Florence, 1481, with 19 illustrations by Sandro Botticelli). Landino ranked Dante’s poetic contribution to the history of human thought as the highest achievement of the human mind because he lauded Dante’s poetry, saying it sprang from Divine wisdom [7]. With these words, he designated Dante the leading Florentine patriot, scholar, and thinker, a true forerunner of the humanistic era but also the catalyst of the modern era, as his verses influenced all European movements in the shaping of the importance of personal individuality. However, he was recognized by followers, scholars, thinkers, philosophers, theologians, and literary critics as the first man to denote the fall of the medieval world and announce the beginning of the Renaissance period. This is one of the reasons why it is questionable that his biography could have been written at all during his lifetime. On the other hand, the later biographies are mainly not enough speculative and not enough imaginative, only the refrain of what Boccaccio lucidly denote as poet’s biographies in the novelistic sense [cf. 2].

Understandably, we can find many biographical attempts from the following centuries, namely, in the form of subjective contributions. Significantly, these partial or “conditional” biographies of Dante are often placed as the prolegomena to later editions of his works, that is, as prefaces or introductions to his poetry. Another question that Dante’s biography poses pertains to the role of national biographers and biographies, among which are crucial contemporary national biographical bases, with the Italian online one being among the most valuable – <https://www.treccani.it/enciclopedia/dante-alighieri> [8]

### **1. 1. 2. St. Maximus the Greek**

The second example of this case study is based on the biographical knowledge of St Maximus the Greek (ca. 1470–1556), the renowned man who lived in more than three different countries, changed (and learned) three different languages and also changed his name – three times. Consequently, he went through the metamorphosis of three identities (today he has three different Wikisources, in three different languages: English, Russian and Greek). Not only was he changed through being immersed in three different cultures, he also experienced many different occasions of cultural ethical encoding. He, for example, also became close to three different Christian religions – Catholic, Western (Latin), and Eastern (Orthodox) – but with the peculiarities of Slavic or Slavonic in that age. However, this was the period of complexity known as the Renaissance.

He was born Michael Trivolis around 1470 in the Greek Epirus town of Arta, into the Trivolis family. His youth education corresponded to the Byzantine system of the wide branched perspective of classical knowledge. His uncle Demetri Trivolis was a well-known bibliophile and a collector of ancient manuscripts. With the Greek colleagues and scholars Ioannos Laskaris (1445–1535), Marko Mousouros and members of Moschus, he travelled to Corfu, Crete and, alongside Croatian islands, to northern Italy. First, he resided in Florence, where the scholar and grammarian (prominent philologist) Ioannos Laskaris, and at that time also Michael’s supervisor, lived. In Florence, he started his philological work as a professional copyist, when he was already acknowledged as a skillful creator of manuscripts. He was introduced to an elite community of scribes, translators, and professional calligraphers, who were carefully carrying out the process of transmitting ancient manuscripts into a new, printed form. He was probably invited into the circle of the scholars who were involved in the shaping of the Medici library. He met Florentine intellectuals such as Marsilio Ficino, Angelo Poliziano, Cristophoro Landino, etc. Michael also

visited other northern Italian cities: Milan, Ferrara, Bologna, and Padua. Twice, for a longer period, he lived at the Mirandola castle where he taught Greek to Gianfrancesco Mirandola, the nephew of the famous Pico. At that time, he was already critical towards Aristotelian thought, but he also started to study mystical theological writers such as John the Ladder and Pseudo-Dionysius the Areopagite. In 1492 Michael met Aldo Manuzio in the Florence apartment of Ioannos Laskaris. Within a few years he was already in touch with the newly established printing house of Aldo Manuzio in Venice and with Greek colleagues Ioannos Grigoropoulos, Zacharias Kalliergis, Cretan calligrapher and founder of the Greek Press in Medici Rome, Nikolas Vlastos, and Scipion Carteromach, the correspondence with whom is most widely preserved from that period. At the end of 1490s, he returned to Florence where he regularly listened to the public sermons of Girolamo Savonarola, which affected and touched him greatly; he was present at Savonarola's public execution. Michael Trivolis, who was a friend also of the Camaldolese monk Pietro Candido (Leuceimon), entered San Marco Monastery in Florence in 1502, exactly four years after the death of Girolamo Savonarola (1498), but he remained there less than a year. He left the monastery before being ordained as a priest, remaining a novice. Indeed, he was one of the members of the second generation of the Greek diaspora, as in Venice precisely at that time the first Orthodox community was taking shape. But as a spiritual refuge he was seeking intellectual support similar to that he had found before only in Aldo Manuzio's printing house in Venice.

Nevertheless, Michael Trivolis decided to return to his native Greece, where in 1506 he joined the Holy Mount Athos the Vatopaidi (Vatopedi) Monastery, dedicated to the Annunciation of Mother of God. He was ordained and given the monastic name Maxim (the monastic example of the Saint Maximus Confessor). In the Vatopaidi Monastery, Maxim developed his extensive writing, translation, and transcribing activities, to which he added his acquisition of the knowledge of Slavic languages. He also wrote several hymnological works in verses and edited the hagiographical manuscripts. As an experienced scribe, with calligraphic and linguistic skills, the Athonite monk Maxim was chosen for the mission to the Orthodox lands. During the monastic period in the Vatopaidi Monastery he carried out several Orthodox missions with the Constantinople Patriarch Niphont II, whose faithful discipline Maxim was at that time. They visited also lands beyond the borders of the Holy Mount Athos, Ochrid, and others places in Macedonia, Albania, Bulgaria (Melnikov) and Moldo-Vlachia (Wallachia). The monk Maxim was sent, therefore, as a translator from Greek to Old Church Slavonic, from Athos to Moscow, at the invitation of the Russian emperor Vasili III. But in Russia, although he was immediately recognized as one of the wisest men of his age, he was spreading a mindset opposite to that of the prevailing church-governing authority since he defended non-privatization and non-ownership of church property (particularly monastic property). Crucial in this sense was, therefore, his opposition to the tendency of the Russian Orthodox Church (autocephaly) towards independence from the church of Constantinople; he sensed in this imperial ideology serious contradictions with the basic Eastern Christian Orthodox doctrine. He was unjustly punished with more than 27 years of imprisonment in the monastery cell at two Moscow church councils.

The only benefit from the second trial against him was that the punishment (to silence, total privatization in the dark and starving conditions, without permission to talk, to communicate, to write, and even to read books) was permitted to be slightly milder: he was allowed to write. From then, around 1536, he began to write his own apologetically marked, and polemically involved theological works, most of them also imbued with a highly distinct monastic attitude. The character of prayers and the humble position he was evoking were formally framed also by the hymnographical and liturgical quality of his texts [9]. However, he was writing in his own personal form of the Old Church Slavonic language (idiolect), which was hard to understand not only for his Russian contemporaries but also for modern scholars. Consequently, many modern treatises do not properly perceive the value and the meaning of his Orthodox writings (often there are one-sided ideological or purely linguistic treatments mainly from Russian and other



pro-Russian scholars). Thus, it was not until 1988 that Maksim Grek was canonized by the official Russian church.

So, as mentioned, as Michael Trivolis he travelled through Greek lands, the islands of Crete and Corfu, and, northern Italian cities, to the multicultural Holy Mount Athos, where he became a monk. Now Maximus (Trivolis), he travelled to Moldo-Vlachia (Wallachia), Albania, Macedonia, Constantinople and finally to Muscovite Russia, where he was called Maksim Grek. Maximus the Greek, as a unique humanist, met people from different social levels and the highest intellectual milieu. Today his manuscripts are housed in the manuscript departments of various European libraries, in Rome (Vatican Apostolic Library), in Florence (Biblioteca Lanurenziana), Cremona, Milan (Biblioteca Ambrosiana), Oxford (Bodleian Library), Russian (Saint Petersburg, Russian National Library-RNB; Moscow: State Historical Museum, Russian Government Library-RGB).

## **2. The synthetic treatise**

These two case studies are represented from two different points of view. The first, on the life of Dante Alighieri, is shown in terms of the authors of his biography. His most important life work was doubtless his Poetics (e.g. all works expressing his personal view on poetic literary submissions) by which he, indeed, passed through times, and was in his vision created for the eternal.

The second, on the life of Maximus the Greek, is presented through different sources in a reconstructive attempt to provide a full picture of his biographical destiny. The first that have to be mentioned are the hagiographic Russian sources from the beginning of the seventeenth century, followed by secondary sources authored by scholars from Russia [10, 11], Italy [12], Slovenia [13, 14], etc. The seminal analysis in terms of synthetic recognition between the three different historical personalities was made by French scholar Elie Denisoff [15] during the Second World War (1943). He managed to identify and to clear up the fact that humanist and philologist Michael Trivolis, Athonite monk Maxim (Trivolis) and Maximus the Greek – alias Maksim Grek in Muscovite Russia – were one and the same person.

## **3. The symmetrical comparison**

To provide the most updated data from reliable sources as the research progresses, the two presented methodological approaches have to be combined. The same procedure is demanded also to address the peer-review process for the researched subject. The basic method necessarily pertains to the reconstruction of the biographical data, but in order to best follow the intention reconstructing the past of an individual's life, the researcher has to be extremely careful in dealing with the selected type of sources. Each of the methodological approaches presented here must include possibilities for allowing a wide range of viewpoints, from introspective attempts at capturing the historical moment during which the subject of the biography lived, to critical consideration of contemporary and other secondary sources. Further biographies from primary sources could prove decisive in terms of valid testimonies.

In fact, autobiographical literary samples, which are decisive in the context of the author's works, have to present the basic material along with, for example, manuscript forms of supplementary biographical sources. For example, Dante's Vita Nuova is considered the author's "autobiography of sorts, since it is a first-person narrative that purports to tell of things that actually happened in the life of the narrator" [4].

By doing this, the researcher must often search for the historical achievements from various historical disciplines, such as epigraphy, onomastics, heraldry, numismatics, paleography, and prosopography. Thus, it is necessary to evaluate also different linguistic (dictionary, lexicographical) possibilities in revealing biographical data. It is also important to pay attention to different cultural values and the reasoning that stems from them. Not surprisingly, this means that the research is interdisciplinary; nevertheless, it remains, in fact, a thoroughly dedicated study. In other words, the researcher often must devote himself to this kind of examination.

Both biographies reveal the complexity of biographical structure and the problems of reception of the individual. Dealing with the identification of the author and the person of the subject, the concretization of the birthplace, the role of names, the distribution of the facts, and the place in the social structure, however, often leads to very different interpretations. Both biographies represent the life of creators-authors (artists, writers), which the biography could render more problematic, because the obligatory perception of them as the author has to be permanent and constant, and subtle as well. These are valid examples of biographies that cross countries and cultures and address also the different historically decisive contexts of awareness. Both presented case studies – the first of the world-famous poet Dante, and the second of a saint who was extremely important for the establishment of the Russian language of literature, original philosophy, and theology – are examples of the problematic historical canonization of people in the history of humankind.

In the age of digitalization, it is, therefore, necessary to be even more careful about what we accept and what we accept as true versus what is only probable, as well as to judge what is relevant at all. The biographical visualization could be accepted only under a condition that could provide deeper insight into or a better understanding of the author's work. Other benefits must be neglected since we are in an age when the brief acknowledgment of the masterpieces of historical personalities does not bring the perhaps expected intellectual comprehension. Unfortunately, the years of digitalization and visualization have not yielded many advances; rather, they have given rise to many misunderstandings and merely superficial, rash insights founded on superficial contact with authors, without great awareness of the proper historical context and background.

Similar analytical topics should only reveal the cultural memory and the author's heritage that need to be carefully investigated and properly presented.

I. Both authors were connected to the same tags/targets/topics/motifs/themes:

- 1) "Pilgrimage" (though neither of them in the literal meaning); both were exiled and died in exile, not in their own homeland; their great (only) wish was to be buried in their homeland.
- 2) both were unjustly accused of something they considered themselves to be not guilty of
- 3) both very critically minded (Dante in *Vulgari eloquentia* opened the "first scientific literary criticism in the modern world" [16] and early modern literary history [17]; both were connected with the ecclesial and government circles of their ages
- 4) Humanists; highly engaged thinkers of their age; Renaissance men, one a harbinger announcing the European Renaissance, the other expressing the collapse of the Western European "old" world.

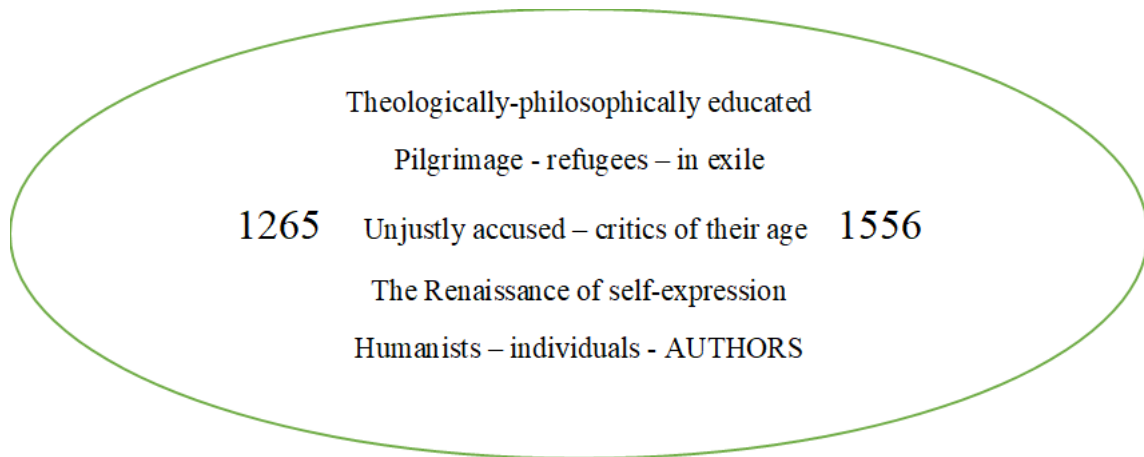


Figure 1

The years 1265 (the birth of Dante Alighieri) and 1556 (the death of Maximus the Greek) could represent the beginning of the collapsing of the medieval world and the final clash in the middle of the sixteenth century to denote the margin of the Renaissance period and the gradual onset of the modern era in the history of human thought, when the European intellectual was able to express his creativity in more isolated, deeply personal, secluded and clearly anthropocentric circumstances. Their deeply personal spiritual vision of the lonesome and poetic destiny of the pious human being came to the fore. Exactly this principle is what Dante was already following, according to the words of Boccaccio: "Studies, and especially those of speculation, to which our Dante [...] entirely surrendered himself, tend to demand solitude, liberty from anxiety and tranquillity of mind." [3]

Indeed, Dante was one of the crucial authors who denounced the transitional period, namely, between the aesthetic theories of the thirteenth century and the Renaissance. During this time artists were aware of their individuality, which they expressed in a new sense, adding to the history of aesthetic feeling and theory [cf. 18, p. 91]. With embryonic 'medieval' humanism of Dante Alighieri [19, p. 224] begins a period which and terminates in the early sixteenth century, when Italian humanists ceased to form part of the mainstream of creative literature in Italy and Europe.

Both Dante and Maximus the Greek searched for a new language of the individual and they dedicated themselves to this apologetically enlightened appeal all their lives

#### 4. The diametrical evaluation

II. Dante Alighieri and Maximus the Greek:

- 1) were talented and extensively productive writers; theologically and philosophically educated.
- 2) were very sensible authors as auctores.
- 3) were prominent individuals.
- 4) had an intellectual heritage that might apply for/to all humans; were known for the transnational character of their works.

The relation between their creative processes and the biographical facts, therefore, presents the main challenge of these biographical case studies (the similar attempt by the usage of the Cushman's PolyCub framework one could find recently [20]). The biographical presentations should thoughtfully and discretely show that neither the external repression from the governmental authorities nor the pressure from the misunderstanding of their enemies

suppressed their inspiration. In fact, the biography of Dante and St Maximus the Greek must lead to recognition of the entire spiritual positioning of their literary work, because they as authors were looking at it only in terms of future circumstances. The increasing literary appeal for the author's creation, in both cases, even ameliorated over time, becoming significantly stronger.

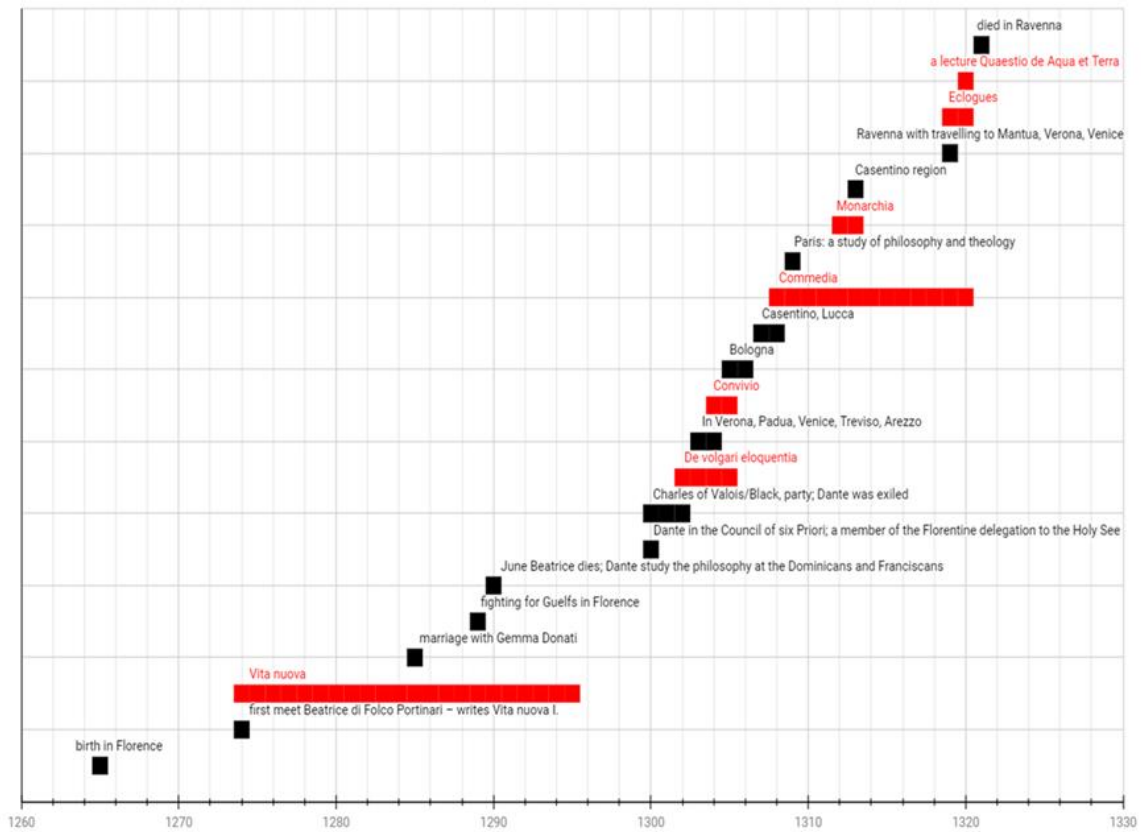


Figure 2  
Dante Alighieri (1265–1321)

Dante defined such trans-temporal attempts and investigations in *Monarchia*, but also in *De vulgari eloquentia*, when he reflected on the etymology of the word author: “And inasmuch as ‘autore’ derives from this verb, it is taken to refer to the poets alone, who with musical art have bound together their words: and with this meaning, we are not concerned at present (4, 6)” [21].

Dante’s Poetics were created hand in hand with his life destiny. For Landino and Marsilio Ficino the goal of the soul’s journey in Dante’s *Commedia*, allegorized in the pilgrim’s journey, was precisely consideration as the highest contemplation of divine matters [cf. 5]. Already in the Renaissance period, echoes and reminiscences of Dante’s verses were evident in many writers, including Petrarch, Boccaccio, Angelo Poliziano, Girolamo Savonarola, Ariosto, Trissino, Folengo, and Michelangelo Buonarroti [cf. 5]. Later we could find “his voice” in the poetry of such prominent and contemporary authors as (also) English-language authors of nineteenth century, Ugo Foscolo, Giosuè Carducci, Mary Shelley, John Ruskin, George Eliot, Charles Eliot Norton, Ralph Waldo Emerson, Leigh Hunt, Byron, Coleridge, Browning, Alfred Tennyson, Keats, etc., and in the twentieth century, T. S. Eliot, Anna Akhmatova, Osip Mandelstam, etc.

Maximus the Greek contributed much to the transmission of Slavonic, linguistic, etymological, grammatical and philosophical topics, alongside the hagiographical, liturgical, patristic terms that permitted him to express the most refined, sophisticated and complex theological expressions in the adequate linguistic form. But since many of his texts reflect his personal monastic prayers and Byzantine hymnography [9, p. 285-318], they are consequently touched by a significant poetic effect. The further result was that St Maxim the Greek made not only an important revision of the Russian liturgical language; in addition, his grammatical surveys and linguistic decisions about the Church language were included in the first printed books on Russian Grammar by M. Smotriskij in the eighteenth century. Consequently, the linguistic types that St Maxim the Greek used in his personal writings and biblical translations appeared in the normative language of the Russian literature of the nineteenth century in the most respectable literary works of A. S. Pushkin, F. I. Tiutchev, and especially in F. M. Dostoevsky, N. V. Gogol, A. P. Chekhov, L. Leskov [22]. He also quite good denoted the problem of the crisis of the modern humanistic consciousness in the Western (European) philosophical and theological investigations, connected with such movements as are the phenomenology and the existentialism.

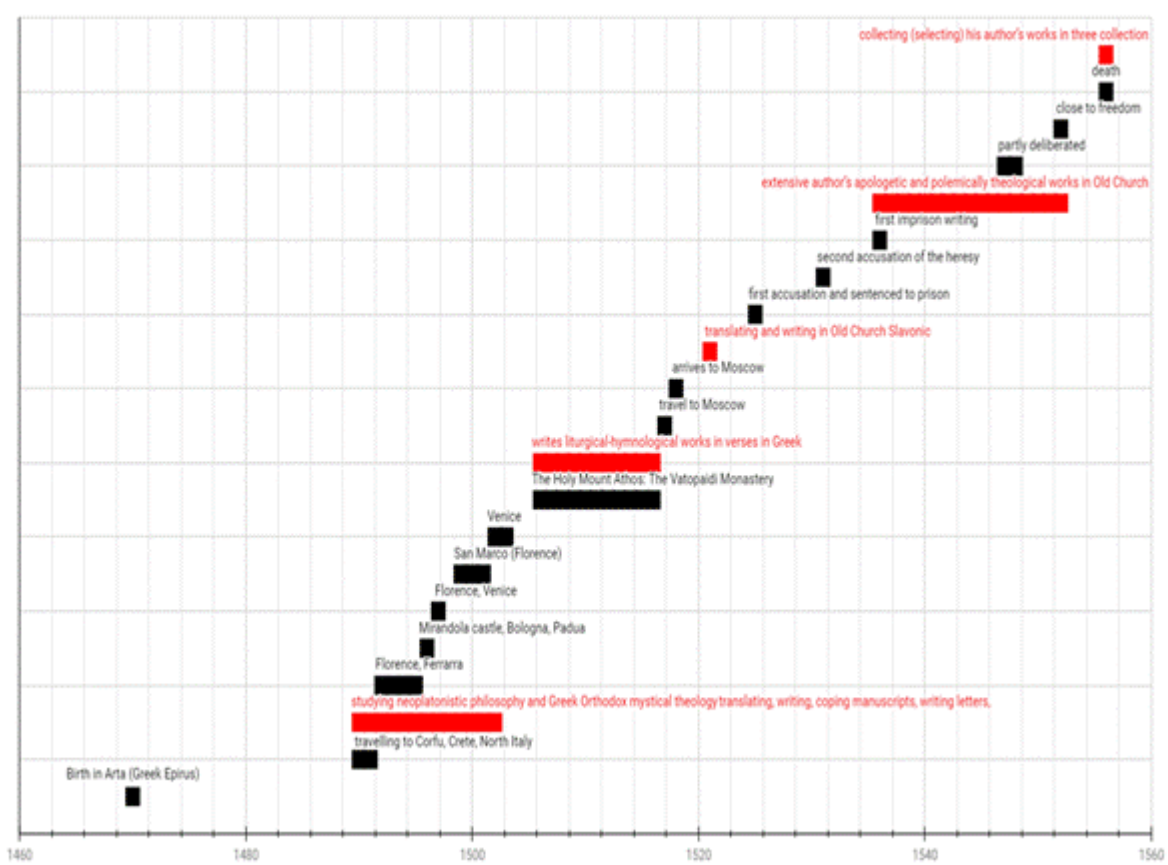


Figure 3

St. Maximus the Greek (cca. 1469/70-1556)

## 5. The author's (theological) view of (the source of) inspiration

Both authors were intrigued by the theocentric worldview that had entirely shaped their vision of the human being in earthly time. Since this is crucial for the reception of their written word, the researcher must be aware of providing a highly prudent evaluation of each “man of letters.” However, although Dante had man in mind, a poet, perhaps himself, was, in his opinion, addressed by the Highest to create poetry that was subsequently not written by his own will. He thought that he was doing something in service to God. The reason? The writer is ever aware that the Highest Instance demands this from the individual: that he creates throughout his lifetime, for as long as he is capable of doing so. This could be, not accidentally, compared to their Christian belief in the Second Coming of Christ and standing before the Last Judgment. Moreover, Dante saw the individual work as the only proper solution for keeping a man humble (and here he was thinking of his own faith and obedience) in the act of ascending to the (earthly) Paradise, which he characterized as a new form of nobility - one that meant individual value (in contrast to ancient nobility, cf. [21, p. 54]. However, this viewpoint was tacitly linked to his contemplation of authorship. In fact, Dante, when he is defining the role of the *autore*, is doing so in impersonal terms, placing himself in the role not of the *auctor* but of one who humbly believes and obeys authoritative words. In other words, he defines himself as a very individual author [21, p. 57].

Maximus the Greek, as an Athonite monk, was not acquainted with contemplating himself as the author, though he remained highly aware that it was only through the written word that he could justify his voice. As long as he was realistic, he also knew that, in fact, his only addressee was God the Son. This is why he refers precisely to “the theology of Jesus Christ” [23, p. 194] – what could be compared to Dante’s Christological view [24, p. 174-175] - alongside establishing the fare mostly sincere, cosmologically shaped, and immutable relationship with the Orthodox Trinitarian system (opposing Lat. Filioque and including also the Holy Virgin Mary and only the Mother of God in the Trinitarian circle [see more 25, p. 399-429, 26]. His humble monastic position enables him to create the purest poetically inspired works, directly from the Holy Spirit (exactly as in the liturgical moment of the invocation of Epiclesis), and to understand God as “the One, Who is mild and essentially Philanthrope,” (lit. one who loves man/men) as can be noticed on the last page of the Liturgical Psalter that Maximus translated four years before his death, in 1552 (fig. 5). Similarly, we could see in his personal prayer “To the Most Holy Spirit Paracletos,” in which the ending in the inverted triangle form contains words that humbly appeal for the redemption of the human soul (fig. 4). His manuscripts often end with a visual recognition of the script in the inverted triangle form, while his literary expression has a single perspective – of, gradually and with full awareness, ascending only as far as the “bottom of Mount Tabor.” Indeed, it could be said, that the basic idea – at both authors - was lying in the specific understanding of the sacred time [27, p. 329-368].



Figure 4

The last page of the manuscript of St. Maximus the Greek's prayer "To the Holy Divine Spirit Paracletos"





Figure 5

The manuscript of the last page of Psalterium, translated by Maximus the Greek in 1552.

This moment of the Lord's Transfiguration could be connected to the same contemplation of Dante (*Paradiso*, Canto 24–26) [28, p. 116] because both authors were thinking a great deal about the vision of the three apostles, Peter, John, and James (which were also the names of the three sons of Dante) as the three biblical *auctore* [21] who were privileged witnesses of the Transfiguration. In the opinion of both authors, this was the only vision that could properly be named theological. In fact, Maximus considered all patristic authors, Church fathers, and the other nine apostles – only three apostles were at the bottom of the mountain – as being very much distant from the insight of the proper theological understanding of God the Son [23, p. 133]. Moreover, one could find, specifically, that both Dante Alighieri and Maximus the Greek spoke not literally but implicitly about the Jesus Prayer in one's heart – Maximus when he placed "the inner man" in literature, Dante when "he is translating" the Lord's Prayer in the opening seven tercets of Canto 11, *Purgatorio* [29]. In the so-called "Monastic cycle" Maxim described the prayer activity of the inner man (Ms. Slave 123, fol. 81 r.), which could be associated not only with the original words of Apostle Paul (Ef 3, 16; 2 Cor 4, 16), but also with the hesychastic and Byzantine patristic practice of the Jesus Prayer in human's heart.



*Figure 6*

A) Dante's **poetic\* theological\*\*** vision: the Highest descends to man.

B) The **theological poetic** vision of Maximus the Greek: ascending toward the Highest.

\* the "poetic" here is used in the sense of "creative".

\*\* the theological is meant here as "theocentric and philosophical".

This kind of evaluation criteria was constructed also in Boccaccio's biography of Dante when the author was seeking for the definition and consequently the difference between the Poetry and the Theology [3, p. 45]. Later examinations of Dante's poetry were also unable to forego juxtaposition with theological aspects [30].

Nevertheless, we could see in the following picture that also Dante in Purgatorio was following the inverted triangle construction as it is ascending. One could argue that the medieval world was mentally constructed in such a triangular manner of thinking. One could also object that palaeographic researchers quite well know that many ancient manuscripts end with such a conclusion. Indeed, all that has been mentioned supports our proposal that these are valid typological pre-case studies (Prototypes) of the biographical approach to the intellectual mindset of the medieval and Renaissance periods.

The triangle form could offer a visual aspect of the Poetics of deeply personal faith. This could represent the structure of the creative process of human thought, ascending to the Highest Instance – and achieving the mysterious and highly mystical messages from the Highest himself (all these could be easily found in the writings of the Eastern Church fathers, in the hagiography, monastic scriptures, the hymnography of Western and Eastern Churches, in the (liturgical) hymnological species, homilies, etc.). This could be a visual rendering of the internal aspect of the individual who strongly believes. Such circumstances finally allowed the human to become more independent, which led to the humanistic period and the Renaissance approach. Consequently, the proposed triangle form could appropriately represent the scheme for one's pious creative process. Only through this kind of understanding of individual creative work could one begin to comprehend the author's artistic vision which shaped his deeply personal Poetics.

This is why we are calling for further examinations of the personal archive, of the author's manuscripts and of personal collections that the author collected and assembled throughout his life.

This is the relationship in between the creative process and biography.

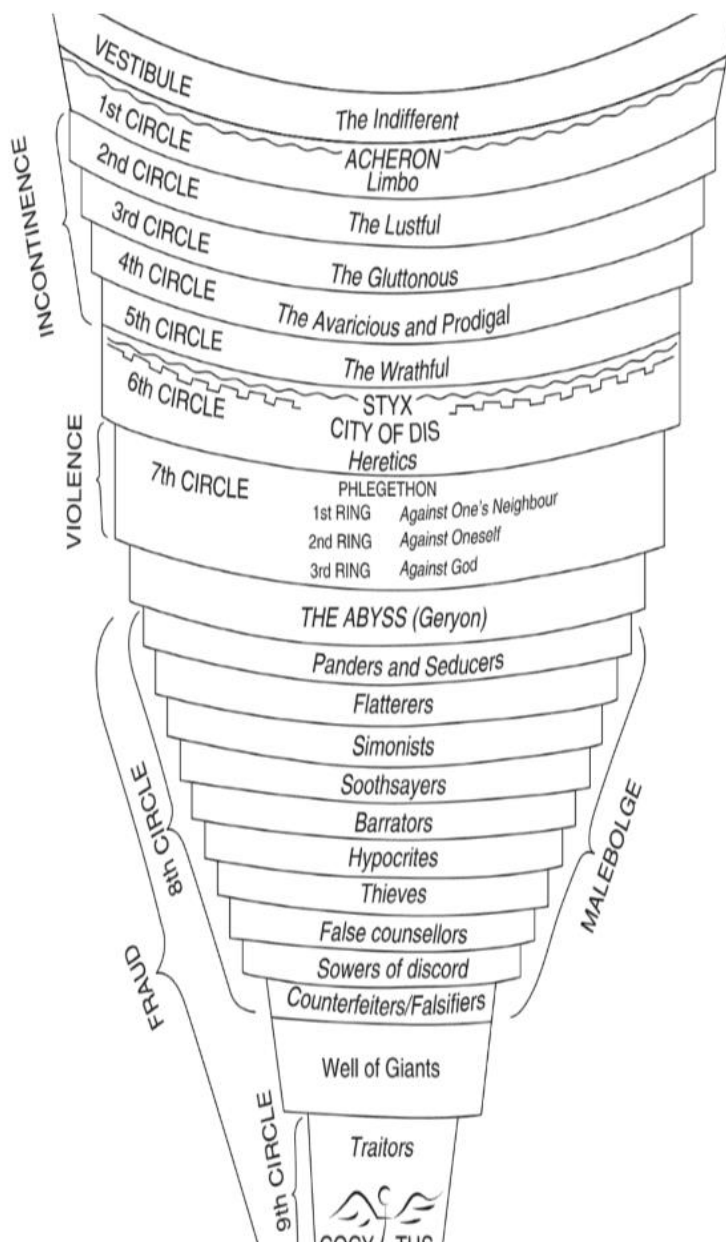


Figure 7

Dante's Hell (Schematic representation of Dante' Hell, source: The Cambridge Companion to Dante' *Commedia*, 2016, ed. Baranski, xix)

## 6. Further researching

A special challenge, however, of such biographical studies could be the project of the visualization in the frame of the following issues: A) visualization of the poetic vision of full spectrum, related to the corpus of works of Dante Alighieri. The enigmatic question: not a single interpretation of Dante's poetry survived as the only valid one. Plus: the source of Dante's individual vision of the Comedy. B) three personal collections of manuscripts of St Maximus the Greek that he collected by his own hand. The enigmatic question remains until nowadays: His own Slavonic idiolect (how could be properly understood) [31] Plus: although there are written

documents (correspondence, copied manuscript, epigrams, short poems, liturgical poems, epitaphs, notes in marginalia of manuscripts) in Greek, and some notes even in Latin, no single page in Old Church Slavonic and fully written with his hand survived until today.

## 7. Acknowledgments

This research was conducted in the frame of the H2020 project InTaVia (<https://intavia.eu>, 2020-2023)

First paragraph in every section does not have first-line indent. Use only styles embedded in the document.

## References

The Manuscripts of St Maximus the Greek:

- Paris, Bibliothèque Nationale: Mss. Coll. Slave 123.
- Moscow, Russian Government/State Library (RGB): Cod. Mss. MDA, fund. 173.I.042.
- The Historical Museum of Moscow (GIM): Mss. Coll. Uvar. 85/14.

Literature:

- [1] D. J. Snider, A Biography of Dante Alighieri. Set Forth as His Life Journey, St. Louis, 1922.
- [2] G. Mazzotta, The Life of Dante, in: R. Jacoff (Ed.), Cambridge Companion to Dante, 2nd. ed., Cambridge University Press, Cambridge, 2007, pp. 1-14.
- [3] G. Boccaccio, Life of Dante. Appendices, L. Bruni, G. Villani, F. Villani, Manuscripts from Boccaccio, Alma Classics, Castle Yard, 2009.
- [4] A. Frisardi (Ed., transl., notes), Introduction, In Dante Alighieri, *Vita Nova*, Northwestern University Press, Evanston, Illinois, 2012, pp. ix-xix.
- [5] F. Ciabattoni, Francesco, Dante Alighieri, In M. Sgarbi (Ed.), Encyclopedia of Renaissance Philosophy, Springer, 2017, pp. 947-958.
- [6] D. Quint, Humanism and modernity: A reconsideration of Bruni's dialogues, *Renaissance Quarterly* 38 (1985): 423-445.
- [7] G. Holmes, Cristoforo Landino, in J. R. Halle (Ed.) Dictionnaire de la Renaissance italienne. Paris-London, Thames&Hudson. 1997.
- [8] <https://www.treccani.it/enciclopedia/dante-alighieri>
- [9] N. Zajc, The Byzantine-poetic path of the works of St. Maximus the Greek (Mikhail Trivolis, Arta, ca. 1470 - St. Maximus the Greek, Moscow, 1556). *Studia Ceranea* (2018): 285-318.
- [10] N. V. Sinitsyna, Maksim Grek v Rossii. Nauka Moscow, 1977.
- [11] N. V. Sinitsyna, Maksim Grek. Moscow. 2008.
- [12] F. Romoli, Maksim Grek na Zapade. *Drevnjaja Rus'* (2020): 32-39.
- [13] N. Zajc, Krogozor slovanske besede, ZRC, ZRC SAZU, Ljubljana, 2011.
- [14] N. Zajc, Some notes on the life and works of Maxim the Greek: (Michael Trivolis, ca 1470 - Maksim Grek, 1555/1556). Part 1: Biography. *Scrinium*, vol. 11, iss. 1 (2015): 314-325.
- [15] E. Denissoff, Maxime le Grec et l'Occident. Contribution à l'histoire de la pensée religieuse et philosophique de Michel Trivolis (Université de Louvain: Recueil de travaux d'histoire et de philologie, 3\* série, 14e fascicule), Desclée De Brouwer, Paris, et Bibliothèque de l'Université, Louvain, 1943.
- [16] Purcell, Sally, The Introduction. In Sally Purcell, *Literature in the Vernacular*, Carcanet New Press Limite, Manchester, 1981, pp. 1-36.
- [17] Appelbaum, Stanley. Ed., Transl. 2006. Introduction. In *La Vita Nuova. The New Life. A dual-language book*. Pages v-xiii.

- [18] Eco, Umberto. *Art and Beauty in the Middle Ages*. Prev. Hugh Bredin. New Haven, London. 2002: Yale Nota Bene, University Press.
- [19] M. L. McLaughlin, *Humanism and Italian Literature*. In Jill Kraye (Ed.), *Cambridge Companion to Renaissance Literature*, Warburg Institute, Cambridge University Press, Cambridge, 1996, pp. 224–246.
- [20] E. Mayr, Eva, S. Salisu, V. A. Filipov, G. Schreder, R. A. Leite, S. Miksch, F. Windhager, *Visualizing Biographical Trajectories by Historical Artifacts: A Case Study based on the Photography Collection of Charles W. Cushman*, in *Biographical Data in a Digital World 2019, Proceedings of the Third Conference on Biographical Data in a Digital World 2019*, Varna, Bulgaria, September 5-6 (2019): 49–56.
- [21] A. R. Ascoli, 2007. From *auctor* to author: Dante before the *Commedia*. In *Cambridge Companion to Dante*, 2nd.ed., In Rachel Jacoff (Ed.), Cambridge University Press, Cambridge, 2007, pp. 46–67.
- [22] L. S. Kovtun, N. V. Sinitsyna, B. L., Fonkich, Maksim Grek i slavjanskaja Psaltyr' (slozhenie norm literaturnogo jazyka v perevodcheskoj praktike XVI v.), In *Vostochnoslavjanskije jazyki. Istochniki dlja ih izuchenija*, Nauka, Moskva, 1973, pp. 99–128.
- [23] Prep. M. Grek, *Sochinenija*, Volume I, Indrik, Moscow, 2008.
- [24] Davies, Dante' *Commedia* and the Body of Christ, In Vittorio Montemaggi (Ed.), Matthew Traherne, *Dante's Commedia. Theology as Poetry*, University of Notre Dame Press, Indiana, 2010, pp. 161–179.
- [25] N. Zajc, 2020. Prep. Maksim Grek i (slovesnyj) obraz Bozhiej Materi v ego sochinenijah. In Tufanova, Olga A. (ed.), et al. *Germenevtika drevnerusskoi literatury: Sbornik 19 = Germenevtika drevnerusskoi literatury [Hermeneutics of old Russian literature]: Issue 19*, IMLI RAN, Moskva, 2020, pp. 399–429.
- [26] N. Zajc, *The Veneration of the Holy Mother of God in the Theology of St. Maximus the Greek*. Analogia:, images, <http://pemptousia.com/2018/02/the-veneration-of-the-holy-mother-of-god-in-the-theology-of-st-maximus-the-greek/#1> (2018), pp. 1-8.
- [27] N. Zajc, *St Maxim the Greek (1470-1556): some notes on his understanding of the sacred time*, *Slavia Meridionalis* (2016): 329–368.
- [28] R. Jacoff. *Introduction to Paradiso*. In *Cambridge Companion to Dante*. Second edition. Ed. Rachel Jacoff. Cambridge University Press, Cambridge, 2007, pp. 107-125.
- [29] J. T. Schnapp, *Introduction to Purgatorio*, in *Cambridge Companion to Dante*. Second edition. Ed. Rachel Jacoff. Cambridge. Cambridge University Press, 2007, pp. 91–107.
- [30] G. Getto, *Poesia e Teologia nel Paradiso di Dante*, in: *Aspetti della Poesia di Dante*, Florence 1966.
- [31] N. Zajc, Neža, *Some notes on the life and works of Maxim the Greek (Michael Trivolis, ca 1470 - Maksim Grek, 1555/1556): Part 2: Maxim the Greek's Slavic idiolect*. *Scrinium*, volume 12 (2016): 375–382.



# Mapping biographies in a Relational Database. Biographies of Luxembourgish soldiers in the Second World War

Nina Janz<sup>1,2</sup>

<sup>1</sup> 1 Luxembourg Centre for Contemporary and Digital History, University of Luxembourg, 11, Porte des Sciences, Esch-sur-Alzette, 4365, Luxembourg

<sup>2</sup> NIOD, Institute for War, Holocaust and Genocide Studies, Herengracht 380, 1016 CJ Amsterdam, Netherlands University

## Abstract

Project WARLUX researches the personal side of the history of Luxembourgers born between 1920 and 1927 who were recruited and conscripted into German services under the Nazi occupation in Luxembourg. In representing different personal testimonies and individual war experiences, the lives of these men, women and families are uncovered. The team establishes a relational database to be able to represent these war experiences but in doing so, also encounters several challenges, such as a data structure that is too rigid and strict to “map” the fluid and unpredictable life patterns of the study subjects. This article proposes “mapping” with a relational database to represent and analyze different life paths departing from the profiles (military unit, prisoner of war (POW) camp etc.). The objects in the database structure can be a person’s individual profile, but also institutions or “life events”. Every single object is treated equally and connected in order to create a separate “biography” of a person and his or her war experiences.

## Keywords

Relational database, nodegoat, biographical data, WWII

## 1. Introduction


People’s biographies are never linear. Especially during times of crisis such as the Second World War individual life paths are unpredictable. A relational database can support a researcher by providing a more sufficient structure to present non-linear biographies. During the Nazi occupation of Luxembourg, around 12 000 women and men (born 1920--1927) were conscripted to serve in the Labour Service (Reichsarbeitsdienst, RAD) and the German Army (Wehrmacht). The WARLUX project collects their biographical data and collates their individual life paths throughout the war and the post-war period.


This study revolves around approximately 1200 recruits and their family members, serving as a case study to delve into their biographies and individual narratives during the wartime period. The primary objective is to present various analytical approaches to understanding the Second World War, offering alternative life stories and personal experiences. Their life experiences were exceptionally diverse and often controversial: post-war, many were labeled as “collaborators” due to their service in the German Army. While some enlisted voluntarily, others were coerced into frontline service, and some resorted to hiding or desertion. In essence, we encounter a complex tapestry of profiles and paths throughout the war, intertwined with life-altering choices – whether to serve, defy orders, or join the resistance. Each of these choices carried profound consequences. For instance, deserters faced arrest and execution, and their families were relocated to the Eastern regions of the German Reich. Thus, the myriad of data points and life “events” is pivotal in reconstructing these individual profiles.

---

Biographical Data in a Digital World 2022 (BD 2022) Workshop, July 25, 2022 online

 [n.janz@niod.knaw.nl](mailto:n.janz@niod.knaw.nl) (N.Janz)

 <https://orcid.org/0000-0001-6251-7740> (N. Janz)

 © 2023 Copyright for this paper by its authors.  
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

[https://doi.org/10.3986/9789610508120\\_03](https://doi.org/10.3986/9789610508120_03)

Our research is not confined to mere biographical data collection; it delves into the personal experiences, narratives, reactions, choices, contradictions, and survival strategies of these individuals during and after the Second World War. This extensive data collection spanned a three-year period, during which the author and a PhD researcher collaborated closely. We organized the data from the WARLUX project, with its complexities, into a relational database, allowing us to chart the wartime experiences of our subjects. We faced the challenge of creating a comprehensive data corpus from diverse materials and institutions, which we intend to utilize exclusively for qualitative analysis. Our objectives extend beyond traditional prosopography or collective biography studies, as undertaken by other researchers [1]. Instead, we focus on the individual life narratives, seeking to move away from generalizations and the collective portrayal of Luxembourgers' life stories and decisions.

In this paper, I will elucidate our data model by providing a structured map of distinct life phases and their consequential impacts within our relational database. I will begin by exploring the historical context of the period, delving into the profiles of individuals and the accompanying datasets, elaborating on our methodology and biographical approach, and providing a comprehensive overview of our data model's construction. Subsequently, I will undertake a critical examination of our preliminary results, analytical insights, and our approach to working with a relational database, all while considering the ongoing nature of our research.

## **2. Historical context**

In tandem with Belgium and the Netherlands, Luxembourg faced invasion on May 10, 1940. Subsequently, the Grand Duchess departed the country, and a civilian occupational administration was established. This administration enforced German laws and regulations on Luxembourg's territory, with the ultimate aim of annexing the former Grand Duchy, inhabited by approximately 200 000 people, into the German Reich under Gau Moselland.

According to National Socialist racial ideology, Luxembourgers were categorized as ethnically "Germanic". Consequently, on May 23, 1941, they were legally conscripted into the Labour Service (Reichsarbeitsdienst, RAD), and on August 30, 1942, into the German army (Wehrmacht) [2]. This conscription of non-citizens clearly violated international law. Faced with coercion and threats from the Nazi regime, over 12 000 Luxembourgers complied with these orders and joined Nazi forces. Many others evaded the draft or joined the resistance, while some deserted during their leave and never returned to their regiments. The occupation gave rise to a "power vacuum," unleashing fresh dynamics that allowed residents to adapt, innovate, and make choices. As historian Tönsmeier [3] aptly noted, "Occupation opens up possibilities for action for those impacted, even as it imposes constraints on their actions". Individuals were compelled to choose their paths: resist, collaborate, join Nazi organizations, volunteer, or quietly adapt to the situation.

The date of August 30, 1942, held particular significance for Luxembourgers during and after the war. Young men who voluntarily joined the German military or police before the August 1942 draft order received state support, and their families gained social prestige within the Nazi administration. Since the date of enlistment held wartime significance, it also remained consequential after the war when these men were eventually tried by Luxembourg civilian courts and convicted as traitors and collaborators [4].

As this brief overview of the historical context illustrates, individual data and choices assume utmost importance in constructing the data model for studying the biographies of Luxembourgers during the war.

## **3. Research Question**

In Luxembourg, the occupation gave rise to a myriad of reactions. The majority of the population stoically endured the imposed restrictions, changes, and intimidation, resulting in



various responses such as acceptance, collaboration, resistance, or passive compliance. These roles assumed by residents were subjected to critical scrutiny by society, leading to the emergence of prejudice, mistrust, and reservations, both during the wartime period and in its aftermath. The allocation of roles within society underwent close examination, with occupied Luxembourg keenly observing the developments and their impact on fellow citizens. Questions about who cooperated with the occupation, who sympathized with the occupiers, who gained advantages in terms of job opportunities or access to goods, who benefited, who suffered, who resisted, and who fought became paramount, particularly in the post-war era when the events of the wartime period were evaluated.

In particular, the largest group directly affected by the occupation, those who were conscripted—whether called to arms or sent to labor camps -- were subjected to close scrutiny, evaluation, categorization, arrest, praise, celebration, or criticism. Their designations, “Ons Jongen” (our boys) or “Zwangsrekrutierten” (Forced Recruits), carried emotional weight, underscoring their coerced enlistment and subsequent fates. Society collectively grappled with these terms, negotiating their significance for those who had been directly impacted. Concepts like “resistance”, “collaboration”, “heroes”, “victims”, and “traitors” were unequivocal, yet Luxembourg recognized the need to challenge and deconstruct these narratives, fostering a deeper self-awareness of its intricate role during the Second World War.

This project aims to contribute to this process by conducting an analysis of individual biographies, with a specific focus on personal records of soldiers in the German Army (Wehrmacht) and women in the Labour Service (RAD).

#### **4. Historiography (research and data)**

The largest group of people “directly” affected by the occupation in Luxembourg has been the subject of recent studies, which have predominantly approached them from a “collective” or “quantitative” perspective [5]. Quantitative studies are undeniably valuable for analyzing war biographies. However, in national works, this group has often been examined collectively, with individual biographies occasionally highlighted and then generalized. Frequently, the patriotism and resistance of specific individuals have been emphasized to serve as representative of the entire group. In prior research, scholars tended to focus on single-actor groups, such as deserters, forced recruits, or collaborators, when reflecting on the wartime period in Luxembourg. Before these studies, these groups were largely perceived as undifferentiated collectives with limited agency and sometimes seen through a fatalistic lens [6]. This perception was influenced by the categories used within the post-war judicial, political, and administrative processes of reckoning with the German occupation. These categories often automatically assigned guilt or innocence and were shaped by post-war policies related to citizenship acquisition [7], compensation [8], and commemoration [9]. The connection between war experiences and discourses on national identity has been highlighted by Peter Quadflieg [10]. Luxembourg's master narrative of being a “nation résistante et martyre” draws significantly from the victim discourse that emerged from the 20th-century wartime experiences [11], which also encompassed the conscripts.

Recent studies have started to delve into Luxembourg's culture of collective remembrance, offering fresh insights into the diversity of wartime experiences. They reveal that many actor groups have been overlooked in the current state of research and public discourse. This oversight is evident in the struggles faced by the “Zwangsrekrutierten” (Forced Recruits) in obtaining post-war compensation [12]. In the current historiographical discourse, the “Forced Recruits” are the predominant actor group and are often viewed as a cohesive collective [13]. However, contrary to the national narrative, not all of them displayed heroism or deserted. The majority served at the front, quietly completed their service, returned home, and often never discussed their experiences. Our goal is to uncover this “average” segment of the majority and illuminate the life stories hidden between the lines. This includes exploring the reasons why a

Luxembourger might have volunteered and understanding the social impact of military service within their communities.

## **5. Sources / Dataset**

### **5.1. Origin**

While biographies are typically structured to encompass key stages of life, including birth, education, work, and ultimately, death, the impact of external events, such as the Second World War, can profoundly alter the trajectory of one's life story. Traditional biographies may deviate, take unexpected turns, diverge from the original life path, and, in some cases, abruptly conclude.

Our subjects have been exposed to various conditions and institutions, such as military service or imprisonment, and have been actively engaged in diverse networks, ranging from those of soldiers to resistance fighters. The pertinent sources documenting their experiences are dispersed across different European countries, originating from various contexts and homelands. The occupying Nazi administration, responsible for registering young men and women for labor and military service, generated organizational and official documents, statistics, and standardized forms and cards. These records are preserved in the Luxembourgish National Archives and other repositories. Beyond basic information like names, birth dates, and places of residence, we require additional data, such as military records found in German archives. Further information on captivity, repatriation, and compensation was gathered in the post-war period. Luxembourgish state surveys and statistics offer comprehensive insights into the experiences of the war generation during the conflict. Consequently, our dataset comprises diverse and heterogeneous sources.

A structured collection of data related to this group exists on a website by the Fédération des Enrôles de Force [14]. Unfortunately, this dataset is known to be inaccurate and incomplete, rendering it insufficient for extensive research. With the WARLUX database, our objective is to create a new and enriched dataset, primarily focused on the case study of Schifflange, an industrial town in the southern region of Luxembourg.

### **5.2 Data Sample**

While the conscription affected a significant number of over 12 000 men and women, Project WARLUX narrows its focus to a case study involving approximately 300 recruits from Schifflange and their families. At the onset of the war, Schifflange boasted a population of around 5 000 inhabitants [15]. In totality, our data sample encompasses roughly 1 200 individuals, including both recruits and their family members [16].

The comprehensive data entry process, a pivotal aspect of this research endeavor, was carried out by the author in collaboration with a PhD researcher and several student assistants. This substantial data collection and entry effort primarily spanned from August 2020 to July 2023. It's worth noting that the PhD researcher will continue to actively engage in the ongoing tasks of analyzing and using the dataset, with these activities expected to persist until possibly 2025. Consequently, given the ongoing work and the possibility of new records being incorporated, the final results of this project are anticipated to be available in 2025 [17].

### **5.3 Ambiguity and Uncertainties**

Handling such a diverse dataset presents a considerable challenge in terms of data collection and constructing biographies. As previously mentioned, a substantial portion of the data is riddled with ambiguity, errors, and gaps. For instance, in the post-war period, we encountered a

statement from a former Wehrmacht soldier claiming desertion. However, we never came across definitive evidence from the German authorities in the form of a warrant, court record, or missing unit report. It is important to note that desertion from the army was often celebrated as an act of “resistance” [18]. However, the absence of data on his desertion in the German military archives does not necessarily negate the event itself. Documents could have been destroyed, lost, or the army might have reported his desertion only towards the end of the war. Therefore, during the data entry process, we must meticulously record both the source of information (such as an assertion in the veteran's memoirs) and whether it aligns with other data (e.g., through additional witness statements).

Cases where information and other potentially inaccurate data are provided must be treated consistently within the data model, with any ambiguity clearly indicated by citing the sources of origin. Ambiguity can arise in terms of conflicting data, particularly regarding the source of the information. Information from the wartime period may differ from self-declarations made in the post-war era, especially regarding the date of enlistment in the Wehrmacht. Some survivors may have provided different or false information about their service duration to avoid being labelled as collaborators. Additionally, memoirs, often written years after the war, may contain details that diverge from what is recorded in military personnel records. Consequently, each data statement is meticulously cross-referenced with its source to ensure the traceability of information and to clarify the context in which the data originated. As Aram and Fernandez et al. elucidate, a relational database proves to be an appropriate solution for gathering data from various sources [19]. Therefore, a relational database appears to be the most suitable approach for managing this complex and often ambiguous dataset.

## 6. Sources / Dataset

This study employs a biographical approach, delving into documents, personal perspectives, and the active decision-making processes that shaped individual life paths. It recognizes that historical actors, within their complex environments, often faced multiple options, and their decisions were not purely self-centered nor solely driven by external constraints. Instead, most decisions were grounded in a situational logic, reflecting what appeared most suitable at the moment [20]. Even in cases where a young man received a mandatory order to join the Nazi forces, the path was not preordained. Our unique method, being used for the first time in Luxembourg, allows for a close examination of individuals, rather than making generalizations about collective groups.

The application of this biographical approach is influenced by Rosenthal's work [21]. She developed methods of objective hermeneutics, drawing from sociological theories, to reconstruct biographies as social constructs. Her analysis of the “Hitler Youth generation” and their experiences during the Third Reich led to the creation of a “life history” approach that offered a new perspective on the history of National Socialism. Rosenthal distinguished between the biographer's perspective in the past and the present, introducing the concept of Gestalt as a phenomenological notion that interconnects experience, memory, and narration, which should be considered in all narrated and written biographies. This underscores the importance of our targeted biographical approach, which focuses on analyzing the individual experiences of soldiers and recruits.

In a similar vein, Fickers and Brüll [22] employ the term “situational opportunism” to emphasize the link between biographical research and sociological decision theory. According to both authors, active decision-making should be contextualized within the individual's situation and the context in which they acted. In our case, we will examine the social environment and the conditions of active decision-making in which each individual actor found themselves. As Schimank notes [23], historical actors typically had several options within the complexities of decision-making processes.

## 6.1 “Life mapping”

Worth, a proponent of social geography, employs life maps as a means to explore the “geography” of the life path, as demonstrated in her article on investigating transitions to adulthood among visually impaired young individuals [24]. In the realm of social geography, life courses encompass a holistic understanding of people's entire life trajectories, their social interactions, and their connections to broader structural forces. It encompasses fateful moments, crossroads, decision-making, records of significant experiences, places, and the people who shape one's life. The term “mapping” is also aptly employed to navigate the intricate life paths and twists and turns experienced by individuals who lived during the war. Within this “Map”, we incorporate information derived from diverse sources, including interviews, documents, personal and public memoirs, and other testimonies. Collectively, this map constructs the life and geography of an individual.

Drawing from Denzin's approach [25], we analyze the narrative in terms of its origins, cultural background, place of residence, identity, life events, and pivotal experiences. Particularly at the turning point, we consider the war as a framing device and a significant turning point in the lives of these individuals. Adhering to a sociological perspective, we follow an inductive process that contextualizes knowledge and human intention [26]. Mapping these lives not only involves connecting different individuals but also linking various life events with one another [27].

## 6.2 Tool

Translating these sources into machine-readable data is a crucial step in our project. The project team has chosen to employ a relational database called nodegoat, which is a web-based research environment tailored for use in the humanities. Developed by Lab 1100, a company based in the Netherlands, nodegoat offers a range of functionalities that facilitate data exploration, including spatial and temporal visualizations. It enables us to construct datasets based on their data model and provides relational modes of analysis with spatial and chronological contextualization. The combination of these features within a single environment allows us to process, analyze, and visualize complex datasets instantly, with a focus on relationships, timelines, and spatial dimensions. Nodegoat operates on an object-oriented framework, drawing inspiration from actor-network theory. In this framework, individuals, events, artifacts, and sources are treated as equals, and their hierarchy depends solely on the composition of the network or their relations [28].

Nodegoat boasts several advantages, including network analysis capabilities, efficient data storage, and a flexible data model. Additionally, our project provides the opportunity to integrate short biographies and personal profiles. However, it also presents challenges, primarily in data modeling complexity (both an advantage and a disadvantage) and the rigidity of the one-set operation for biographical data. More details about these challenges are discussed in the following section. An illustrative project that utilizes nodegoat and follows a similar biographical approach is the “Forced Academic Migration” project at the University of Bern [29]. This project collects biographical data on German and Jewish academics who fled to Switzerland from Nazi Germany, aiming to depict academic relations and career paths in exile.

The object-oriented approach employed levels all entities, treating them without privileging the human over the non-human. Individuals are categorized as objects, actors (or agents) within the broader system of networks. Nodegoat primarily focuses on the creation and contextualization of individual objects as they traverse time and space. Nonetheless, queries and selections can also be made for network analysis outside of nodegoat or for multivariate analysis within the context of prosopography. Since we do not directly rely on statements from the actors themselves, the biographical profiles are naturally constructed and can be distorted due to the source situation. This distortion is particularly evident in the case of political systems like the Nazi regime, where personal data about individuals is tainted by ideology, as it is created within

the context of this ideology. For example, someone's name might be associated with labels like “anti-German” or “traitor”.

## 7. Data model

The primary objective is to gather crucial data, including biographical profiles, and to depict the various stages of individuals' lives. As previously mentioned, nodegoat utilizes the concept of “objects” to represent diverse data. To comprehend and interpret biographical choices effectively, they must be contextualized within the life-worlds of the individuals. Historians encounter challenges when reconstructing past decision-making and action options due to the multiple identities and roles of historical actors. Depending on their role (private, professional, social), these actors had varying degrees of freedom in specific situations. The core motivation is to comprehend past actions or inactions within their situational logic and to explore the interplay between subjective intentions and structural factors, individual decision-making possibilities, and social role patterns.

In the context of Luxembourg, the application of these terms seeks to move away from established labels such as “volunteers”, “forced recruits”, or “heroes”. WARLUX aims to emphasize the individual over group categorizations. A volunteer, for instance, may not necessarily be a staunch Nazi supporter, just as a recruit may have joined voluntarily. The perspective adopted is not one of absolutes; it acknowledges that people's actions are influenced by their circumstances, backgrounds, and various external factors. Moreover, the existing framework, including the country of birth and the prevailing political system—in this case, the Nazi occupation—sets the stage for individuals' possibilities for action and decision-making. These structures provide a horizon that both limits and enables our actions without definitively defining or constraining them.

For different types of information and data, nodegoat utilizes the concept of “objects”. Two specific objects are highlighted as examples.

**Gaasch Roger** [edit](#)  
( ngMy5D176MCbyNCQeL3v5cSDuL.f )

<b>Name</b>	Gaasch
<b>First Name</b>	Roger
<b>Sample Community</b>	Schiffflange (LU)
<b>Whereabouts</b>	returned
<b>Gender</b>	male
<b>Religion</b>	Katholisch
<b>Nationality</b>	LUX
<b>Recruit</b>	yes

[Birth]

▼ ▲ 25 1 - 1 of 1 ◀ 1 ▶

Date Start ▲	Date End ▲	Locati...
03-09-1922	-	-

**Figure 1:** Object Person with standard biographical data

Object Person:

The person's profile encompasses standard data points such as birth and death dates, residence, education, and profession (see *Figure 1*). Information regarding military unit and

imprisonment could have been directly incorporated here. However, we made the deliberate choice to house this information within another object. First, this decision was made to avoid over-determining this object. Second, we aimed to present individuals' lives as “chapters” or separate entities (objects), allowing the Wehrmacht experience to be a distinct chapter in the lives of those involved, but not a mandatory one. By using objects, we represent different life situations.

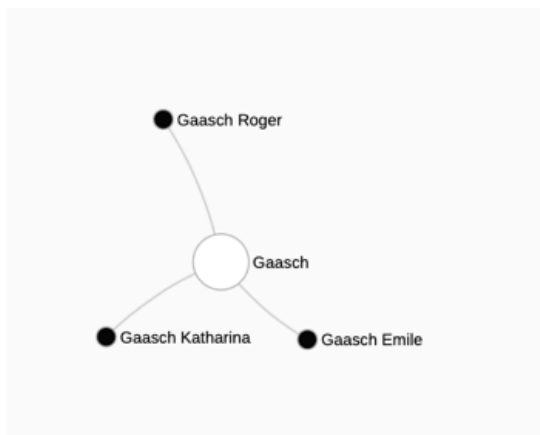
Object Wehrmacht:

The “Wehrmacht” object is not applicable to everyone, as not every individual experienced this life event. Some men voluntarily joined the Nazi forces, while others chose to go into hiding, resulting in the omission of the Wehrmacht object from their “biography”. Instead, we created a scenario or event-like context and traced various decisions and outcomes.

With these straightforward data models, we track the course of individuals' lives and their decision-making processes. We enriched the biographies, following the suggestion of Hyvönen et al. [30], by incorporating additional data such as information from public encyclopedias about military units, geolocations of operational locations, and personal data from ego-documents like letters from the front and memoirs. Nevertheless, these objects and individual profiles are linked to dates, places, and other profiles. As demonstrated by Fokkens et al. [31], our goal is to extract relationships between people and events, specifically events with their participants, timing, and location. In our small-scale project, we established these links ourselves or through data import using CSV files, rather than employing Natural Language Processing (NLP) as Fokkens did.

## 8. Discussion

Through the use of various situations or events as distinct objects, we construct a comprehensive life history map. These objects, such as the Wehrmacht experience, the decision to desert or go into hiding, and the ensuing consequences like family resettlement, are not merely components within a person's profile; they exist as separate entities in their own right (see *Figure 2*). This approach allows us to create a nuanced and detailed representation of individuals' life trajectories, highlighting the pivotal moments and decisions that shaped their experiences.



**Figure 2:** Linked members of family Gaasch

### 8.1 Mapping of *Biographies*

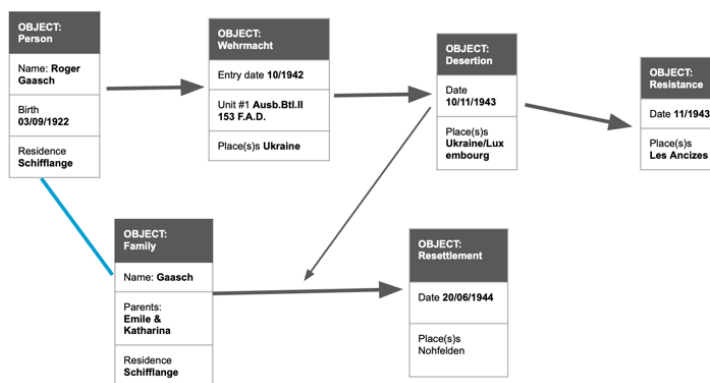
Let's illustrate this approach using the example of Roger Gaasch, born in 1922. In our database, Roger Gaasch is represented within the “Person” object, which contains standard data about him,



including his birth and place of residence. Additionally, he is represented within the “Wehrmacht” object, which includes data about his enlistment in the Wehrmacht and the unit in which he served. His family members, including his parents and sister, also have entries within the “Person” object, with links connecting them to Roger Gaasch (see *Figure 2*).

As we delve into Roger Gaasch's life, we encounter different phases or events. For example, his desertion in 1943 is captured in the “Desertion” object, while his involvement with the French resistance in 1944 is documented in the “Resistance” object. Simultaneously, the consequences of his actions ripple out to affect his family members, who are linked to the “Repression/Resettlement” object (see *Figure 3*).

By structuring the data in this way, we aim to create a comprehensive map of individual biographies and the interconnected relationships between people and events, allowing us to trace the complex trajectories of individuals like Roger Gaasch and the repercussions their decisions had on themselves and their families.



**Figure 3:** Data model WARLUX – Single objects and different life “chapters” and consequences for specific decisions such as desertions (Resettlement for the family)

Depicting the decision-making process within the database poses a complex challenge. Let's consider the scenario of a young Luxembourger who volunteered for the Wehrmacht. Instead of simply labeling him as a “volunteer” in the database, we adhere to the Fickers/Brühl principle of “situational opportunity”. This means that we must examine various conditions to comprehensively represent and capture the nuances of this active decision-making process.

The presumed “volunteer” is connected to his family members, and these connections can be linked to membership in Nazi organizations. By mapping these relationships and affiliations, we can gain insights into the proximity to Nazi ideology, and it becomes apparent that the young man may have volunteered for the Wehrmacht out of genuine conviction. If we include additional information, such as the family's financial status and potential German kinship, we can further investigate the family's social and economic context. It becomes apparent that the young man's decision could have been influenced by financial hardship and/or a strong belief in the National Socialist cause.

However, this example underscores the complexity of the biographies and lives of individuals living under occupation. Their lives were marked by twists and turns, much like other occupied and persecuted people during the Nazi occupation. To accurately represent these multifaceted life trajectories, we need a database that can adequately convey their complexity within their personal and political contexts. We utilize different objects to map various life routes and their subsequent consequences, such as resettlement or material and financial advantages. The date August 30, 1942, assumes particular significance as it marks the distinction between voluntary and forced recruitment of Luxembourgers into the German Army, although it's important to note that some soldiers continued to join voluntarily after August 1942.

As Worth suggested, this data model and life map provide researchers with the means to analyze individual lives, decisions, and their consequences [33]. However, it is important to

acknowledge that this model has its limitations and challenges. Using different life “events” as individual objects or profiles is time-consuming, requiring researchers to navigate between objects for every data entry. Moreover, the data often contains ambiguities, especially in the context of violence and occupation, where every decision can have life-altering consequences and where post-war accounts may present different versions and contradictory statements.

## 9. Conclusion

WARLUX faces the challenging task of transforming individual and unstructured biographical data into structured and interpretable data within the nodegoat database. To avoid imposing a linear structure on individual life stories, the data model needed to be flexible and adaptable, accommodating unstructured and lifeworld models of individuals. This was achieved by treating different life “events” or “situations” as equal objects, placed on the same level as the individuals themselves. This approach allows for a nuanced understanding of each individual's unique life trajectory, rather than imposing a rigid and inflexible schema.

However, it's important to acknowledge that difficulties persist, particularly in the data collection process, where ambiguities and a lack of structure in the data pose challenges. Nevertheless, by treating different life situations and events as equal objects and employing objects equally, WARLUX is able to visualize the lives and individual stories of Luxembourgers during the war as intricate and multifaceted maps, capturing the complexity of their experiences.

### 9.1 Outlook

The next steps involve conducting a qualitative analysis of first-person documents, such as diaries and letters, with the aim of enriching the biographical dataset. This will be achieved by introducing various “experience” objects based on these first-person documents. Notably, a collection of letters has been gathered through a crowdsourcing campaign [34], and these letters will undergo analysis and integration into the database in the upcoming phase of the project.

To incorporate these letters into the database, transcriptions of the letters are generated using HTR/Transkribus [35] and then imported as text files into nodegoat. Within the nodegoat environment, these transcriptions are linked to corresponding objects, individuals, dates/events, and locations. Additionally, text clustering is employed to identify and group relevant keywords and themes within the letters.

This dataset presents a valuable opportunity for further enrichment by adding additional information about individuals, following a methodology similar to that used by Koho et al. in their study on Linked Data based on Finnish POWs in the Soviet Union [36]. This approach could contribute to a more comprehensive understanding of the experiences and biographies of Luxembourgers during the war.

## Acknowledgements

This research is part of the project “WARLUX – Soldiers and their communities in WWII: The impact and legacy of war experiences in Luxembourg” at the Luxembourg Centre for Contemporary and Digital History, University of Luxembourg [37], funded by Fonds National de la Recherche (FNR) [38].

## References

- [1] Roberta Hawkins et al., ‘Practicing Collective Biography’, *Geography Compass* 10, no. 4 (2016): 165–78, <https://doi.org/10.1111/gec3.12262>.

- [2] VBl. CdZ Luxemburg, Verordnung über die Reichsarbeitsdienstpflicht in Luxemburg, 23 May 1941, p. 232; VBl. CdZ Luxemburg, Verordnung über die Wehrpflicht in Luxemburg), 31 August 1942, p. 253.
- [3] Tatjana Tönsmeier, 'Besatzungsgesellschaften. Begriffliche Und Konzeptionelle Überlegungen Zur Erfahrungsgeschichte Des Alltags Unter Deutscher Besatzung Im Zweiten Weltkrieg', Docupedia-Zeitgeschichte, 18.12.2015, 2020, <http://dx.doi.org/10.14765/zzf.dok.2.663.v1>.
- [4] Peter M. Quadflieg et al., 'Mal Blumenstraus, mal Handschellen : Luxemburgische und ostbelgische Wehrmachtrückkehrer zwischen gesellschaftlicher Teilhabe und sozialer Ausgrenzung', in Identitätsbildung und Partizipation im 19. und 20. Jahrhundert : Luxemburg im europäischen Kontext, *Études luxembourgeoises / Luxemburg-Studien* (Frankfurt am Main: Peter Lang, 2016), 293–307.
- [5] Rass, Christoph. 'Sampling Military Personnel Records: Data Quality and Theoretical Uses of Organizational Process-Generated Data'. *Historical Social Research* 34, no. 1 (2009): 172–96; Quadflieg, Peter M. 'Die Zwangsrekrutierung von Luxemburgern Zur Deutschen Wehrmacht Im Spiegel von Wehrmachtspersonalunterlagen'. *Hémecht. Revue d'histoire luxembourgeoise. Zeitschrift Für Luxemburger Geschichte* 4, no. 59 (2007): 401–28.
- [6] Marc Buck, 'Les Jeunes Luxembourgeois "Enrôlés de Force" Dans La Wehrmacht (1940-1945)' (Bruxelles: École royale militaire, 1969).
- [7] Denis Scuto and Gérard Noiriel, *La nationalité luxembourgeoise (XIXe-XXIe siècles) ; Histoire d'un alliage européen* (Bruxelles: Ed. de l'Université de Bruxelles, 2012).
- [8] Peter Helmlinger, 'Ausgleichsverhandlungen“ Der Bundesrepublik Mit Belgien, Den Niederlanden Und Luxemburg" in Grenzen Der Wiedergutmachung', Hockerts, Hans Günter. 2007. *Grenzen Der Wiedergutmachung: Die Entschädigung Für NS-Verfolgte in West- Und Osteuropa : 1945-2000*. Göttingen: Wallstein Verl. Vol.30(2) (2007): pp.447-449; Norbert Franz, 'Der deutsch-luxemburgische Vertrag vom 11. Juli 1959 und die westliche Reparationspolitik nach dem Zweiten Weltkrieg', in ... .. et wor alles net esou einfach : Questions sur le Luxembourg et la Deuxième Guerre mondiale : contributions historiques accompagnant l'exposition: Fragen an die Geschichte Luxemburgs im Zweiten Weltkrieg : ein Lesebuch zur Ausstellung, *Publications scientifiques du Musée d'histoire de la Ville de Luxembourg*, X (Luxembourg: Musée d'histoire de la Ville, 2002), 304–16; Lena Bonifas, 'Le dédommagement des enrôlés de force luxembourgeois après la deuxième guerre mondiale' (Bruxelles, Université libre de Bruxelles, 2007).
- [9] Gilbert Trausch, 'Le Combat Des Enrôlés de Force Luxembourgeois- Mémoire de La Seconde Guerre Mondiale', in *Mémoire de La Seconde Guerre Mondiale. Actes Du Colloque de Metz. 06-08 Octobre 1983, Centre de Recherche Histoire et Civilisation de l'Université de Metz*, vol. 16 (Metz, 1984), 181–200.
- [10] Peter M. Quadflieg, 'Luxemburg - Zwangsrekrutiert ins Grossdeutsche Reich: Luxemburgs nationale Identität und ihre Prägung durch den Zweiten Weltkrieg', in *Kriegserfahrung und nationale Identität in Europa nach 1945*, by Kerstin von Lingen, vol. 49, *Krieg in der Geschichte* (Paderborn: Ferdinand Schöningh, 2009), 170–88.
- [11] Denis Scuto, 'Mémoire et histoire de la Seconde Guerre mondiale au Luxembourg: réflexions sur une cohabitation difficile', *Hémecht : Zeitschrift für xer Geschichte = revue d'histoire luxembourgeoise* 58, no. 4 (2007): 499–513; Benoît Majerus, 'Besetzte Vergangenheiten Erinnerungskulturen an den Zweiten Weltkrieg in Luxemburg — eine historiografische Baustelle', *Hémecht : Zeitschrift für Luxemburger Geschichte = Revue d'Histoire luxembourgeoise* 64, no. 3 (2012): 23–43.
- [12] Eva Klos, 'Umkämpfte Erinnerungen. Die Zwangsrekrutierung Im Zweiten Weltkrieg in Erinnerungskulturen Luxemburgs, Ostbelgiens Und Des Elsass (1944-2015)' (Luxembourg, 2017).
- [13] Peter M Quadflieg, 'Zwangssoldaten' Und 'Ons Jongen'. Eupen-Malmedy Und Luxemburg Als Rekrutierungsgebiet Der Deutschen Wehrmacht Im Zweiten Weltkrieg (Aachen: Shaker Verlag, 2008); Majerus, 'Besetzte Vergangenheiten Erinnerungskulturen an den Zweiten Weltkrieg in Luxemburg — eine historiografische Baustelle'; Trausch, 'Le

- Combat Des Enrôlés de Force Luxembourgeois- Mémore de La Seconde Guerre Mondiale'; Scuto, 'Mémoire et histoire de la Seconde Guerre mondiale au Luxembourg'.
- [14] Fédération des Enrôlés de force (FEDDEF) <http://www.ons-jongen-a-meedercher.lu> [31.08.2023]
- [15] Jérôme Courtoy, Schifflingen im Ausnahmezustand: zwischen Kriegsvorbereitungen, Evakuation und Besatzung: ein Rekonstruktionsversuch (Schiffflange: Administration communale de Schiffflange, 2019).
- [16] Janz, Nina, & Vercruyssen, Sarah Maya. (2023). WARLUX nodegoat database, on recruits of Schiffflange/Luxembourg, Luxembourg Centre for Contemporary and Digital History/University of Luxembourg [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.8138202>.
- [17] Sarah Maya Vercruyssen's doctoral research project is focusing on the social environments of Luxembourgish soldiers and recruits during World War II, (University of Luxembourg).
- [18] Peter M. Quadflieg, 'Luxemburg - Zwangsrekrutiert ins Grossdeutsche Reich: Luxemburgs nationale Identität und ihre Prägung durch den Zweiten Weltkrieg', in *Kriegserfahrung und nationale Identität in Europa nach 1945*, by Kerstin von Lingen (Paderborn, 2009), 170–88, here 181.
- [19] Bethany Aram, Aurelio López Fernández, and Daniel Muñoz Amian, 'The Integration of Heterogeneous Information from Diverse Disciplines Regarding Persons and Goods', *Digital Scholarship in the Humanities* 36, no. 2 (2021): 255–67, <https://doi.org/10.1093/lc/fqaa021>.
- [20] Uwe Schimank, *Die Entscheidungsgesellschaft Komplexität und Rationalität der Moderne* (Wiesbaden: VS, 2005).
- [21] Gabriele Rosenthal, *Interpretive Social Research - An Introduction* (Universitätsverlag Göttingen, 2018).
- [22] Andreas Fickers and Christoph Brühl, 'Situativer Opportunismus Und Kumulative Herorisierung. Ein Experiment Kollektiver Gewissensprüfung', in *Grenzerfahrungen. Eine Geschichte Der Deutschsprachigen Gemeinschaft Belgiens. Vol. 4: Staatenwechsel, Identitätskonflikte, Kriegserfahrungen (1919-1945)* (Eupen, 2019), 8–39.
- [23] Schimank, *Die Entscheidungsgesellschaft Komplexität und Rationalität der Moderne*.
- [24] Nancy Worth, 'Evaluating Life Maps as a Versatile Method for Lifecourse Geographies', *Area* 43, no. 4 (2011): 405–12, <https://doi.org/10.1111/j.1475-4762.2010.00973.x>.
- [25] Norman Denzin, *Interpretive Biography* (Thousand Oaks California: SAGE Publications Inc., 1989), <https://doi.org/10.4135/9781412984584>.
- [26] Sharan B. Merriam, *Qualitative Research: A Guide to Design and Implementation*, Fourth edition, Jossey-Bass Higher and Adult Education Series (San Francisco, CA: Jossey-Bass, a Wiley Brand, 2016), 35.
- [27] Antske Fokkens, Serge ter Braake, Niels Ochelo et. al., 'BiographyNet: Extracting Relations Between People and Events', in *Europa Baut Auf Biographien*, ed. Á.Z. Bernád, C. Gruber, and M. Kaiser (Berlin: new academic press, 2017), 193–227.
- [28] <https://nodegoat.net/about> [31.08.2023]
- [29] <https://en.forced-academic-migration.net/projekt/project> [31.08.2023]
- [30] Eero Hyvönen et al., 'Life Stories as Event-Based Linked Data: Case Semantic National Biography', in *Proceedings of the 2014 International Conference on Posters & Demonstrations Track - Volume 1272, ISWC-PD'14* (Aachen, DEU: CEUR-WS.org, 2014), 1–4.
- [31] Antske Fokkens, Serge ter Braake, Niels Ochelo et. al., 'BiographyNet: Extracting Relations Between People and Events', in *Europa Baut Auf Biographien*, ed. Á.Z. Bernád, C. Gruber, and M. Kaiser (Berlin: new academic press, 2017), 193–227.
- [32] Aimé Knepper, *Vie Ou Mort Des Réfractaires* (Luxembourg: Saint Paul, 1992), 154.
- [33] Nancy Worth, 'Evaluating Life Maps as a Versatile Method for Lifecourse Geographies', *Area* 43, no. 4 (2011): 405–12, <https://doi.org/10.1111/j.1475-4762.2010.00973.x>.

- [34] Janz, Nina, 'THE PARTICIPATORY ASPECT OF CREATING A COLLECTION ON WWII COLLECTING EGO-DOCUMENTS FROM LUXEMBOURGISH RECRUITS AND THEIR FAMILIES', *Etica & Politica / Ethics & Politics*, XXV (2023), pp. 81–103.
- [35] <https://readcoop.eu/transkribus/> [31.08.2023].
- [36] Koho, M., Ikkala, E., & Hyvönen, E. (2022). Reassembling the Lives of Finnish Prisoners of the Second World War on the Semantic Web. In *Proceedings of the Third Conference on Biographical Data in a Digital World (BD2019)* (pp. 31-39). (CEUR Workshop Proceedings; Vol. 3152). CEUR. [http://ceur-ws.org/Vol-3152/BD2019\\_paper\\_5.pdf](http://ceur-ws.org/Vol-3152/BD2019_paper_5.pdf)
- [37] <https://www.c2dh.uni.lu/projects/warlux-soldiers-and-their-communities-wwii-impact-and-legacy-war-experiences-luxembourg> [31.08.2023]
- [38] <https://www.fnr.lu/projects/warlux-soldiers-and-their-communities-in-wwii-the-impact-and-legacy-of-war-experiences-in-luxembourg/> [31.08.2023]

# Creating and Using Biographical Dictionaries for Digital Humanities Based on Linked Data: A Survey of Web Services in Use in Finland

Eero Hyvönen<sup>1,2</sup>

<sup>1</sup>*Semantic Computing Research Group (SeCo), Aalto University, Finland, <https://seco.cs.aalto.fi>*

<sup>2</sup>*Helsinki Centre for Digital Humanities (HELDIG), University of Helsinki, Finland, <https://heldig.fi>*

## Abstract

This paper overviews work on creating Linked Open Data (LOD) ontology and data services and applications for publishing and using biographical collections on the Semantic Web, both for Digital Humanities research and for the general public. The focus is on presenting a series of six LOD services and related biographical portals in use in Finland, based on biographies and person registries of historical people. These web services were developed based on a vision of four technological generations of publishing biographies. The lessons learned in the presented work contributed to the so-called Sampo model for publishing and using Cultural Heritage data in a collaborative way using a shared ontology infrastructure. The six biographical Sampo systems discussed have had up to a million of users on the Web suggesting feasibility of the proposed model and tools used in creating them.

## Keywords

Semantic Web, Biographies, Prosopography, Digital Humanities, Linked Open Data

## 1. Biographies as Linked Data on the Semantic Web

Biographical data are typically strongly interlinked, but published in heterogeneous, distributed local data silos, making it difficult to utilize the data on a global level. Both tangible and intangible Cultural Heritage (CH) data is involved. Furthermore, the content is usually available only for humans to read, and not as data for Digital Humanities (DH) analyses and application development.

This paper argues that these challenges can be addressed effectively using the *Sampo model* [1] and related user interface (UI) tools, such as the SPARQL Faceter [2] and Sampo-UI [3] framework in our case. The paper surveys a line of research where this approach was applied to developing a series of six biographical/prosopographical applications available on the Semantic Web in Finland: WarSampo<sup>1</sup>[4], Norssi Alumni<sup>2</sup> [5], U.S. Congress Proposographer<sup>3</sup> [6], Biogra-

---

*Biographical Data in a Digital World (BD 2022), July 25th, 2022, virtual, Tokyo, Japan*

✉ [eero.hyvonen@aalto.fi](mailto:eero.hyvonen@aalto.fi) (E. Hyvönen)

🌐 <https://seco.cs.aalto.fi/u/eahyvone/> (E. Hyvönen)

🆔 0000-0003-1695-5840 (E. Hyvönen)

© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

[https://doi.org/10.3986/9789610508120\\_04](https://doi.org/10.3986/9789610508120_04)

<sup>1</sup>Available at: <https://sotasampo.fi>; Project home: <https://seco.cs.aalto.fi/projects/sotasampo/en/>

<sup>2</sup>Portal: <https://norssit.fi/semweb/>; Project home: <https://seco.cs.aalto.fi/projects/norssit/>

<sup>3</sup>Portal: <https://semanticcomputing.github.io/congress-legislators/>



phySampo<sup>4</sup> [7, 8], WarVictimSampo 1914–1922<sup>5</sup> [9], and AcademySampo<sup>6</sup> [10, 11]. Table 1 presents a summary of these systems online with their year of publication, application domain, number of users, size of the knowledge graph, and a list of primary data owners.

**Table 1**

Six biographical Sampo portals and LOD services for Digital Humanities; distinct user counts (site visits) by Google Analytics in 2021 October

Portal	Year	Domain	# Users	# Triples	Primary data owners
WarSampo	2015–2019	World War II	1 000 000	14M	National Archives, Defense Forces, and others, Finland
Norssi Alumni	2017	Person registry	unknown	0.47M	Norssi High School alumni organization Vanhat Norssit
U.S. Congress Prosopographer	2018	Politicians	unknown	0.83M	U. S. Congress Legislator data
BiographySampo	2019	Biographies	50 000	5.56M	Finnish Literature Society
WarVictimSampo 1914–1922	2019	Military history	29 000	9.96M	National Archives of Finland
AcademySampo	2021	Finnish Academics	8200	6.55M	University of Helsinki and National Archives, Finland

This paper overviews these systems by relating them to generations of publishing biographies and by explaining their underlying design principles. In the following, related works are first discussed (Section 2) and a vision of four consecutive generations of publishing biographical information is presented. The work on the biographical Sampo systems has been driven by this vision of developing more and more useful and intelligent biographical publishing systems for both researchers and the general public. The lessons learned have gradually evolved into the so-called Sampo model, a set of six principles for developing LOD services and semantic portals on top of them, discussed in Section 4. The paper ends with a summary on contributions and challenges of the presented work.

## 2. Related Work

Biographies are an important source of information for researchers across various disciplines with an interest in the history. [12] Biographical dictionaries are scholarly resources used not only by the academic community but also the by the public. Such dictionaries typically start with narrative text describing the life of a biographee followed by a structured synopsis of his/her basic biographical facts, such as family relations, education, works, career events, and so on.

In addition textual dictionaries there are also prosopographical databases or data services

<sup>4</sup>Portal: <https://biografiasampo.fi>; Project home: <https://seco.cs.aalto.fi/projects/biografiasampo/en/>

<sup>5</sup>Portal: <https://arkisto.sotasurmat.fi> ; Project home: <https://seco.cs.aalto.fi/projects/sotasurmat-1914-1922/en/>

<sup>6</sup>Portal: <https://akatemiasampo.fi>; Project home: <https://seco.cs.aalto.fi/projects/yo-matrikkelit/>

available on the Web. For example, in Austria there is the dictionary OEBL<sup>7</sup> online serving biographical texts and also the Austrian Prosopographical Information System APIS<sup>8</sup> as a further advancement with structured, linked data. From databases and data services online, structured data be exported from the service and/or reused via application programming interfaces (API).

An example of a biographical dictionary is the Oxford Dictionary of National Biography (ODNB)<sup>9</sup> with more than 60 000 lives. It was published in print and online in 2004. Today many dictionaries are available on the Web. These include USA's American National Biography<sup>10</sup>, Germany's Neue Deutsche Biographie<sup>11</sup>, Biography Portal of the Netherlands<sup>12</sup>, The Dictionary of Swedish National Biography<sup>13</sup>, and the National Biography of Finland<sup>14</sup> (NBF). There are also many "who is who" services online, and Wikipedia contains lots of short biographies with lots of data available in DBpedia<sup>15</sup> and Wikidata<sup>16</sup>.

Biographical collections can be used to study the underlying historical world. However, the texts, the language used, and the biographical collection as a whole can also be studied from a different, historiographical perspective as an artifact reflecting its own time, the editorial values and biases in selecting the biographees, the authors' perspectives, and also from linguistic points of view. Such analyses have been already made for some national dictionaries of biography, e.g., for the British ODNB [14], the Irish Ainm [15], Biography Portal of the Netherlands/BiographyNet [16], APIS in Austria [17], and the National Biography of Finland/BiographySampo [7, 8]. There are also related studies using, e.g., Wikipedia articles as the data source [18, 19].

Aside publishing biographical dictionaries in print and on the Web, representing and analyzing biographical data has grown into a new research and application field. In 2015, the first Biographical Data in Digital World workshop BD2015 was held presenting several works on studying and analyzing biographies as data [20], and the proceedings of BD2017 contain more similar works [21]. In [22], analytic visualizations were created based on U.S. Legislator registry data. The idea of biographical network analysis was developed in the Six Degrees of Francis Bacon system<sup>17</sup> [23, 24] that utilizes data of the Oxford Dictionary of National Biography. Network analyses and visualizations based on biographies have been presented also in [25, 10].

Extracting Linked Data [26] from texts [27] has been studied in several works, cf., e.g., [28, 29]. In [16] language technology was applied for extracting entities and relations in RDF using Dutch biographies in the BiographyNet<sup>18</sup>. This work was part of the larger NewsReader project<sup>19</sup> extracting data from news [30]. The problem of extracting (linked) data from biographical texts has also been studied when transforming the biographies of AcademySampo [31] and

---

<sup>7</sup><https://www.biographien.ac.at/>

<sup>8</sup><https://apis.acdh.oeaw.ac.at/>

<sup>9</sup><http://global.oup.com/oxforddnb/info/>

<sup>10</sup><http://www.anb.org/aboutanb.html>

<sup>11</sup>[http://www.ndb.badw-muenchen.de/ndb\\_aufgaben\\_e.htm](http://www.ndb.badw-muenchen.de/ndb_aufgaben_e.htm)

<sup>12</sup><http://www.biografischportaal.nl/en>

<sup>13</sup><https://sok.riksarkivet.se/Sbl/Start.aspx?lang=en>

<sup>14</sup><http://kansallisbiografia.fi> [13]

<sup>15</sup><https://www.dbpedia.org/>

<sup>16</sup><https://www.wikidata.org/>

<sup>17</sup><http://www.sixdegreesoffrancisbacon.com>

<sup>18</sup><http://www.biography.net/>

<sup>19</sup><http://www.newsreader-project.eu/>

BiographySampo [32] into LOD. BiographyNet focuses more on the challenges of natural language processing and managing the provenance information of data from multiple sources, while our focus in Sampo systems is on providing the end user with intelligent search and browsing facilities, enriched reading experience, and easy to use data-analytic tooling for biography and prosopography. The Austrian Prosopographical Information System (APIS) [17, 33, 34] is a virtual research environment that transforms text collections to machine readable formats and enables the use of natural language processing based methods to enrich the documents by extracting and linking information in them. The system has been used to transform and to study the collection of Austrian Biographical Dictionary 1815–1950 (ÖBL). Similarly to the Sampo systems, the APIS can be used to analyze and visualize datasets using for example network analysis methods.

### 3. Generations of Publishing Biographies

**Table 2**

Generations of Publishing Biographies

1. Generation	Engravings and printed texts
2. Generation	Biographies online for close reading
3. Generation	Biographies as data for data analysis and distant reading
4. Generation	Automatic knowledge discovery and AI

The idea of publishing biographies has evolved in generations [35] (cf. Table 2). First, life stories were published as texts engraved, e.g., in tomb stones in China and in rune stones in Scandinavia, and later as hand-written or printed texts (1. generation). Publishing biographies on the Web for close reading can be seen as the next 2. generation in the 90's. These systems can be referred to as dictionaries of biography on the Web. Here the biographies are provided for humans to read independently from place and time. Search engines are used for finding persons and texts of interest, and by browsing hypertext links additional recommended sources of information can be found. However, in 2. generation systems the data can be read only by the human user, and not by machines that only communicate the contents: the underlying data is not provided for computational analyses and application development. As a remedy, the data underlying the biographies can be published as a structured prosopographical database to be used in applications via APIs, e.g., for 2. generation dictionaries. In BiographySampo [7] the idea of publishing and using biographies as linked data was argued as a new paradigm change. Linked data can not only be used for data search and exploration as in 2. generation systems but also for data-analytic Digital Humanities (DH) research [36]. (Linked) data-based biographical publications can be seen as 3. generation systems. Arguably knowledge discovery, based on Artificial Intelligence, could be the next step ahead to the 4. generation of publishing biographies. Here the computer by itself is able to find new research questions in the data, solve them, and even explain the solutions to the human. First steps towards 4. generation systems in the case of BiographySampo are discussed in [35, 37].

Our work on biographical linked data started 2013 by designing a demonstrator [38] based on

the short biographies of the National Biography of Finland (NBF)<sup>20</sup>. The research hypotheses of this system was that “the reading experience can be enhanced by enriching the biographies with additional life time events, by providing the user with a spatio-temporal context for reading, and by linking the text to additional contents in related datasets”. The demonstrator was a 2. generation system although some map-based visualizations were created. The lives of the biographees were modeled as sequences of spatio-temporal events using linked data and CIDOC CRM<sup>21</sup> [39], an approach that has been used constantly in the later biographical Sampo systems.

The NBF demonstrator and its approach of using an event-based model for biographical linked data lead us to develop the system *WarSampo – Finnish World War II on the Semantic Web*<sup>22</sup> (online since 2015 with several new application perspectives published in 2016–2019) [4]. A key idea in WarSampo is to reassemble the life stories of the WW2 soldiers using data linking from different data sources. Biographical/prosopographical data was represented using Bio CRM [40], an extension of CIDOC CRM for biographical data. WarSampo took first steps towards 3. generation systems as it included some data analytic tools for, e.g., visualizing the casualties of troops and other prosopographical groups, such as officers, on a timeline. The system also presented various statistics pertaining to the fallen soldiers buried in the over 600 Finnish war cemeteries [41].

The idea of integrating data-analysis with semantic faceted search and browsing was developed further in the *Norssi Alumni* portal<sup>23</sup> (online since 2017) using a historical registry of ca 10 000 students of the prominent Finnish high school “Norssi” in 1867–1992. Here faceted search was used for filtering our groups of people and a number of statistical tools and visualizations could be applied to the result set. For example, most common later vocations, work places, or hobbies of the students in different times could be found and studied. The Norssi Alumni system was re-used and developed further in the *U.S. Congress Prosopographer* system<sup>24</sup> (online since 2018) [6] where a registry<sup>25</sup> of all U.S. Congress legislators from the 1st through 115th Congresses (1789–2018) was used. In this case, legislator data could be visualized of maps, and there were specific tabs for comparing statistics and map views of democratic and republican legislators.

The idea of 3. generation biographical systems and using biographies as linked (open) data was fully developed in the system *BiographySampo – Finnish Biographies on the Semantic Web*<sup>26</sup> (online since 2018) [42], a popular web service with tens of thousands of users. The system is based on mining out a large knowledge graph from the ca. 13 100 Finnish national biographies of the Finnish Literature Society, authored by some 940 scholars. The data is interlinked and enriched internally by 16 external data sources and by reasoning, e.g., by inferring family relations [31] and connections of interest between people and places [37]. In addition, a large linguistic knowledge graph of some 120 million triples of the biography texts was created and used for linguistic analyses about the biographies and their authors. For example, it was found that family-related words are widely used in biographies of female Members of Parliament but not for male.

---

<sup>20</sup><https://kansallisbiografia.fi/english/national-biography>

<sup>21</sup><https://www.cidoc-crm.org/>

<sup>22</sup>Project: <https://seco.cs.aalto.fi/projects/sotasampo/>; portal: <https://www.sotasampo.fi/>

<sup>23</sup>Portal: <http://www.norssit.fi/semweb/>

<sup>24</sup>Portal: <https://semanticcomputing.github.io/congress-legislators/>

<sup>25</sup><https://github.com/unitedstates/congress-legislators>

<sup>26</sup>Project: <https://seco.cs.aalto.fi/projects/biografiasampo/>; portal: <https://biografiasampo.fi/>

A set of data analyses from different perspectives of the BiographySampo dataset is presented in [8], using both the portal user interface and the underlying SPARQL service via the YASGUI editor<sup>27</sup> [43], Google Colab<sup>28</sup>, and Jupyter notebooks<sup>29</sup>. The analyses showed, for example, various statistics on charts, graphs and matrices on how the vocations of biographees change in time and correlate between parents and children, what places are mentioned in biographies and when on maps, and visualization of networks of biographees on how they relate to each other based on mentions in the biographies and on family and other relations.

One application perspective of the BiographySampo portal, based on relational search [37], can be seen as an example of a 4. generation system: Here the user first constrains freely people, professions, and places of interest to her/him using faceted search, and the system then finds “interesting” semantic connections between them and creates natural language explanations for the relations found. For example, when the user selects “Italy” from the place facet and “artist” in the profession faceted, one of the answers to the query is “Elin Danielson-Gambogi got the Florence City Award in 1899” based on an event extracted from her biography text and the place ontology telling that Florence is part of Italy.

After BiographySampo, the idea of integrating data publishing with data analysis was reused in the system *WarVictimsSampo 1914–1922* (online since 2019) [9] with data about the 41 500 victims and 1200 battles of the Finnish civil war and kindred wars. In this system, a new tool for interface design, Sampo-UI [3] was utilized. The application included an automatically generated animation on how the deaths in battles spread in Finland as time goes by in 1918.

The latest biographical Sampo system is *AcademySampo* (online since 2021) [31, 11], a biographical in-use LOD service and semantic portal based on 28 000 short biographies of all known Finnish academic people educated in Finland in 1640–1899. The system includes a rich set of data-analytic tools for DH research [10].

## 4. Sampo Model for Publishing Biographies

**Table 3**

Sampo Model principles for LOD publishing (P1–P3) and portal logic design (P4–P6)

P1	Support collaborative data creation and publishing
P2	Use a shared open ontology infrastructure
P3	Make clear distinction between the LOD service and the user interface (UI)
P4	Provide multiple perspectives to the same data
P5	Standardize portal usage by a simple filter-analyze two-step cycle
P6	Support data analysis and knowledge discovery in addition to data exploration

The work on biographical applications described above contributed to the development of the so called *Sampo Model* and the *Sampo Series* of semantic portals and LOD services<sup>30</sup> [1]. Based

<sup>27</sup><https://yasgui.triplay.cc>

<sup>28</sup><https://colab.research.google.com/notebooks/intro.ipynb>

<sup>29</sup><https://jupyter.org>

<sup>30</sup>See <https://seco.cs.aalto.fi/applications/sampo/> for a complete list of “Sampo portals”, videos, and further information.

on the six principles listed in Table 3, the model is a kind of consolidated approach for creating LOD services and semantic portals, something that the field of the Semantic Web is arguably still largely missing [44].

Principles P1–P3 can be seen as a foundation for developing LOD services; P4–P6 are related to creating semantic portals.<sup>31</sup> The model is based on the idea of collaborative content creation (P1) from multiple data sources. The data is aggregated from local data silos into a global service, based on a shared ontology [45] and publishing infrastructure (P2). The local data are harmonized and enriched with each other by linking and reasoning. In this model everybody can arguably win, including the data publishers by mutually enriched linked data and by re-using shared publishing infra, and the end users by richer global content and services. The model argues (P3) for the idea of separating the underlying Linked Data service *completely* from the user interface via a SPARQL API. This arguably simplifies the portal architecture and the data service can be opened for data analysis research and application development in Digital Humanities for everybody.

The general idea of principles P4–P6 is to “standardize” the UI logic so that the portals are easier to use for the end users and for the programmers to develop [3]. Principle P4 articulates the idea of providing different thematic *application perspectives* by re-using the data service. The application perspectives can be provided on the landing page of the Sampo portal system or be completely separate applications by third parties. According to P5 the perspectives can be used by a two-step cycle for research: First the focus of interest, the target group, is filtered out using faceted semantic search [46, 47, 48]. Second, the target group is visualized or analyzed by using ready-to-use data analytic tools of the application perspectives. At this point, it is also possible to select a particular member in the target group for a closer look and explore the data by browsing related links. Finally, the Sampo model aims not only at data publishing with search and data exploration [49] but also to data analysis and knowledge discovery with seamlessly integrated tooling for finding, analyzing, and even solving research problems in interactive ways (P6). The Sampo model principles are compatible with the FAIR principles for creating Findable, Accessible, Interoperable, and Re-usable data<sup>32</sup>, but were developed in the context of publishing and using LOD.

The Sampo model has been used in the biographical Sampo systems listed in Table 1. They make use of several data sources that enrich each other (P1). A set of shared ontologies have been used for harmonizing and amalgamating the datasets (P2). The aggregated data is then published as a separate LOD services including a SPARQL endpoint (3). On top of the data service different applications were created using only SPARQL for accessing the data. For example, the WarSampo portal initially contained six application perspectives and three more were added later. A Sampo system can also be re-used by external applications, such as other Sampo systems, leading to a kind of “Sampo cloud”. For example, the *WarMemoirSampo* system<sup>33</sup> [50] for publishing video memoirs of the Second World War veterans is based on and enriched by the WarSampo data infrastructure and portal pages.

---

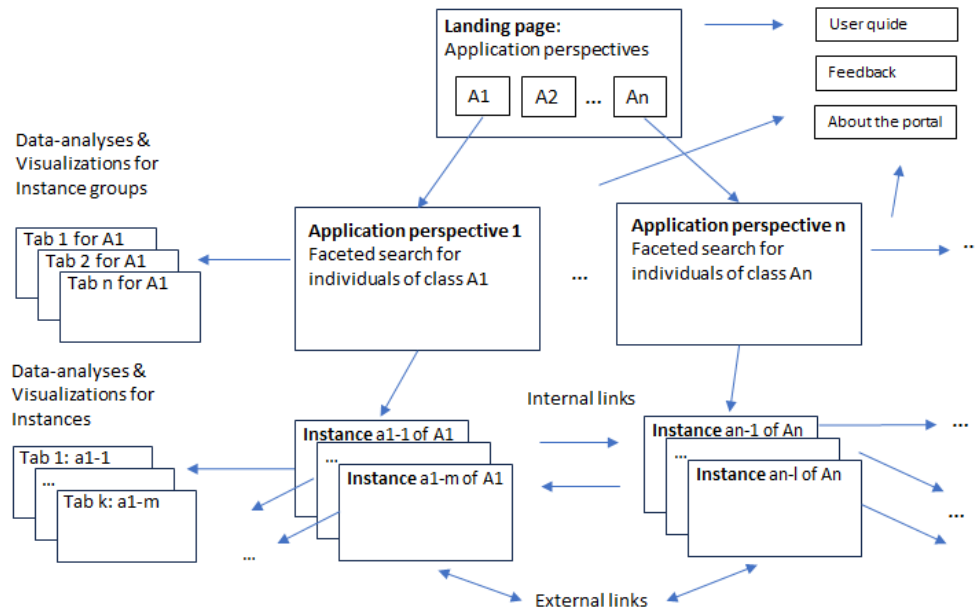
<sup>31</sup>The model is called “Sampo” according to the Finnish epic Kalevala, where Sampo is a mythical machine giving riches and fortune to its holder, a kind of ancient metaphor of technology according to the most common interpretation of the concept.

<sup>32</sup><https://www.go-fair.org/fair-principles/>

<sup>33</sup>Portal: <https://sotamuistot.arkisto.fi>; Project home: <https://seco.cs.aalto.fi/projects/war-memoirs/>



For publishing the LOD services, the Linked Data Finland service LDF.fi<sup>34</sup> [51] has been used. In LDF.fi, the 5-star model<sup>35</sup> of Tim Berners-Lee is extended to a 7-star model. The 6th star is given to a data publication if it includes not only the 5-star data but also the schemas of the data with documentation. This makes re-use of data easier. The 7th star is given to a data publication, if the publication includes some kind of evaluation that the data actually conforms to the provided schemas using, e.g., SHACL<sup>36</sup> or ShEx<sup>37</sup> [52]. The idea here is to encourage publishers to publish high quality LOD, which is a severe issue on the Semantic Web.



**Figure 1:** Navigational page structure of a Sampo portal based on Sampo-UI

The Sampo model principles P4–P6 are used for designing the portal user interfaces: the idea is to “standardize” the UI logic of Sampo portals to be created on top of a LD service SPARQL endpoint. The goal is to make the portals easier to use and implement. Fig. 1 illustrates the navigational structure of a Sampo portal. The user first lands on the *landing page* with several *application perspectives* to the data. The landing page that introduces the application perspectives as clickable boxes; by selecting one the corresponding perspective is opened. Each perspective typically provides a faceted search engine for filtering out a target group of individual instances of a class, such as people, places, or events. The search result can be visualized and studied on separate tabs. The default is to list results as a table, but the results can also be studied by data analytic tools on maps, using statistical charts, on timelines, or as networks for network analysis. Each individual in the system, say a person or a place, has a “home page” on which data related to

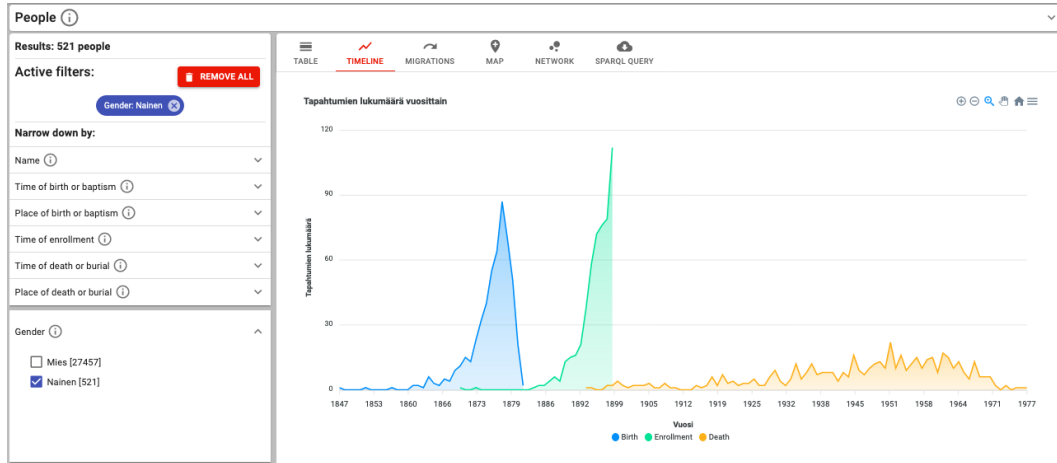
<sup>34</sup><https://ldf.fi>

<sup>35</sup><https://www.w3.org/community/webize/2014/01/17/what-is-5-star-linked-data/>

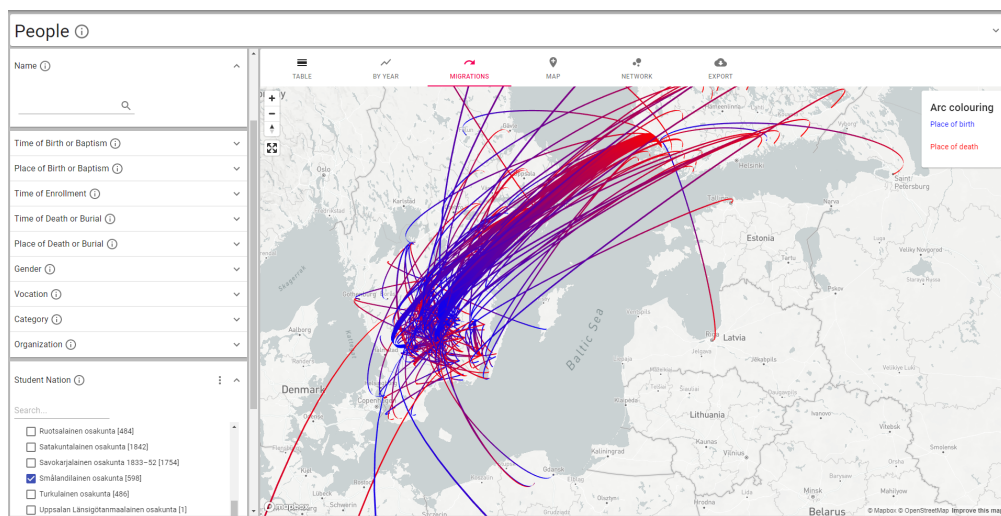
<sup>36</sup><https://www.w3.org/TR/shacl/>

<sup>37</sup><https://shex.io/>

it is automatically aggregated for providing a rich contextualized representation of the individual. Also on the home page, a set of tabs can be provided as data-analytic views of the individual. For example, for a person home page an egocentric network of related other persons can be shown or events in which the person participated in different roles can be visualized on a map or timeline.

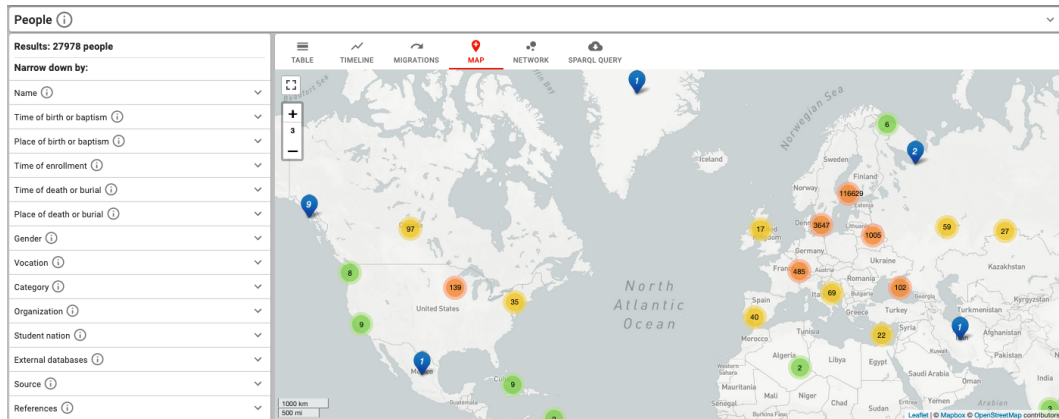


**Figure 2:** The annual births, enrollments, and deaths of female students on the TIMELINE tab. The target group has been filtered out by the facets shown on the left. Six alternative tabs for showing and analyzing the group can be selected on the top.



**Figure 3:** Using AcademySampo for prosopographical research by visualizing life charts of the members of the Student Nation of Småland as arcs from the place of birth (blue end of the arc) to the place of death (red end)

In practice, the UIs of the biographical Sampos have been implemented using the tools SPARQL



**Figure 4:** The MAP tab visualizes the lifetime places mentioned in about 175 000 events.

Faceter [2] (in WarSampo, Norssi Alumni, U.S. Congress Prosopographer, and BiographySampo) and since 2018 with Sampo-UI [3] (in WarVictimSampo and AcademySampo). For example, in the People application perspective of AcademySampo [10], prosopographical analysis tools are available in the TIMELINE, MIGRATION, MAP, and NETWORK tabs in addition to the default TABLE tab listing the search results. The TIMELINE tab shows the annual births, university enrollments, and deaths of the filtered people. For instance, Fig. 2 shows the charts for all 521 female university students in Finland in 1640–1899. The MIGRATION tab (Fig. 3) visualizes the mobility and immigration of students (597) of the Swedish Småland Student Nation with arcs depicting the life cycles. The blue end of the arc indicates the place of birth and the red the place of death, which is most often in the territory of present-day Finland, and the thickness of the arc reflects the number of people associated with the arc. If a person was born and died in the same place, the arc is not displayed. By clicking on the arc, one will find related links to people’s home pages. The MAP tab (Fig. 4) shows the approx. 3000 locations to which students are connected by approximately 175 000 events. For example, clicking on a marker in Ireland finds two related people, the other being the famous Johan Gadolin (1760–1855), who later discovered a new element, Yttrium. Finally, the NETWORK tab allows to explore the internal academic network of a group of people specified by the facet selections, for example, the teacher-student network of 1480 male students born in Helsinki.

## 5. Discussion

This paper presented a vision on how publishing and using biographies has evolved from engravings and written texts (1. generation) to Web-based publishing (2. generation), publishing biographies as Linked Open Data with seamlessly integrated data-analytic tools (3. generation), and finally to knowledge discovery-based systems (4. generation) based on AI. As an attempt to realize this vision, a series of biographical Sampo systems in use in Finland were created. These systems make use of the Sampo model, a set of principles for creating LOD services

and biographical portals on top of them. Our empirical experiences suggests that the model is feasible both from the end-user's and data publisher's points of view. More information about the applications as well and the underlying technical challenges and solutions can be found in the papers and web addresses given in relation to the applications.

During the work also several challenges of using linked data and the Sampo model have been encountered. Using explicit ontologies and linked data sets more demands on data quality than before. Any problems of data modelling or quality are highlighted in the user interfaces and data analyses. In most cases automatic annotation and linking had to be used for knowledge extraction in biographies, which lowers data quality, but on the other hand, manual annotations are costly and do not scale up. The Sampo model also requires more collaboration between the data publishers regarding interoperability, which complicates work. Integration of semantic portals with legacy systems can be a practical challenge in many organizations as well as sustainable maintenance of interlinked knowledge graphs [53]. From the end user, more source criticism<sup>38</sup> and understanding the characteristics and limitations of data are needed [54, 55]. However, the challenges in my mind seem to be smaller than the benefits and potential of utilizing linked open data in publishing and using biographical contents on the Web.

## Acknowledgements

Our work on biographies is partly supported by the EU project InTaVia: In/Tangible European Heritage<sup>39</sup>, and is related to the EU COST action Nexus Linguarum<sup>40</sup> on linguistic data science. Thanks to the Finnish Cultural Foundation for an Eminentia grant and to CSC – IT Center for Science for providing computational resources.

## References

- [1] E. Hyvönen, Digital humanities on the semantic web: Sampo model and portal series, *Semantic Web – Interoperability, Usability, Applicability* 14 (2023) 729–744. doi:10.3233/SW-223034.
- [2] M. Koho, E. Heino, E. Hyvönen, SPARQL Faceter – Client-side Faceted Search Based on SPARQL, in: *Joint Proc. of the 4th International Workshop on Linked Media and the 3rd Developers Hackshop, CEUR Workshop Proceedings, Vol. 1615, 2016*. URL: <http://ceur-ws.org/Vol-1615/semdevPaper5.pdf>.
- [3] E. Ikkala, E. Hyvönen, H. Rantala, M. Koho, Sampo-UI: A Full Stack JavaScript Framework for Developing Semantic Portal User Interfaces, *Semantic Web – Interoperability, Usability, Applicability* 13 (2022) 69–84. doi:10.3233/SW-210428.
- [4] E. Hyvönen, E. Heino, P. Leskinen, E. Ikkala, M. Koho, M. Tamper, J. Tuominen, E. Mäkelä, WarSampo data service and semantic portal for publishing linked open data about the Second World War history, in: H. Sack, E. Blomqvist, M. d'Aquin, C. Ghidini, S. P. Ponzetto,

---

<sup>38</sup><https://ranke2.uni.lu/define-dsc/#%20,%20Universit%C3%A9%20du%20Luxembourg>

<sup>39</sup><https://intavia.eu/>

<sup>40</sup><https://nexuslinguarum.eu/the-action>

- C. Lange (Eds.), *The Semantic Web – Latest Advances and New Domains (ESWC 2016)*, Springer, 2016, pp. 758–773. doi:10.1007/978-3-319-34129-3\_46.
- [5] E. Hyvönen, P. Leskinen, E. Heino, J. Tuominen, L. Sirola, Reassembling and enriching the life stories in printed biographical registers: Norssi high school alumni on the Semantic Web, in: *Proceedings, Language, Technology and Knowledge (LDK 2017)*, Springer, 2017, pp. 113–119. doi:10.1007/978-3-319-59888-8\_9.
- [6] G. Miyakita, P. Leskinen, E. Hyvönen, Using linked data for prosopographical research of historical persons: Case U.S. Congress Legislators, in: *Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection: 7th International Conference, EuroMed 2018, Nicosia, Cyprus, October 29–November 3, 2018, Proceedings. Part II*, volume 11197 LNCS, Springer International Publishing, 2018, pp. 150–162. doi:10.1007/978-3-030-01765-1\_18.
- [7] E. Hyvönen, P. Leskinen, M. Tamper, H. Rantala, E. Ikkala, J. Tuominen, K. Keravuori, BiographySampo – publishing and enriching biographies on the semantic web for digital humanities research, in: *The Semantic Web - 16th International Conference, ESWC 2019, Portorož, Slovenia, June 2–6, 2019, Proceedings*, volume 11503 LNCS, Springer International Publishing, 2019, pp. 574–589. doi:10.1007/978-3-030-21348-0\_{ }37.
- [8] M. Tamper, P. Leskinen, E. Hyvönen, R. Valjus, K. Keravuori, Analyzing biography collection historiographically as linked data: Case National Biography of Finland, *Semantic Web – Interoperability, Usability, Applicability 14 (2023)* 385–419. URL: <https://doi.org/10.3233/SW-222887>.
- [9] H. Rantala, I. Jokipii, E. Ikkala, E. Hyvönen, WarVictimSampo 1914–1922: a national war memorial on the semantic web for digital humanities research and applications, *ACM Journal on Computing and Cultural Heritage* 15 (2022). URL: <https://doi.org/10.1145/3477606>. doi:0.1145/3477606.
- [10] P. Leskinen, H. Rantala, E. Hyvönen, Analyzing the lives of finnish academic people 1640–1899 in Nordic and Baltic countries: AcademySampo data service and portal, in: *6th Digital Humanities in Nordic and Baltic Countries Conference, Proceedings, CEUR Workshop Proceedings, 2022*. URL: <https://seco.cs.aalto.fi/publications/2022/leskinen-et-al-academysampo-dhnb-2022.pdf>, forth-coming.
- [11] P. Leskinen, E. Hyvönen, Reconciling and using historical person registers as linked open data in the AcademySampo knowledge graph, in: *The Semantic Web – ISWC 2021. 20th International Semantic Web Conference, ISWC 2021, Proceedings*, Springer, 2021, pp. 714–730. doi:10.1007/978-3-030-88361-4\_42.
- [12] T. Keith, *Changing conceptions of National Biography*, Cambridge University Press, 2005. doi:10.1017/cbo9780511497582.
- [13] M. Klinge (Ed.), *Suomen kansallisbiografia 1–10*, Suomalaisen Kirjallisuuden Seura, Helsinki, Finland, 2003–2007.
- [14] C. N. Warren, Historiography’s two voices: Data infrastructure and history at scale in the oxford dictionary of national biography (ODNB), *Journal of Cultural Analytics* 1 (2018) 1–31. doi:10.22148/16.028.
- [15] Ú. Bhreathnach, C. Burke, J. M. Fhinn, G. Ó. Cleircín, B. Ó. Raghallaigh, A quantitative analysis of biographical data from Ainm, the Irish-language biographical database, in:

- BD2019 Biographical Data in a Digital World 2019. Proceedings of the Third Conference on Biographical Data in a Digital World 2019, volume 3152, CEUR Workshop Proceedings, 2019. URL: <http://ceur-ws.org/Vol-3152/>.
- [16] A. Fokkens, S. ter Braake, N. Ockeloën, P. Vossen, S. Legêne, G. Schreiber, V. de Boer, *BiographyNet: Extracting Relations Between People and Events*, New Academic Press, Berlin, Germany, 2017, pp. 193–224.
- [17] M. Schlögl, K. Lejtovicz, A prosopographical information system (APIS), in: Proceedings of the Second Conference on Biographical Data in a Digital World 2017 Linz, Austria, November 6-7, 2017, volume 2119, CEUR Workshop Proceedings, 2018. URL: <http://ceur-ws.org/Vol-2119/>.
- [18] A. Jatowt, D. Kawai, K. Tanaka, Time-focused analysis of connectivity and popularity of historical persons in Wikipedia, *International Journal on Digital Libraries* 20 (2019) 287–305. doi:10.1007/s00799-018-0231-4.
- [19] D. Metilli, V. Bartalesi, C. Meghini, A Wikidata-based tool for building and visualising narratives, *International Journal on Digital Libraries* 20 (2019) 417–432. doi:10.1007/s00799-019-00266-3.
- [20] S. ter Braake, A. Fokkens, R. Sluijter, T. Declerck, E. Wandl-Vogt (Eds.), *BD2015 Biographical Data in a Digital World 2015*, volume 1399, CEUR Workshop Proceedings, 2015. URL: <http://ceur-ws.org/Vol-1399/>.
- [21] A. Fokkens, S. ter Braake, R. Sluijter, P. Arthur, E. Wandl-Vogt (Eds.), *BD-2017 Biographical Data in a Digital World 2017*, volume 2119, CEUR Workshop Proceedings, 2017. URL: <http://ceur-ws.org/Vol-2119/>.
- [22] R. Larson, *Bringing lives to light: Biography in context. Final project report*, 2010. URL: [http://metadata.berkeley.edu/Biography\\_Final\\_Report.pdf](http://metadata.berkeley.edu/Biography_Final_Report.pdf), University of Berkeley.
- [23] C. Warren, D. Shore, J. Otis, L. Wang, M. Finegold, C. Shalizi, Six Degrees of Francis Bacon: A Statistical Method for Reconstructing Large Historical Social Networks, *Digital Humanities Quarterly* 10 (2016) 1–16.
- [24] A. Langmead, J. Otis, C. Warren, S. Weingart, L. Zilinski, Towards Interoperable Network Ontologies for the Digital Humanities, *International Journal of Humanities and Arts Computing* 10 (2016). doi:<http://dx.doi.org/10.3366/ijhac.2016.0157>.
- [25] M. Tamper, E. Hyvönen, P. Leskinen, Visualizing and analyzing networks of named entities in biographical dictionaries for Digital Humanities research, in: Proceedings of the 20th International Conference on Computational Linguistics and Intelligent Text Processing (CICling 2019), Springer, 2021. Preprint: <https://seco.cs.aalto.fi/publications/2021/tamper-et-al-cicling-2021.pdf>.
- [26] T. Heath, C. Bizer, *Linked Data: Evolving the Web into a Global Data Space* (1st edition), Morgan & Claypool, Palo Alto, California, 2011. doi:10.2200/S00334ED1V01Y201102WBE001.
- [27] J. L. Martinez-Rodriguez, A. Hogan, I. Lopez-Arevalo, Information extraction meets the semantic web: A survey, *Semantic Web – Interoperability, Usability, Applicability* 11 (2020) 255–335. doi:10.3233/SW-180333.
- [28] A. Gangemi, V. Presutti, D. R. Recupero, A. G. Nuzzolese, F. Draicchio, M. Mongiovì, Semantic web machine reading with FRED, *Semantic Web – Interoperability, Usability, Applicability* 8 (2017) 873–893. doi:10.3233/sw-160240.



- [29] M. C. Pattuelli, M. Miller, L. Lange, H. K. Thorsen, Linked Jazz 52nd Street: A LOD Crowdsourcing Tool to Reveal Connections among Jazz Artists., in: 8th Annual International Conference of the Alliance of Digital Humanities Organizations, DH 2013, Lincoln, NE, USA, July 16-19, 2013, Conference Abstracts, Alliance of Digital Humanities Organizations (ADHO), 2013, pp. 337–339.
- [30] M. Rospocher, M. van Erp, P. Vossen, A. Fokkens, I. Aldabe, G. Rigau, A. Soroa, T. Ploeger, T. Bogaard, Building event-centric knowledge graphs from news, *Web Semantics: Science, Services and Agents on the WWW* 37 (2016) 132–151. doi:10.2139/ssrn.3199233.
- [31] P. Leskinen, E. Hyvönen, Linked open data service about historical Finnish academic people in 1640–1899, in: DHN 2020 Digital Humanities in the Nordic Countries. Proceedings of the Digital Humanities in the Nordic Countries 5th Conference, volume 2612, CEUR Workshop Proceedings, 2020, pp. 284–292. URL: <http://ceur-ws.org/Vol-2612/short14.pdf>.
- [32] M. Tamper, P. Leskinen, K. Apajalahti, E. Hyvönen, Using Biographical Texts as Linked Data for Prosopographical Research and Applications, in: *Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection. 7th International Conference, EuroMed 2018, Nicosia, Cyprus*, Springer-Verlag, 2018, pp. 125–137. doi:10.1007/978-3-030-01762-0\_11.
- [33] Á. Z. Bernád, M. Kaiser, The biographical formula: Types and dimensions of biographical networks, in: *Proceedings of the Second Conference on Biographical Data in a Digital World 2017 Linz, Austria, November 6-7, 2017.*, volume 2119, CEUR Workshop Proceedings, 2018. URL: <http://ceur-ws.org/Vol-2119/>.
- [34] V. Gunter, S. Matthias, G. Vogeler, Data exchange in practice: Towards a prosopographical api (preprint), in: *Proceedings of the Third Conference on Biographical Data in a Digital World (BD 2019), Varna, Bulgaria, 2019.*
- [35] E. Hyvönen, Using the semantic web in digital humanities: Shift from data publishing to data-analysis and serendipitous knowledge discovery, *Semantic Web – Interoperability, Usability, Applicability* 11 (2020) 187–193. doi:10.3233/SW-190386.
- [36] E. Gardiner, R. G. Musto, *The Digital Humanities: A Primer for Students and Scholars*, Cambridge University Press, New York, NY, USA, 2015. doi:10.1017/CBO9781139003865.
- [37] E. Hyvönen, H. Rantala, Knowledge-based relational search in cultural heritage linked data, *Digital Scholarship in the Humanities (DSH)*, Oxford University Press 36 (2021) 55–64. doi:10.1093/llc/fqab042.
- [38] E. Hyvönen, M. Alonen, E. Ikkala, E. Mäkelä, Life stories as event-based linked data: Case semantic National Biography, in: *Proceedings of the ISWC 2014 Posters & Demonstrations Track, a track within the 13th International Semantic Web Conference (ISWC 2014) Riva del Garda, Italy, October 21, 2014.*, volume 1272, CEUR Workshop Proceedings, 2014, pp. 1–4. URL: [http://ceur-ws.org/Vol-1272/paper\\_5.pdf](http://ceur-ws.org/Vol-1272/paper_5.pdf).
- [39] M. Doerr, The CIDOC CRM – an ontological approach to semantic interoperability of metadata, *AI Magazine* 24 (2003) 75–92.
- [40] J. Tuominen, E. Hyvönen, P. Leskinen, Bio CRM: A Data Model for Representing Biographical Data for Prosopographical Research, in: *Proceedings of the Second Conference on Biographical Data in a Digital World 2017 Linz, Austria, November 6-7, 2017*, volume 2119, CEUR Workshop Proceedings, 2018. URL: <http://ceur-ws.org/Vol-2119/>.



- [41] E. Ikkala, M. Koho, E. Heino, P. Leskinen, E. Hyvönen, T. Ahoranta, Prosopographical views to finnish ww2 casualties through cemeteries and linked open data, in: Proceedings of the Workshop on Humanities in the Semantic Web (WHiSe II), volume 2014, CEUR Workshop Proceedings, 2017. URL: <http://ceur-ws.org/Vol-2014/>.
- [42] E. Hyvönen, P. Leskinen, M. Tamper, H. Rantala, E. Ikkala, J. Tuominen, K. Keravuori, BiographySampo – Publishing and enriching biographies on the Semantic Web for digital humanities research, in: The Semantic Web. 16th International Conference, ESWC 2019, Springer, 2019, pp. 574–589. doi:10.1007/978-3-030-21348-0\\_37.
- [43] L. Rietveld, R. Hoekstra, The YASGUI family of SPARQL clients, Semantic Web – Interoperability, Usability, Applicability 8 (2017) 373–383. doi:10.3233/SW-150197.
- [44] P. Hitzler, A review of the semantic web field, Commun. ACM 64 (2021) 76–83. doi:10.1145/3397512.
- [45] S. Staab, R. Studer (Eds.), Handbook on Ontologies (2nd Edition), Springer, 2009.
- [46] E. Hyvönen, S. Saarela, K. Viljanen, Application of ontology-based techniques to view-based semantic search and browsing, in: Proceedings of the First European Semantic Web Symposium, Springer, 2004. doi:10.1007/978-3-540-25956-5\\_7.
- [47] D. Tunkelang, Faceted search, Morgan & Claypool, Palo Alto, California, 2009. doi:10.2200/S00190ED1V01Y200904ICR005.
- [48] Y. Tzitzikas, N. Manolis, P. Papadakos, Faceted exploration of RDF/S datasets: a survey, Journal of Intelligent Information Systems 48 (2017) 329–364. doi:10.1007/s10844-016-0413-8.
- [49] G. Marchionini, Exploratory search: from finding to understanding, Communications of the ACM 49 (2006) 41–46. doi:10.1145/1121949.1121979.
- [50] E. Hyvönen, E. Ikkala, M. Koho, R. Leal, H. Rantala, M. Tamper, How to search and contextualize scenes inside videos for enriched watching experience: Case stories of the second world war veterans, in: Proceedings of the 19th Extended Semantic Web Conference (ESWC 2022), Poster and Demo Papers, 2022. URL: <https://seco.cs.aalto.fi/publications/2022/hyvonen-et-al-wms-2022.pdf>, forth-coming.
- [51] E. Hyvönen, J. Tuominen, M. Alonen, E. Mäkelä, Linked Data Finland: A 7-star model and platform for publishing and re-using linked datasets, in: The Semantic Web: ESWC 2014 Satellite Events, Springer, 2014, pp. 226–230. doi:10.1007/978-3-319-11955-7\\_24.
- [52] J. E. Labra Gayo, E. Prud'hommeaux, I. Boneva, D. Kontokostas, Validating RDF Data, volume 7 of *Synthesis Lectures on the Semantic Web: Theory and Technology*, Morgan & Claypool Publishers LLC, 2017. URL: <https://doi.org/10.2200/s00786ed1v01y201707wbe016>. doi:10.2200/s00786ed1v01y201707wbe016.
- [53] C. Gutierrez, J. F. Sequeda, Knowledge graphs, Communications of the ACM 64 (2021) 96–104. doi:10.1145/3418294.
- [54] T. Koltay, Data literacy for researchers and data librarians, Journal of Librarianship and Information Science 49 (2015) 3–14. doi:10.1177/0961000615616450.
- [55] E. Mäkelä, K. Lagus, L. Lahti, T. Säily, M. Tolonen, M. Hämäläinen, S. Kaislaniemi, T. Nevalainen, Wrangling with non-standard data, in: Proceedings of the Digital Humanities in the Nordic Countries 5th Conference, CEUR Workshop Proceedings, CEUR-WS.org, Germany, 2020, pp. 81–96.

# Biographical Research and Digital Mapping

Paul Longley Arthur<sup>1</sup> and Isabel Smith<sup>1</sup>

<sup>1</sup> *Edith Cowan University, Western Australia*

## Abstract

This paper provides a mid-project report on an Australian Research Council-funded project that focuses on the development of digital tools to map biographical data relating to the history and legacies of British slavery in Australia from the 1830s onward. Through developing innovative methods for biographical research and digital mapping, this data-intensive project is tracing the movement of capital, people, and culture from slave-owning Britain to Western Australia. The paper will begin with an introduction to the Western Australian Legacies of British Slavery (WALBS) project, and then the Time-Layered Cultural Map (TLCMap), which has been used to visualise the project's data, before explaining how these two projects have come together. It will then discuss some of the major questions and tensions that have emerged out of gathering and visualising this data so far. The aims of the paper are to underscore the potential for data collection and analysis specific to this research such as mapping biographies and slavery networks, to demonstrate the benefits of TLCMap for humanities researchers, and to highlight key challenges we encountered with this data including questions around how to navigate bias and power tied up within data, and how to map lives.

## Keywords

Digital mapping, biography, slavery

## 1. Background: The Western Australian Legacies of British Slavery project and the Time-Layered Cultural Map

The WALBS project is tracing the movement of people, property, capital and culture from Britain to Western Australia, exploring the links between slavery in the British Empire and settler colonialism. In particular, research has focused on detailed biographical investigations into a series of individuals connected to slavery who were some of the earliest colonists in Western Australia. This research has grown out of the University College London's Legacies of British Slavery (or LBS) project, which started by tracing the 20 million pounds in compensation money paid out to British slavers following the abolition of slavery in 1833. By examining the records of claims made for this compensation money, the LBS project identified around 46,000 claimants, as well as thousands of plantations and estates. This data is publicly available on the LBS online database. Users can search the database to find information including the names, professions, activities and affiliations of men and women who were slave-owners, attorneys, mortgagees and legatees, as well as the names and locations of plantations and estates. The WALBS project builds on this research by tracing individuals who moved to the settler colonies in Australia. Research so far has revealed some striking patterns, including the migration of a number of British slavers and slave-trade beneficiaries from select Caribbean estates to Western Australia in the early 1830s. Some of the key data being gathered in the project include the names, professions, family relations and business associates of individuals, their capital, and connections to

---

Proceedings of the First Conference on Metaverse Research (ICMR 2023), July 14, 2023, Kuala Lumpur, Malaysia



© 2023 Copyright for this paper by its authors.  
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

[https://doi.org/10.3986/9789610508120\\_05](https://doi.org/10.3986/9789610508120_05)

particular estates. It also traces individuals' movements, activities, and influence in Western Australia, including amounts and locations of land occupied, and affiliations with particular businesses, institutions, and government.

This data – looking at the movement of people, the distribution of networks, and land – lends itself to mapping and geographic visualisations. In 2021, the WALBS project team began talks with lead investigators and creators of the TLCMap based at the University of Newcastle in Australia. The TLCMap offered a useful tool in visualising and analysing the lives and movements of individuals connected to slavery who migrated to Western Australia. The TLCMap is a set of online tools that allows humanities researchers to compile historical and cultural data using spatio-temporal coordinates. It is not a singular and literal map per se, but a range of accessible software – or 'software ecosystem' – that allows researchers with minimal programming skills to upload, gather, analyse and visualise data themselves.

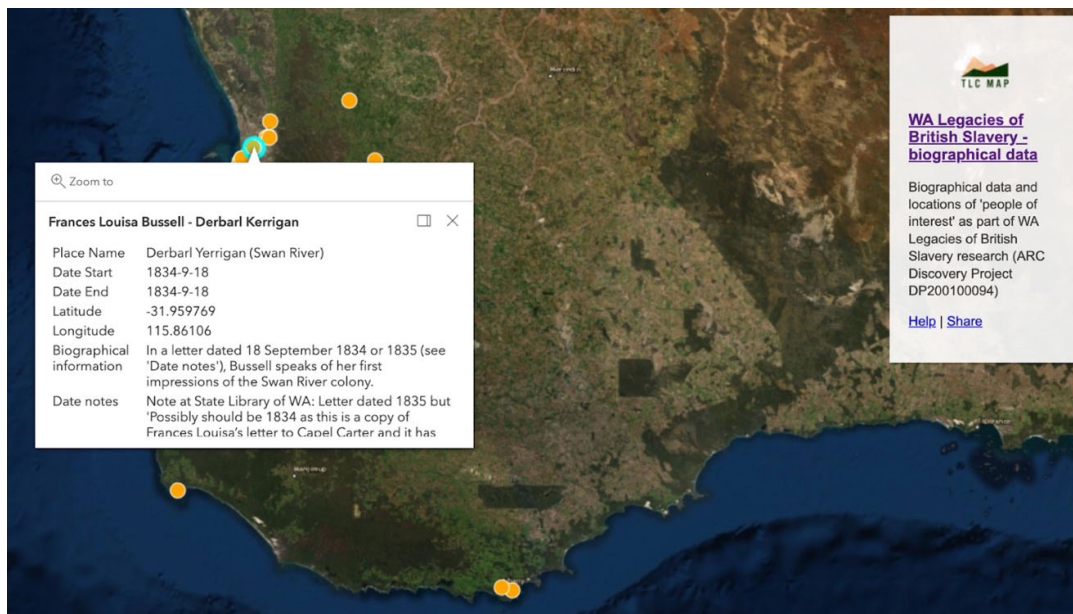


**Figure 1:** The user interface of the Time-Layered Cultural Map software, displaying the entry point for the Western Australian British Legacies of Slavery biographical data.

It was decided that the best TLCMap tool to employ for the WALBS project was the Gazetteer of Historical Australian Places. The Gazetteer allows researchers to create layers on a 3D map which pinpoint locations with accompanying information. This tool allowed us to plot out the journeys of slavers and slave-trade beneficiaries as they moved across Britain, the Caribbean and Western Australia. Figure 1 shows the entry point for users navigating this data. For our trial data we selected six individuals, based on those with the most available data and digital assets, and also those that offered a range in gender, geographical destinations and biographical events. For each individual we created a separate layer, uploaded as CSV files converted from Excel spreadsheets. Data on each individual included latitude and longitude for each geographical place they visited or lived in, place names, short biographical summary text, and links to data stored on other databases and archives (see Figure 2 for a selection of these fields). In the 3D map this translated to a staged journey for each individual, represented by a series of points on the map accompanied by brief narrative text regarding time spent at that location, links to images and further resources (Figure 3 gives an example of one of these points of an individual's journey).

Title	Placename	Latitude	Longitude	Location notes	Date	Date notes	Biographical information	Images
Frances Louisa Bussell - Portsmouth	Portsmouth, England	50.801389	-1.109861	Selected location near the Portsmouth Historic Dockyard, though have not pinpointed the exact dock that the James Pattison departed from.	10/2/1834	We do not have a record of the ship departure date, however in a letter dated 5 May 1834 Bussell refers to 'those dear hearts who have thought and prayed for us in all their positions for the last 84 days and nights', suggesting that she departed England on 10 February. [5]	Frances Louisa Bussell (née Yates) was born in England in 1782. She married Reverend William Marchant Bussell and had nine children: John Garrett, Charles, Joseph Vernon, Alfred Pickmore, Lenox, Frances Louisa (Fanny), Elizabeth (Bessie) Capel and Mary. Reverend William Marchant Bussell died in 1820. The following decade Frances Louisa and seven of her children left England and became prominent colonists in the southwest of WA. [13]  Bussell left for WA in early 1834 with her eldest daughter Mary. [6] In a letter written to her sisters while anchored at Table Bay, Cape Town, on 5 May, she refers to 'those dear hearts who have thought and prayed for us in all their positions for the last 84 days and nights', suggesting that she departed England on 10 February. She said of the journey: 'We are in health and spirits and have heard favourable accounts of our dear Augustine. The voyage has been tedious for the impatient but remarkably propitious. Not one passenger would we wish away we have had no storms and very few annoyances or "fikes" as some of our scotch friends term.' [5]	
Frances Louisa Bussell - Kincinnup	Kincinnup (Albany)	-35.048278	117.968028		19/6/1834		Bussell and her daughter Mary arrived at King George Sound in Kincinnup (Albany) aboard the James Pattison on 19 June 1834. [7] They soon sailed on to Walyalup (Fremantle). Sons John Garrett, Charles, Joseph Vernon, Alfred Pickmore had arrived in 1830, and Lenox, Fanny and Bessie in 1833.	
Frances Louisa Bussell - Walyalup	Walyalup (Fremantle)	-32.056861	115.741389		19/8/1834		After delays due to bad weather, the James Pattison arrived at Walyalup (Fremantle) on Tuesday 19 August 1834. [8, 9, 10] In a letter to Elizabeth Capel Carter dated 27 August, Bussell writes: 'Perplexed and worried as I am in landing and collecting my property and the ships on the point of [sinking/sailing?] I snatch a moment to say we arrived safe and well. We are now at Fremantle with the McDermots and were to have been conveyed to Perth to day to stay with Lady Stirling till the arrival of the Colonial Schooner which is to embark us for Augusta.' [11]	Newspaper notice advising of delays to James Pattison: <a href="https://trove.nla.gov.au/newspaper/article/6413937?searchTerm=%22james%20pattison%22">https://trove.nla.gov.au/newspaper/article/6413937?searchTerm=%22james%20pattison%22</a>  Newspaper notice advising of James Pattison arrivals: <a href="https://trove.nla.gov.au/newspaper/article/6413547?searchTerm=%22james%20pattison%22">https://trove.nla.gov.au/newspaper/article/6413547?searchTerm=%22james%20pattison%22</a>

**Figure 2:** For each individual of interest we created a spreadsheet, with a new row for each new location, detailing coordinates of the location, place name, date ranges, narrative biographical text, and links to assets such as images. Additional fields of data are discussed later in the paper.



**Figure 3:** The user interface of WALBS biographical data once uploaded to TLCMap software, the Gazetteer of Historical Australian Places.

For the WALBS team, the TLCMap offered a number of advantages. Embedding links to records in other collections and databases allowed us to point to entries in the LBS database. It also offered unique opportunities for visualisations and analysis of patterns. While other databases – such as People Australia, created by the National Centre of Biography at the Australian National University – enable network analysis to reveal family, business and other connections between individuals, the TLCMap is

the only purpose-designed research tool in Australia to visually represent these connections, allowing researchers to identify geographical clusters and parallel journeys by sight.

## 2. Mapping biography: Maps as life stories

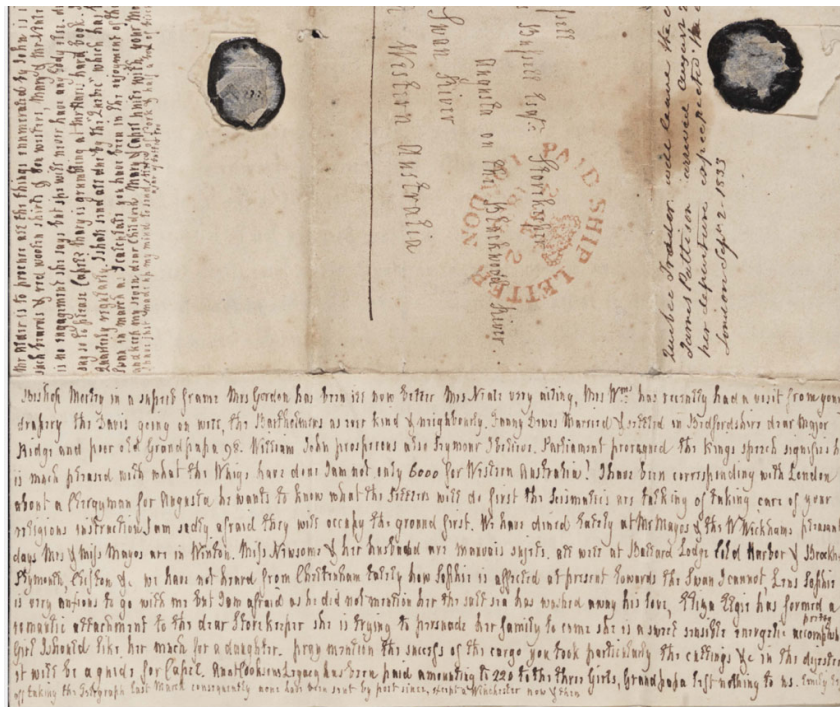
Plotting out these geographical and temporal journeys based on biographical details constructs a sort of visual life story, representing the movements, encounters, and places called home by an individual over time. Maps are indeed forms of storytelling. Robert T. Tally contends that ‘In many ways, telling a story is like drawing a map, and vice-versa’ (Tally 2016, 26). The field of literary cartography explores this relationship, examining the visual and spatial potential of literature to function as a figurative map. In his analysis of contemporary American fiction, Fabrizio Di Pasquale states that ‘every novel is a map, every map tells a story, and every story is connected to a territory.’ (Di Pasquale 2016, 47). Similarly, John McCrystal writes that

*The earliest cartographic technology is story. Ancient stories contained useful information, such as the relation of changes in the night sky to the rhythms of nature; they also contained navigational and cartographic information.* (McCrystal 2019, 11)

Mapping can offer a methodology in tracing and recording an individual’s life story. Writing about ‘Aboriginal mapping’ in Australia, Justin Butler describes a process of asking Aboriginal people a series of questions about their identity, including questions regarding connections to family and land (Butler 2017). For Butler, ‘mapping’ does not refer to the creation of literal maps per se, but is a practice concerned with narrative therapy and storytelling, which places at its heart the individual and custodian of that life story.

There are many challenges and complexities in mapping a life. This begins with choices around what information is presented and what is left out. In selecting biographical data to map out the WALBS people of interest, we were guided in part by parameters regarding the functionality of the TLCMap software. In particular, data was structured around geographical and temporal coordinates. Due to the research theme, another mandatory field of data was ‘links to slavery’, where brief narrative text outlined the particular family or business connections an individual held with the slave trade. However, as individuals and journeys dated back approximately 200 years, furnishing their stories with rich biographical data was not always possible. Often, records were limited and the issue of missing or ‘fuzzy’ data was a recurrent one. At times we could identify accurate spatial coordinates, for example by looking at ship records to locate particular ports of departures and arrivals, or identifying specific estates or plantations. But much of the time this information was less precise – for example an individual’s diary entry referring to a broad region, or a former place name that no longer exists. (Figure 4 shows an excerpt of a letter written by one of the individuals selected) In these instances we made estimates, and included fields for notes – ‘Date notes’ and ‘Location notes’ – where we could explain the limitations of our data and our process for devising estimations. An example of this is the journey of James Walcott from Western Australia to Mauritius in 1837 – while a record was obtained indicating Walcott’s date of departure, no date of arrival was available. An estimated time of one month was calculated based on a proposal written in 1826 by James Stirling, in which Stirling estimates three weeks’ travel from the Swan River Colony to Mauritius (Project Gutenberg Australia 2014). This calculation is detailed under the ‘Date notes’ field.





**Figure 4:** Frances Louisa Bussell’s letter to Frances and Charles Bussell, dated August 30, 1833. Letters written by the WALBS individuals of interest were rich in detail but still often made only vague reference to key data such as locations and dates of events. Courtesy State Library of Western Australia.

### 3. Mapping influence: Deconstructing power and bias in data

Tracing the lives of these early colonists in Australia, tensions emerged around power and bias tied up within the data. Olaf Berg reminds us that data are never found but always created, through many moments of complex interpretation and decision-making (Berg 2020). The content and storage of data reflects relations of power at the times of it being recorded as well as retrieved. In the WALBS research, we noticed the influence of colonial authorities and frameworks underlying our data. Frequently, the names, lives and experiences of Aboriginal people, migrants, and people of colour were obscured or written out of colonial records. Documents such as diaries, paintings, and indeed maps, presented narratives that glorified colonial exploits and laid European claims over lands. For example, when researching the life of Frances Louisa Bussell – an early Western Australian colonist and cousin to Thomas Legal Yates, one of Jamaica’s biggest slave owners – a book detailing the family history downplayed the events of a significant massacre led in part by one of the Bussell sons (Shann 1926). Similarly, surveying maps that outline which land was allocated to specific colonists, land is described as being ‘granted’ to Europeans when in fact it was never granted by Aboriginal people but only by colonial authorities. In using and presenting these sources to map out biographical data, we face the potential of inadvertently reconstructing colonial narratives. Thomas J. McGurk and Sébastien Caquard underscore the potential for mapmaking to serve as a ‘tool of dispossession’ (McGurk and Caquard 2020, 52). They explain that through colonisation, the primary function of maps quickly shifted from a means of communication and passing on of cultural knowledge, to claiming and carving up territories, thus framing land and resources in economic terms. When creating new maps based on data regarding powerful colonial figures, we have been mindful of these legacies. This is especially the case when drawing upon colonial maps themselves.

On the other hand, displaying these materials allows them to be interrogated and deconstructed. The narrative text fields were useful for these commentaries, for instance offering spaces where it could be explained that land was not in fact 'granted' by Indigenous people to colonists. Closely examining the biographies and narratives of the WALBS people of interest also offered opportunities to deconstruct the relations of power and discourses they embody. Explicitly mapping out the movements and activities of colonists and slavers - detailing their connections to particular institutions, businesses, and one another, as well as identifying hives of activity - visually represents their networks and spheres of influence. Catherine Hall, Principle Investigator on the LBS project, stresses the significance in interrogating slavers' and colonists' creations of narratives around empire and 'race', and frameworks of racial thinking and racial logics that construct 'otherness', to explore racisms and cultures that are still operating today (Hall 2018). We therefore targeted data such as first-hand comments regarding race, class or labour, as well as contextual biographical information that could shed light on individuals' ideological frameworks. For instance, we identified a report written by James Stirling - Western Australia's first governor, with ties to the slave trade - in which he describes Aboriginal people under the heading of 'Animal Productions', as well as a description in Stirling's biography that refers to his 'enviable Scottish pedigree and enormous family pride.' (Statham-Drew 2003, 1)

Despite the colonial legacies of maps, many have pointed to the decolonising potential of mapmaking. In her analysis of Sami activist and artist Hans Ragnar Mathisen, Maria Therese Stephansen illustrates how Mathisen's maps and mapmaking practices construct counternarratives that reinstate cultural knowledge of the Sami people - Indigenous people of the Sápmi region in northern Europe - in the face of official histories from Norway, Finland and Sweden (Stephansen 2017). She refers to Mathisen's well-known map of Sápmi, produced in 1975, which presents the region through the Sami people's cultural and geographical terms. This includes the use of Indigenous language - it features 920 Sami place names. The mobilisation of language can be a key means by which maps challenge colonial frameworks. Reflecting on the development of TLCMap for a project mapping out colonial frontier massacres in Australia from 1788 to 1930 (the 'Massacre Map'), System Architect Bill Pascoe refers to questions around the naming of particular sites and massacres that have both Aboriginal and non-Aboriginal names. Pascoe also highlights the choice of words such as 'perpetrators', 'murderers', 'attackers', 'killers' and so on, underscoring the politicised nature of particular terms (Pascoe 2022). In terms of data used in the WALBS project, similar questions are faced. When locating individuals at particular geographical locations, we use Aboriginal places names where possible, with non-Aboriginal names in brackets. Similarly, the narrative biographical text describes individuals of interest as 'colonists' and 'slavers' rather than 'settlers', and chronicles their acts using terms such as 'colonised' rather than 'discovered'.

The use of imagery and aesthetics also offers a powerful means by which to navigate and challenge colonial bias. Pascoe explains that rather than presenting a warning to Aboriginal audiences about photographs of deceased people, the Massacre Map project team chose not to include these images at all. He also speaks of considerations around the overall 'tone', for example which colour to make the dots placed on the map (Pascoe 2022). After initially selecting the colour red, Pascoe explains that to some this appeared too sensationalist and even violent, while others argued that this was in fact apt. At the same time, certain Indigenous communities view and use red as a colour of celebration, making it seem an inappropriate choice (the colour yellow was ultimately chosen). Imagery and tone in Mathisen's work is a key feature. His 1975 Sápmi map is detailed with references to Sami mythology and culture including a sun representing the gods, a shaman drum, a milk bowl (nappi), and old rock carvings, as well as traditional Sami designs and patterns (Stephansen 2017). When surveying available digitised images to illustrate the biographies of the WALBS individuals of interest, the majority were records such as early colonial paintings, maps, and photographs of colonists. Taken as a whole, and accompanying the biographies of slavers and slavery beneficiaries, these created a traditionally



European aesthetic reminiscent of early colonial histories. In an attempt to counter the weight of these archival materials, the WALBS team investigated the use of alternative images, for instance contemporary Aboriginal artworks. The work of Chris Pease, an Australian artist of Indigenous and European heritage, takes the traditional aesthetics of early European landscape paintings of Australia and superimposes imagery such as bullseyes, or land allocations created by white authorities, to underscore the impacts of colonisation on Indigenous people and land (see Figure 5). Juxtaposing biographical materials obtained from colonial and state archives with such contemporary imagery and narratives highlights the power, politics and omissions in archival data.



**Figure 5:** Chris Pease, *Land Release 3*, 2008. Contemporary artworks such as this might be used to provide counternarratives to early colonial paintings, maps and state archival materials.

#### 4. Mapping digital lives: Embracing process and non-linearity

When navigating the complexities of mapping lives, endeavouring to present rich biographical data as well as conflicting narratives of the past, one approach might be to present multiple layers of data rather than singular journeys, fixed points, or static maps. This aligns with an increasing emphasis in cartography in thinking of mapping as a process rather than of maps as final objects or outcomes. Some have suggested this process-oriented approach is more common in Indigenous communities. Referring to the work of Gwilym Eades, McGurk and Caquard explain:

*According to Eades (2015) ... while Indigenous persons were “mapping,” which is a process-oriented activity indicative of an Indigenous way of seeing, Europeans were “mapmaking,” which is an object-focused activity that reflects a Eurocentric worldview. (McGurk and Caquard 2020, 52)*

McGurk and Caquard also draw upon Margaret Wickens Pearce’s assertion that by viewing maps as processes, this allows for ‘more experimental forms of mapping that can better mobilize the strengths of oral and performative formats as a means of transmission of Indigenous knowledge.’ (McGurk and Caquard 2020, 52). Tristan Schultz, designer and Gamilaraay man with both Aboriginal and European Australian heritage, argues for a decolonising method that focuses ‘on experimentation, creative insight, iteration and reflection of *how* mapping with people in situated contexts can occur, rather than *what* has been articulated.’ Schultz points to the value in creating and recording conversation, narrative, workshops, and encounters (Schultz 2019, 2).

Finding ways to present the processes underlying mapmaking demands a rethinking of what maps look like and how they function. Sébastien Caquard and William Cartwright explain that ‘Telling the

story about how maps are created and how they come to life in a broad social context and in the hands of their users has become a new challenge for mapmakers' (Caquard and Cartwright 2014, 101). Digital formats might be particularly well-suited to capturing these iterative processes by displaying layers of multiple, fluid, and diverse data. In their analysis of Indigenous web-mapping sites in Canada, McGurk and Caquard argue that online mapping offers significant potential in underscoring the complexities of place names and how they come to be, by allowing for the presentation of diverse multimedia content accompanying maps (McGurk and Caquard 2020). They cite Pearce, who highlights the ways that digital technologies allow for the linking of place names with their oral origins by embedding audio and visual files, thus contextualising places and names within broader historical and cultural stories (Pearce 2008, cited in McGurk and Caquard 2020). That is, digital modes might be particularly useful in presenting not only data such as geographical and temporal coordinates, place names, and so on, but data that contextualises how this geotemporal data was obtained and constructed. In terms of the WALBS data, an instance where this contextualisation might be useful is the contested location of the 'Peel settlement'. This is where one WALBS person of interest – Adam Wallace Elmslie – was based as agent to one of Western Australia's early colonists, Thomas Peel. Archaeologist Shane Burke has located the Peel settlement at a different location from historians Pamela Statham-Drew and Ruth Marchant James (Burke, Di Marcho and Meath 2010). This debate has been included in the narrative text, but digital content might also allow for layers of alternate locations or links to oral and written discussions around the debate. Digital layers of content could also be used to underscore the complexities of biographical data and narratives. For example, the life and history of Stirling has been the subject of debate due to Stirling's leading of the notorious Pinjarra Massacre. This debate intensified recently with calls to rename the city of Stirling (which were subsequently rejected). Within TLCMap, Stirling's biographical entries might include links to discussions around this history, including radio episodes and print media articles (ABC News 2021).

The fluidity of digital content can also enable non-linear approaches to the presentation of data. Though it remains in a theoretical stage, we have looked at organising data in non-hierarchised formats. In this conception, rather than data appearing through a series of stages in which users begin by clicking on place names which then lead to more layers or hierarchies of information, data would instead all appear at once. For example, when exploring an individual's mapped life story, users would be able to dive in through any piece of data, not only geographical points: all place names, images, links, and pieces of text would be given equal prominence. Pascoe describes this democratising approach using the metaphor of a tree: rather than presenting data in a structured hierarchy akin to branches and trunks, presenting it like a collection of leaves. Future research could experiment with how this might look in practice. The next step in our project is to extend this prototype to include a wider selection of individuals, incorporating further findings from the WALBS project, which will be ongoing until the end of 2023.

## **5. Acknowledgements**

We acknowledge support for this research from Australian Research Council grants DP200100094 Western Australian Legacies of British Slavery, and LE230100079 Time Layered Cultural Map of Australia: Advanced Techniques and Big Data.

## 6. References

- [1] G. Arnott, Commemorating James Stirling? ABC News, 2021. <https://www.abc.net.au/radionational/programs/the-history-listen/commemorating-james-stirling/13642650>.
- [2] O. Berg, Capturing displaced persons' agency by modelling their life events: A mixed method digital humanities approach. *Historical Social Research* 45. 4 (2020) 263-89. doi: 10.12759/hsr.45.2020.4.263-289.
- [3] S. Burke, P. Di Marco and S. Meath, The land 'flow[ing] ... with milk and honey': Cultural landscape changes at Peel Town, Western Australia, 1829-1830. *Australasian Historical Archaeology* 28 (2010) 5-12.
- [4] J. Butler, Who's your mob? Aboriginal mapping: Beginning with the strong story. *The International Journal of Narrative Therapy and Community Work* 3 (2017) 22-26.
- [5] Fabrizio Di Pasquale, Cartography and the contemporary American novel. Nic Pizzolatto: An Example of Geocritical Analysis, in: E. Peraldo (Ed.), *Literature and Geography: The Writing of Space Throughout History*, Cambridge Scholars Publishing, Newcastle Upon Tyne, 2016, pp. 37-49.
- [6] C. Hall, 2018. Doing reparatory history: bringing 'race' and slavery home. *Race & Class* 60. 1 (2018) 3-21. doi: 10.1177/0306396818769791.
- [7] J. McCrystal, *Singing the trail: The story of mapping Aotearoa New Zealand*. Allen & Unwin, Sydney, 2019.
- [8] T. J. McGurk and S. Caquard. 2020. To what extent can online mapping be decolonial? A journey throughout Indigenous cartography in Canada. *The Canadian Geographer* 64. 1 (2020) 49-64. doi: 10.1111/cag.12602.
- [9] A. R. Novaes, 2020. Mapping Cross-Cultural Exchange: Jaime Cortesão's Dialogues and Documents on the Role of Indigenous Knowledge in Brazilian Exploration, in: B. Schelhaas, F. Ferretti, A. R. Novaes and M. Schmidt di Friedberg (Eds.), *Decolonising and Internationalising Geography: Essays in the History of Contested Science*, Springer, Cham, 2020, pp. 1-16.
- [10] B. Pascoe, *The Massacre Map, Space, Data, Place: Digital Tools for Australia's Deep Past*, 2022.
- [11] M. W. Pearce, Framing the days: Place and narrative in cartography. *Cartography and Geographic Information Science* 35. 1 (2008) 17-32.
- [12] Project Gutenberg Australia. *Official Papers Relating to the Settlement at Swan River West Australia. December, 1826-January, 1830, 2014.* <https://gutenberg.net.au/ebooks14/1402751h.html>.
- [13] T. Schultz. *Decolonising Design: Mapping Futures*, PhD thesis, Queensland College of Art, 2019. doi: 10.25904/1912/2253.
- [14] E. O. G. Shann. *Cattle Chosen: The Story of The First Group Settlement in Western Australia*. Oxford University Press, London, 1926.
- [15] P. Statham-Drew, *James Stirling: Admiral And Founding Governor Of Western Australia*. University of Western Australia Press, WA, 2003.
- [16] R. T. Tally Jr., Adventures in literary cartography: Explorations, representations, projections, in: E. Peraldo (Ed.) *Literature and Geography: The Writing of Space Throughout History*, Cambridge Scholars Publishing, Newcastle Upon Tyne, 2016, pp. 20-36.
- [17] M. T. Stephansen, A hand-drawn map as a decolonising document: Keviselie (Hans Ragnar Mathisen) and the artistic empowerment of the Sami Movement. *Afterall: A Journal of Art, Context, & Enquiry*, 44. 1 (2017) 112-121. doi: 10.1086/695520.

# Annotation of Named Entities in Medieval and Early Modern Epigraphic Texts

Gregor Pobežin<sup>1,2</sup>

<sup>1</sup> ZRC SAZU, Novi trg 2, 1000 Ljubljana, Slovenia

<sup>2</sup> University of Primorska, Faculty of Humanities, Titov trg 5, 6000 Koper, Slovenia

## Abstract

Relying on the annotation scheme proposed by Álvarez-Mellado et al. [1], this paper attempts to refine the proposed model for the annotation of named entities and adapt it to the needs of (medieval and early modern) epigraphy, exemplified in this article by the case of the MEMIS corpus, which brings together medieval and early modern inscriptions from the area of present-day Slovenia. Digital humanities (DH) tools and protocols provide us with ways to access and process elements of historical evidence on epigraphic monuments as documents: In addition to actual events, they include, in particular, names of persons and places. Named Entity Recognition (NER) is therefore of paramount importance for the extraction of biographical, prosopographical, etc. data. Building on the previous work of DH researchers in the field of encoding standards for humanities texts, this paper focuses on the previously unexplored medieval and early modern inscriptions in the northern Istrian (now Slovenian) towns of the former Republic of Venice.

## Keywords

Medieval epigraphy, Early modern epigraphy, Named-entity annotation, MEMIS

## 1. Introduction

As brief and seemingly silent as they may seem, epigraphic monuments are a rich source of historical data. So opulent and numerous they are, in fact, that they may even resemble an inexhaustible quarry of ... dead and crumbling stone. More often than not, they are written in half-legible long-faded script – and usually in Latin or some other presumably dead language. As a rule, they bear witness of people long dead and most certainly forgotten by the general audience of everyday passers-by.

And yet, epigraphic documents enjoy – perhaps now more than ever – *the* status of one of the most important historiographical sources: a huge portion of new information about the distant past is actually retrieved from inscriptions on stone or other durable materials. Even the most modest of inscriptions can convey a cartload of information on the political and/or cultural history of official structures as well as everyday life of people from all social strata. Individual efforts of researchers suggest that epigraphic materials represent a valuable historical resource; in the case of Greek and Roman inscriptions these prove to be an important contribution to the historical findings on the dynamics and dimensions of ethnic, economic, military etc. development in a given area, and also helped to define the dynamics of communication of a certain area with the neighboring (as well as more distant) regions. Studies of medieval and early modern inscriptions will produce important insights with the potential to either confirm or refute the currently valid historical findings and notions of cultural specifics of a particular area.

As a discipline, epigraphy has been around since at least the 15<sup>th</sup> century: one of the ground breakers in this discipline, Ciriaco d'Ancona (1391–1452), started compiling inscriptions soon after 1420 (his compilations are, unfortunately, all but lost). Systematic and scientific compiling of inscriptions and their scientific interpretation developed in the 16<sup>th</sup> century, but they took off

---

*Biographical Data in a Digital World 2022 (BD 2022) Workshop, DH 2022, 25.-29. July, 2022*

✉ gregor.pobezin@zrc-sazu.si (G. Pobežin, ZRC SAZU); gregor.pobezin@fhs.upr.si (G. Pobežin, UP FHŠ)

ORCID 0000-0002-3418-9767

© 2023 Copyright for this paper by its authors.  
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

[https://doi.org/10.3986/9789610508120\\_06](https://doi.org/10.3986/9789610508120_06)

particularly in the 17<sup>th</sup> and 18<sup>th</sup> century when top scientific papers started being published on the subject – and even became a popular object of epistolary distribution.

However, the bulk of this scientific production focused primarily on Greek and Roman epigraphic monuments so eminently researched and represented by the largest epigraphic project *Corpus inscriptionum Latinarum* (CIL), the still active undertaking of tracking epigraphic monuments, first published by the Berlin Academy of Sciences under the auspices of Theodor Mommsen in 1863 (the plans for the publication had already been drafted in 1815).

## 2. Digital Epigraphy Databases

In the past few decades – at least since the late 1980s –, Greek and Roman epigraphy made stellar breakthroughs in the world of digital humanities [2, 3, 4]. Several major digital epigraphic projects have come to represent just how energetically the epigraphic community embraced the digital tools: apart from CIL, which made its way online in 2007 [5], or the Clauss-Slaby Epigraphik Datenbank [6], an especially noteworthy project<sup>2</sup> is the Europeana network of Ancient Greek and Latin Epigraphy (EAGLE) [8], a collaborative database whose inscriptions search engine searches through several other existing collections: Arachne<sup>3</sup>, Archaia Kypriaki Grammateia Digital Corpus (AKGDC), Epigraphic Database Bari (EDB) [9], Epigraphic Database Heidelberg (EDH) [10], Epigraphic Database Rome (EDR) [11], Hispania Epigraphica Online (HE) [12], PETRAE [13], The Last Statues of Antiquity and Ubi erat lupa [14].

Medieval and early modern epigraphy is making similar progress though far less intensive and with a far lesser tradition (although the 17<sup>th</sup> century Dutch diplomat, historian, philologist and antiquarian Gisbert Cuper (1644–1716) dedicated his scientific work *Harpocrates, sive Explicatio imagunculae argenteae* (Utrecht, 1676) to epigraphical and numismatic problems – among them a medieval inscription of dubious origin from Iustinopolis – now Koper, Slovenia). In terms of the development of medieval epigraphy as a stand-alone scientific discipline, the thesauri of inscriptions from Italy, Spain [15] and Germany stand out. Collecting and studying inscriptions from the Italian countries have a particularly long tradition [16, 17]; addressing the important Christian centres, the scientific organisation of the *Monumenta epigraphica Christiana* strongly hints at CIL. Lately, these publications have only intensified in number; one such example (among many) is Paola Guerini's *Inscriptiones Medii Aevi Italiae* [18]. The medieval and early modern inscriptions, collected in Spain, are being systematically published in the so-called *Corpus inscriptionum Hispaniae medievalium* [19]; similarly, the medieval and early modern inscriptions, collected in the area of France, are published in the *Corpus des inscriptions de la France médiévale* [20]. However, these corpora remain, at least for the time being, on paper only. Only two noteworthy major corpora exist online i.e. the Deutsche Inschriften online [21] and the Epigraphica Europea [22].

Most of the above-mentioned projects (except the Epigraphica Europea, which is essentially a searchable image databank featuring some metadata) are based on the Leiden encoding system<sup>4</sup>. However, since the introduction of the EpiDoc initiative in the late 1990s [4, 24], several EpiDoc based epigraphic projects have been launched, perhaps most notably the Vindolanda Tablets Online<sup>5</sup> and the Inscriptions of Aphrodisias (IAph2007) [25, 26] or lately for instance the Cretan Institutional Inscriptions [27]. More or less obvious advantages of the system such as controlled vocabularies<sup>6</sup>, metadata and a wide variety of possibilities for encoding semantically rich information – e.g. expanding abbreviations or supplying missing text with possible tagging of

---

<sup>2</sup> See [2], [3] & [7] for a more detailed list of online epigraphy databases.

<sup>3</sup> <https://arachne.dainst.org/>.

<sup>4</sup> First published in 1932, the Leiden system harmonized various styles of editing and publishing inscriptions and papyri; see <https://labs.brill.com/sedev/sego/leidenplus/> for basic information and sigla. Timothy Finney [23] proposed a set of guidelines for converting Leiden-based editions into XML.

<sup>5</sup> <http://vindolanda.csad.ox.ac.uk/index.shtml>

<sup>6</sup> See EAGLE/EpiDoc vocabularies (<https://www.eagle-network.eu/resources/vocabularies/>) describing the types of a) material, b) execution technique, c) type of inscription, d) object type, e) decoration, f) dating criteria and g) state of preservation, featuring up to 13 languages.

details as to why the text is missing – make EpiDoc “clearly the way forward ... [when] compiling or contemplating compiling a database of Greek or Latin inscriptions ...” [3], despite some voices of skepticism [28]. It may well be the case that for trained scholars – epigraphers, paleologists, papyrologists etc. – Leiden is still “easier” to produce and read, however, it is also beyond doubt that EpiDoc allows for better abstraction of the Leiden conventions into digital form [4].

### 3. The Slovenian corpus of medieval and early modern inscriptions (MEMIS) and EpiDoc

In Slovenia, medieval and early modern inscriptions received little attention by the professional epigraphic discipline so far. Consequently, medieval and early modern epigraphic material has been severely neglected in comparison with the Greek and Roman epigraphic material, which means that until recently there were no catalogues or systematically compiled corpora of inscriptions (or parts of them) in Slovenia.

The Epigraphic Corpus of Mediaeval and Early Modern Inscriptions in Slovenia (MEMIS) [29] is a growing online corpus that collects Latin inscriptions located or discovered in the territory of Slovenian coastal towns (with emphasis on Koper and Piran). Its aim is to provide a methodological basis for the processing of mediaeval and early modern inscriptions in Latin (and in the vernacular languages), focusing on the study of epigraphic material (in the broadest sense of the word) belonging to the insufficiently researched area of medieval and early modern epigraphy.

One of its main challenges is to create an appropriate standard for the compilation, cataloguing and encoding of medieval and early modern inscriptions. As with in any other epigraphic project, there are a number of difficulties (or rather: peculiarities) to overcome. Apart from the fact that most of these inscriptions are written in Latin, which requires a clear distinction of morphological structures (especially noun cases) there is also the fact that these epigraphic documents, already faded and damaged, as they are, feature specific orthographic conventions and norms as well as possible errors. The corpus brings together inscriptions that are either still located in their primary context or have been moved or even destroyed and are therefore only accessible in transcriptions; the material was collected through fieldwork i.e., the recording and documentation of inscriptions *in situ*.

At this stage, the corpus features individual entries containing (in the order listed below):

- inscription ID
- physical description based on the EAGLE/EpiDoc vocabularies:
  - object description
  - material description
- history of the inscription
  - findspot/original location
  - current location
- links to photographs
- related bibliography if available
- the transcription of the text
- the (Slovenian) translation

Ligatures and abbreviations in the inscriptions have been properly expanded, missing text was supplied and the named entities were annotated. For example, a Leiden-based transcription of a hitherto little-known inscription of the Vergerii brothers from Iustinopolis (now Koper, Slovenia) – all high-profile clergy –, is supplied below. Located (until recently) in the Koper Assumption Cathedral, the inscription (Figure 1) was hidden from sight in the walled staircase leading to the choir. It commemorates Aurelio Vergerio (?–1532) the oldest of the three Vergerii brothers, the youngest of whom was Pier Paolo *il Giovane* (‘the Younger’; 1498– 1565), Justinopolitan bishop and famous apostate. Because of his apostasy, Pier Paolo suffered the fate of *damnatio memoriae* shared by his older brother Giovanni Battista (1492–). The most striking



feature of the otherwise exceptionally beautiful inscription with a very complex history [30] is the *damnatio memoriae*<sup>7</sup>:



**Figure 1:** The Vergerio inscription (Photo: T. Benedik, Restoration Centre)

Aurelio Vergerio Iacobi f(ilio)  
Ro(mani) pont(ificis) a secr(etis) Romae mortuo  
dum id munus cum summa omnium  
laude et admiratione obiret  
[[Ioannes Bap(tista) pius beneficio Polae]]  
[[et Pet(rus) Paulus Iust(inopolitanus) episc(opus)]]  
fratres posuere  
MDXLVIII  
Vita Christus et mors lucrum

‘To Aurelio Vergerio, the son of Giacopo. He died in Rome while performing his duties as the secretary to the Pope, much to the general admiration. [This inscription was erected by] his brothers Gianpaolo Vergerio, blessed by his noble deeds for Pola, and Pierpaolo Vergerio bishop of Justinopolis, 1548. To live is Christ, to die is gain.’

In comparison, this is how the text is rendered in EpiDoc:

```
<lb n="1"/><persName>Aurelio Vergerio</persName>  
<persName>Iacobi</persName> <expan>f<supplied  
reason="abbreviation">ilio</supplied></expan>  
<lb n="2"/><expan>Ro<supplied  
reason="abbreviation">mani</supplied></expan>  
<expan>pont<supplied  
reason="abbreviation">ificis</supplied></expan> a  
<expan>secr<supplied reason="abbreviation">etis</supplied></expan>  
<placeName>Romae</placeName> mortuo  
<lb n="3"/>dum id munus cum summa  
<lb n="4"/>laude et admiratione omnium obiret
```

<sup>7</sup> For details about the inscription and the circumstances of its defacing, see [31].



```

<lb n="5"/><damage><supplied
reason="damnatioMemoriae"><persName>Ioannes <expan>Bap<supplied
reason="abbreviation">tista</supplied></expan></persName>
<unclear>pius</unclear> beneficio
<placeName>Polae</placeName></supplied></damage>
<lb n="6"/><damage><supplied reason="damnatioMemoriae">et
<persName><expan>Pet<supplied
reason="abbreviation">rus</supplied></expan> Paulus</persName>
<expan>Iust<supplied
reason="abbreviation">inopolitanus</supplied></expan>
<expan>episc<supplied
reason="abbreviation">opus</supplied></expan></supplied></damage>
<lb n="7"/>fratres posuere
<lb n="8"/><date>MDXLVII</date>
<lb n="9"/>Vita Christus et mors lucrum

```

#### 4. Named Entity Annotation

One of the major and, indeed, fundamental [32] tasks is the annotation of medieval inscriptions relates to named entities (NE) – in our case particularly the person names – which seems straightforward enough, but isn't without its specific problems [33]. In Greek and Roman inscriptions person names can be anything from first names (*praenomen*, e.g. *Gaius*), family names (*nomen*, *nomen gentilicium*, e.g. *Iulius*), a nickname (*cognomen*, e.g. *Caesar*) or a combination thereof: *Gaius Iulius Caesar*. In the case of medieval and early modern inscriptions from the Venetian towns in present-day Slovenia, it is generally a combination of the first name and surname: *Antonius Zarottus*.

As a rule, Venetian-controlled cities of northern Istria had to be organised according to the Venetian legislature, featuring therefore a so-called *maggior consiglio* (Figure 2) consisting of all the major aristocratic families who participated in the administration business, contributing over time a vast number of family members' names to the long roster of the city councils, consuls, *syndaci* etc. Often, these names become repetitive to a point where confusion may occur, leading to potential faulty structuring, mismatching and misinterpretation in their ensuing prosopographical/biographical processing.

**M**illesimo quadringentesimo trigesimo primo, Indictione ix. die primo Mensis Martii infrascripti sunt Nobiles de Majori Consilio Justinopolis hic inferiusseriatim registrati Mandato Sp., & Gen. Viri Hominisboni Grittis in ejus secundo Regimine Potestatis, & Capitanei Justinopolis.

Serenissimus, & Excell. D. Thomas Mocenigo D. G. inclitus Dux Venetiarum	-	Exc. Philippus de Pola
Sp., & Gen. Vir D. Antonius Contareno Proc. Ecclesie S. Marci	-	Exc. Marcus Farello
Sp., & Gen. Vir D. Jacobus de Ripa Miles	-	* Exc. Variendus de Tarfia
Sp. D. Vitalis Miani cum Fratribus suis	-	Exc. Ambrosius Lugnano
Sp. D. Castellanus Minio	-	Exc. Grimaldus Telta
Sp. D. Julianus Lauredano Castellanus Castri Leonis	-	Exc. Nicolaus de Elio
Excell. Franciscus Beuazano Cancell. Venetiarum	-	Exc. Nicolaus Bonacurtio
Exc. de Colmanus de Vergetiis	-	Exc. Jeremias Malgranello
Exc. Christophorus de Senis	-	Exc. Damianus Canis
Exc. Natalis de Baldeno	-	- Exc. Philippus de Gauardo
Exc. Augustinus de Serenis	-	* Exc. Franciscus de Almerigoto
Exc. Bertus de Vanto	-	Exc. Michael de Nedelo
* Exc. Johannes Belgramonus	-	Exc. Cristophorus Spataris
Exc. Gaspar de Bratis	-	Exc. Joannes de Trefoldo
Exc. Nicolaus Grixonius	-	Exc. Hieronymus Albanensis
Exc. Dominicus de Muxela	-	Exc. Vincentius de Fino
Exc. Sanctorius de Bonzanigo	-	Exc. Zan nus de Chofa
* Exc. Basilius de Basilio	-	- Exc. Nicolaus de Spelatis
* Exc. Joannes de Sabinis	-	Exc. Jacobus de Languschiis
Exc. Victorius de Victore	-	* Exc. Hieronymus Joannis
Exc. Joannes de Ingaldeo	-	* Exc. Almericus de Verziis
Exc. Dominicus de Petronio	-	Exc. Bernardus de Pellegrino
Exc. Nazarius de Oliua	-	* Exc. Petrus de Azzo
Exc. Georgius de Luciato	-	- Exc. Jacobus del Seno
Exc. Joannes de Guizardo	-	- Exc. Victor de Rino
Exc. Dominicus de Oclatio	-	- Exc. Jacobus de Barbo
Exc. Lucas Scribano	-	Exc. Joannes de Rauenna
		Exc. Nicolaus de Plateo
		Exc. Joannes de Roma
		- Exc. Petrus de Salo
		Exc. Nazarius de Salo
		Exc. Petrus Paulus de Zarotti
		Exc. Cristophorus della Corte
		Exc. Nicolaus Agresta

Exc.

**Figure 2:** The list of families constituting the *maggior consiglio* of Iustinopolis.

For instance, several inscriptions on the main square (the once *Piazza del Duomo*) of the Venetian city of *Iustinopolis* (present-day *Koper*) bear the name *Pietro Loredan*. Others will name several *Maffei* – not all of them the same person. The above-mentioned case of a *Pier Paolo Vergerio* (Lat. *Petrus Paulus Vergerius*) may refer to two persons from Iustinopolis i.e., *Pier Paolo Vergerio il Vecchio* ('the Elder'; 1370–1444/45), the famous 15th century humanist, or *Pier Paolo Vergerio il Giovane*, the famous 16th century humanist and apostate. In MEMIS we are generally dealing with all the categories of named entities from the expanded list [1]: person names, location names, organization names, role names and, to a lesser extent, miscellanea. The introduction of **role names** is a particularly welcome addition, since the societal and inter-

personal roles make a highly represented category on inscriptions, making them an important source of information on the relationships between different persons mentioned on either a single inscription or several seemingly non-related monuments. Furthermore, when the MEMIS database grows to a more considerable extent, annotated role names will make it possible to analyze the occurrence and relevance of occupations, positions, military expertise etc.

As we've already mentioned, the EpiDoc guidelines (the latest version 9.5 was released on 26 April 2002) provide a solid and comprehensive system of controlled vocabularies, metadata, and a variety of ways of encoding semantically rich information, which is particularly important for one of the most common epigraphic features, i.e. the expansion of the ubiquitous abbreviations and/or the provision of missing text with the provided TEI tags, which is common practice in EpiDoc. So far, this approach seems to be error-proof and only requires possible further refinement - if and where necessary.

Apart from providing a consistent format, there is another crucial aspect to working with the EpiDoc/TEI schemas. As we've already mentioned, several epigraphic databases already work with them; this ensures not only their interoperability – as manifested by the EAGLE project – but also the fulfilment of all the FAIR components (findable, accessible, interoperable and reusable), since they are accessible via bulk EpiDoc XML/TEI downloads [34].

#### 4.1. A more finely granulated version of NE annotation

In this particular instance we are especially concerned with person names, for which Álvarez-Mellado et al. propose the user-friendly annotation scheme, suggesting the use of the simple XML-TEI tag `<persName>`. An additional nested tag `<addName>` is suggested for epithets, regnal numbers and nicknames (*cognomina*) as part of the official name [1], which is a welcome addition. An additional attribute to `<persName>` is suggested in the case of lone nicknames, deities and divine figures. According to this scheme it will suffice to annotate the entire name as:

```
<persName>Aurelio Vergerio</persName>
```

This is a great starting point; however, in the case of MEMIS – or other corpora of Latin inscriptions –, this scheme doesn't quite cover all the aspects, particularly with the problem of inflection where the proper noun (a name) occurs in any of the oblique cases e.g., dative: Aurelius Vergerius > Aurelio Vergerio. For Latin and especially for medieval and early modern inscriptions, a more granular annotation is therefore desirable, with the nominative case as the attribute as well as reference for later computer manipulations of the text:

```
<persName key="Aurelius Vergerius" ref="AuVerg"><name  
type="forename">Aurelio</name> <name  
type="surname"/>Vergerio</persName>
```

It is beneficial to think of combined named entities, particularly `personName` and `roleName`, the latter nested inside the main `<persName>` tag. For this purpose, let us take a look at a heavily damaged and linguistically interesting inscription for one Antonio Zarotto [35, 36] from Justinopolis who died during the Ottoman–Venetian War (1537–1540), perhaps during the battle in the Ambracian gulf or later during the siege of Castelnuovo (1539). PersonNames are rendered in bold characters, roleNames in italics and – very important to the mapping of interpersonal relations – relation types are rendered in underlined characters:

**Antonio Zarotto**  
*equiti splendidior(i)*  
qui b[el]lo contra  
Turcas suscepto  
tiremi Venet(ae) pro

Iustinopolitanis  
 praefectus Cret<a>e  
 sum<m>o cu(m) totius  
 classis m<a>erore  
 de qua optime  
 meritus erat  
 e vita decessit  
 an<n>o D(omini)  
 MDXXXIX aetatis LV  
**Franc(iscus) frater**  
 et ex hoc nepotes  
**Nicolaus eques**  
**Leander doctor**  
**Zar(otus)**  
 et **Io(annes) Paulus**  
 m<a>estiss(imi) posuerunt

‘To Antonio Zarotto, the most splendid knight who, in the name of Justinopolitans, went to war with the Turks as a captain of a Venetian trireme and died aged 60 in 1539, much to the dismay of the whole fleet, off the shore of Crete for which he so valiantly fought. [This monument was] erected by his sorrowful brother Francesco and, by him, his nephews knight Nicolo, barrister Leandro Zarotto and Gianpaolo.’

In EpiDoc, the transcription is far more granulated, with familial relations sketched under the tag <listRelation>:

```
<lb n="1"/><persName key="Antonius Zarottus" ref="AntZar"><name
type="forename">Antonio</name> <name
type="surname"/>Zarotto</persName><lb n="2"/><roleName
type="honorific"><expan>equit<supplied
reason="abbreviation"/>i</expan></roleName>
<expan>splendidior<supplied reason="abbreviation"/>i <...>
<persName key="Franciscus Zarottus"
ref="FrZar"><expan>Franc<supplied
reason="abbreviation">iscus</supplied></expan></persName>
<expan>fr<supplied reason="abbreviation">ater</supplied></expan>
<lb n="16"/>et ex hoc nepotes
<lb n="17"/><persName key="Nicolaus Zarottus" ref="NicZar"><name
type="forename"><expan>Nicol<supplied
reason="abbreviation">aus</supplied></expan></name></persName><ro
leName type="honorific">eques</roleName>
<lb n="18"/><persName key="Leander Zarottus" ref="LeZar"><name
type="forename">Leander</name> <roleName
type="occupation">doctor</roleName>
<lb n="19"/><name type="surname"><expan>Zar<supplied
reason="abbreviation">otus</supplied></expan></name></persName>
<lb n="20"/>et <persName key="Ioannes Paulus Zarottus"
ref="GPZar"><name type="forename"><expan>Io<supplied
reason="abbreviation">annes</supplied></expan>
Paulus</name></persName>
<lb n="21 m<supplied reason="missing">a</supplied>estiss<supplied
reason="abbreviation">imi</supplied>
```

```

</expan> <expan>p<supplied
reason="abbreviation">osuerunt</supplied></expan>
</ab>
<listRelation type="personal">
<relation name="sibling" mutual="#AntZar #FrZar"/>
<relation name="parent" active="#FrZar" passive="#NicZar #LeZar
#GPZar"/>
</listRelation>

```

None of the tags of this very “verbose” description [4] are actually displayed. In fact, converted by the IJS interface<sup>8</sup> that runs some recent version of the TEI Stylesheets with various local profiles, it looks (quite unattractively) something like this:

Antonio Zarotto  
equiti splendidiori  
qui bello contra  
Turcas suscepto  
triremi Venetae pro  
Iustinopolitanis  
praefectus Cretae  
summo cum totius  
classis maerore  
de qua optime  
meritus erat  
e vita decessit  
anno Domini  
MDXXXIX aetatis LV  
Franciscus frater  
et ex hoc nepotes  
Nicolaus eques  
Leander doctor  
Zarottus  
et Ioannes Paulus  
maestissimi posuerunt

The elements of Antonio Zarotto’s name as well as names of other family members mentioned on the inscription are annotated so that they may be searched for other occurrences in the corpus. The main person of interest on this inscription is clearly Antonio Zarotto, but the inscription links him with other persons (Francesco, Nicolo, Leandro, Gianpaolo) mentioned, apart from this particular case, by at least two other monuments. On the other hand, Antonio Zarotto must be disambiguated from at least one other known contemporary and Anotnio’s namesake from Parma.

For this purpose, ‘key’ attribute – the name in the nominative case – is added to the <persName> tag as well as a reference attribute (name acronym in this case) element, in order to standardize the lists for searching, lest the machine searches for “Antonio” or even “Nicol”, “Zar” and “Io” only.

## 5. Conclusion

---

<sup>8</sup> <http://nl.ijs.si/tei/convert/>

The above-listed solutions reached with the TEI/EpiDoc solutions being only a fragment of the vary vast array of possibilities at hand, “things are coming together” [3] for digital epigraphy – and particularly the Slovenian project MEMIS.

But it is precisely because of the nature of the material that some very interesting and indeed fundamental research questions have already been raised, among them perhaps the most obvious: is it possible to map the networks of people mentioned in medieval and early modern inscriptions, which are far more numerous than Greek and Roman monuments? Each church contains dozens, if not hundreds, of them.

The solutions to the problems of NE annotation proposed by the article cited in the introduction to this paper [1] are indeed useful, but as we've pointed out, they require further refinement and granularity because of the linguistic differences in the material.

There is another aspect of the more finely granulated annotation this paper proposes i.e., the elements of the `<listRelation>` tag, which will expectedly – with time – yield the possibility to visualise a network of family relationships, professional relationships, commercial patterns etc., or even build prosopographical profiles of otherwise less known but historically important individuals like Antonio Zarotto.

The `<persName>` tag will only enable searching for connections of the inscriptions that mention the same name (but not necessarily the same individual!), whereas other attributes like `<roleName>` and particularly the elements of the `<listRelation>` tag give hope of exciting new discoveries as well as exciting new possibilities of data visualisation.

## Acknowledgements

This paper owes its existence to the work conducted within the framework of the project ‘INTAVIA: In/Tangible European Heritage Visual Analysis, Curation & Communication’, has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 101004825, as well as the project ‘The Pretorial Palace of Koper/Capodistria: Form, Meaning and Function’ (J6-2588), which has received funding from the Slovenian Research Agency (ARIS) and research program P-0440 funded by the Slovenian Research Agency (ARIS).

## References

- [1] E. Álvarez-Mellado, M.L. Díez-Platas, P. Ruiz-Fabo et al. 2021. “TEI-friendly Annotation Scheme for Medieval Named Entities: a Case on a Spanish Medieval Corpus.” *Language Resources & Evaluation* 55 (2021): 525–549. doi: 10.1007/s10579-020-09516-2
- [2] A. Babeu. 2011. ‘Rome Wasn’t Digitized in a Day’: Building a Cyberinfrastructure for Digital Classics. Council on Library and Information Resources, Washington. URL: <http://www.clir.org/pubs/abstract/pub150abst.html>.
- [3] J. Bodel, *Latin Epigraphy and the IT Revolution*, in: J. K. Davies & J. J. Wilkes (Eds.), *Epigraphy and the Historical Sciences*, OUP, Oxford, 2012, pp. 275–296.
- [4] H. Cayless, C. Roueché, T. Elliott & G. Bodard, *Epigraphy in 2017*, *Digital Humanities Quarterly* 3 (2009). URL: <http://www.digitalhumanities.org/dhq/vol/3/1/000030.html>
- [5] C. Marksches, *Corpus inscriptionum Latinarum*. URL: <https://cil.bbaw.de/>.
- [6] M. Clauss et al., *EDCS – Epigraphik-Datenbank Clauss/Slaby*. URL: <http://www.manfredclauss.de/gb/index.html>.
- [7] K. Steiner, S. Mahony. 2016. “How are Digital Methods Changing Research in the Study of the Classical World? An EpiDoc Case Study.” *Panta Rei. Revista Digital de Ciencia y Didáctica de la Historia* (2016): 125–148. doi: 10.6018/pantarei/2016/8
- [8] S. Orlandi et al., *Europeana Eagle Project*. URL: <https://www.eagle-network.eu/>.
- [9] C. Carletti et al., *Epigraphic Database Bari project (EDB)*. URL: <http://www.edb.uniba.it/>.
- [10] G. Alföldy et al., *Epigraphic Database Heidelberg (EDH)*, 1986. URL: <https://edh.ub.uni-heidelberg.de/>.

- [11] Epigraphic Database Roma (EDR), 2003. URL: <http://www.edr-edr.it/default/index.php>.
- [12] G. Pantoja et al., Hispania epigraphica online. URL : <http://eda-bea.es/>.
- [13] O. Devillers et al., PETRAE – Programme d’Enregistrement, Traitement et Reconnaissance Automatique en Épigraphie, 2012. URL : <https://petrae.huma-num.fr/fr/>.
- [14] F. und O. Harl, Ubi erat lupa. URL: <http://lupa.at/>.
- [15] J. Vives, Inscripciones cristianas de la España romana y visigoda, 2nd. ed., A. G. Ponsa, Barcelona, 1969.
- [16] G. B. De Rossi, A. Silvagni, Inscriptiones Christianae Urbis [Urbis] Romae Septimo Saeculo [Saeculo] Antiquiores [Antiquiores]. Vol. 1 (1861), Pont. Inst. Archaeologiae Christianae, Rome, 1978.
- [17] A. Silvagni, Monumenta epigraphica christiana saeculo XIII antiquiora quae in Italiae finibus adhuc exstant, Pont. Inst. Archaeologiae Christianae, Roma, 1943.
- [18] P. Guerini, Inscriptiones Medii Aevi Italiae (saec. VI-XII) II: Umbria, Terni, Fondazione Centro italiano di studi sull’alto Medioevo, Spoleto, 2010.
- [19] M. E. Martín López, Las Inscripciones de la Catedral de León (SS. IX-XX), Corpus Inscriptionum Hispaniae Mediaevalium, León, 2014.
- [20] C. Treffort et al., Corpus des inscriptions de la France médiévale 23 : Côtes-d’Armor, Finistère, Ille-et-Vilaine, Morbihan (région Bretagne), Loire-Atlantique et Vendée (région Pays de la Loire), CNRS, Paris, 2008.
- [21] Deutsche Inschriften online. URL: <https://www.inschriften.net/>.
- [22] F. A. Bornschlegel et al., Epigraphica Europea. URL: <http://www.epigraphica-europea.uni-muenchen.de/>.
- [23] T. Finney, Converting Leiden-style Editions to TEI Lite XML, 2001. URL: <https://www.tfinney.net/Leiden/index.html>.
- [24] G. Bodard, EpiDoc: Epigraphic Documents in XML for Publication and Interchange, in: F. Feraudi-Gruénais (Ed.), Latin On Stone: Epigraphic Research and Electronic Archives, Lexington Books, Lanham, 2010, pp. 101–118.
- [25] G. Bodard, C. Roueché, & J. Reynolds, Inscriptions of Aphrodisias (IAph2007), 2007. URL: <http://insaph.kcl.ac.uk/iaph2007>.
- [26] G. Bodard, “The Inscriptions of Aphrodisias as Electronic Publication: A User’s Perspective and a Proposed Paradigm.” Digital Medievalist 4 (2008). doi: 10.16995/dm.19.
- [27] I. Vagionakis, Cretan Institutional Inscriptions Dataset. CLARIN-IT repository hosted at Institute for Computational Linguistics “A. Zampolli”, National Research Council, in Pisa, 2021. URL: <http://hdl.handle.net/20.500.11752/OPEN-548>.
- [28] FL. Álvarez et al., Sharing Epigraphic Information as Linked Data, in: S. Sánchez-Alonso, I. N. Athanasiadis (Eds.), Metadata and Semantic Research. MTSR 2010. Communications in Computer and Information Science, vol 108, Berlin/Heidelberg, pp. 222–234.
- [29] G. Pobežin, Epigraphic corpus of Medieval and Early Modern inscriptions in Slovenia MEMIS 1.0. Slovenian language resource repository CLARIN.SI, 2020. URL: <http://hdl.handle.net/11356/1376>.
- [30] Z. Mileusnić, “Nuove conoscenze sugli inizi urbani della città tardoantica di Capodistria.” Quaderni friulani di archeologia 31 (2021): 55–60.
- [31] G. Pobežin, “Napis Vergerijev v koprski stolnici: kratka zabeležka in rekonstrukcija napisnega polja – The Inscription of Vergerii from the Koper Cathedral: a Short Epigraphic Note with the Reconstruction of the Damnatio Memoriae.” Studia universitatis hereditati 8/1 (2020): 97–102.
- [32] D. Nadeau, S. Sekine, “A survey of named entity recognition and classification.” Linguisticae Investigaciones 30/1 (2007): 3–26.
- [33] K. Fort et al., Towards a Methodology for Named Entities Annotation, in: Proceedings of the Third Linguistic Annotation Workshop (LAW III), Association for Computational Linguistics, Singapore, 2009, pp. 142–145.
- [34] P. Heřmánková, V. Kaše A. & Sobotkova, “Inscriptions as data: digital epigraphy in macro-historical perspective.” Journal of Digital History, 1/1(2021). doi: 10.1515/JDH-2021-1004?locatt=label:JDHFULL



- [35] A. Cherini, *Bassorilievi araldici ed epigrafi di Capodistria. Dalle origini al 1945*, Fama Capodistriana, Trieste, 2001.
- [36] G. Radossi, *Monumenta heraldica iustinopolitana: stemmi di rettori, di famiglie notabili, di vescovi e della città di Capodistria*, Centro di Ricerche Storiche, Rovigno, 2003.

# Biographical and Prosopographical Analyses of Finnish Academic People 1640–1899 Based on Linked Open Data

Petri Leskinen<sup>1,2</sup>, Eero Hyvönen<sup>1,2</sup>

<sup>1</sup>*Semantic Computing Research Group (SeCo), Aalto University, Finland*

<sup>2</sup>*Helsinki Centre for Digital Humanities (HELDIG), University of Helsinki, Finland*

## Abstract

This paper presents work on prosopographical data analyses using the AcademySampo linked data service and portal. The original primary data, based on ten man-years of digitization work, covers a significant part of the Finnish university history based on the student registries in 1640–1852 and 1853–1899. They contain biographical descriptions of 28 000 students of the University of Helsinki, originally the Royal Academy of Turku. AcademySampo also sheds light to the academic history of Sweden and Baltic countries through their shared history with Finland in the larger Swedish empire. The Finnish student registries have been widely used by genealogists and historians by close reading. The main focus of this article is on the networks connecting the students and on knowledge discovered using this methodology. Networks connecting the students as well as their relatives mentioned in the data can be constructed based on various criteria, e.g., by genealogical relations, or by similarities on career by common vocations and employees. The student records already have a linkage to related Wikidata resources, which have been earlier used for enriching, e.g., the information about the relatives mentioned in the register descriptions. In this paper the biographical data is further extended by using Wikidata and by extracting further information and connections from the textual descriptions in Finnish, Swedish, or English Wikipedia. Although the descriptions in AcademySampo provide detailed data about the academic careers and family relations, the related Wikipedia entries can provide more details about their lifetime events with, e.g., their known locations of work or residence, topics of interest, lifetime events, or acquaintances. Topic of specific interest in this paper include: 1) Inheritance analysis of vocations and social classes in families. This analysis uses correlation matrices based on vocations of the students and their parents. 2) Quantitative analyses and visualizations of the family lines of students, based on automatically created family trees of the students and their parents. Family lines as long as eight generations can be found.

## Keywords

Linked Data, Data Analysis, Digital Humanities, Network Analysis, Cultural Heritage

---

*Biographical Data in a Digital World 2022 (BD 2022) Workshop, DH 2022, 25.-29. July, 2022*

✉ petri.leskinen@aalto.fi (P. Leskinen)

ORCID 0000-0003-2327-6942 (P. Leskinen); 0000-0003-1695-5840 (E. Hyvönen)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

[https://doi.org/10.3986/9789610508120\\_07](https://doi.org/10.3986/9789610508120_07)

## 1. Introduction

ACADEMYSAMPO<sup>1</sup> [1, 2] consists of two parts: 1) a Linked Open Data (LOD) service<sup>2</sup> published on the Linked Data Finland platform [3] and 1) a semantic portal<sup>3</sup> based on it. The ACADEMYSAMPO Portal provides intelligent capabilities for searching and browsing with seamlessly integrated data analytical tools and visualizations for biographical and prosopographical [4] research using statistics, networks, timelines, and maps. The open Application Programming Interfaces (API) of the LOD service and its SPARQL endpoint, in turn, provide an easy-to-use opportunity to implement data analyses for DH researchers with some experience in the SPARQL query language<sup>4</sup> and programming. For example, the Yasgui editor [5], Python scripts, Jupyter<sup>5</sup>, and Google Colab notebooks<sup>6</sup> can be used.

ACADEMYSAMPO is part of the Sampo portal series<sup>7</sup> [6] and uses the Linked Open Data Infrastructure for Digital Humanities in Finland (LODI4DH)<sup>8</sup> [7], a part of the Finnish FIN-CLARIAH infrastructure initiative<sup>9</sup>. This paper describes and compares a set of four networks created from the AcademySampo actor data, and introduces various results from analyzing the relationships found in genealogical or academic connections or in lifetime events.

## 2. Primary Data and Knowledge Graph

ACADEMYSAMPO's data form an extensive knowledge graph that has been produced algorithmically from the digitized student registers of the Royal Academy of Turku and the University of Helsinki in 1640–1852 and 1853–1899<sup>10</sup> by extracting information from the texts and database structures. The data has been enriched by linking it both internally by artificial intelligence-based reasoning, and externally to other open datasets [8]. The student registers describe all people who have received academic education in Finland in 1640–1899, as there were no other universities in Finland at that time. The descriptions of students tell not only about their studies, but also about their career after studies and relatives, as well as references to the literature. The original register of the Royal Academy of Turku was destroyed in the Great Fire of Turku in 1827, but it was reconstructed in the late 19th century by Vilhelm Lagus. The register was supplemented in the 20th century from various sources, and in the end the information was edited by Yrjö Kotivuori and Veli-Matti Autio in an effort of ca. ten man years.

Since the registers 1640–1852 and 1853–1899 were provided by different authors their tabular CSV data differ to some extent. The source information found in the table of records 1640–1852 includes, in addition to some technical information in the database: 1) the person's registration number, 2) HTML text showing the person's name, places and times of birth and death, parents,

---

<sup>1</sup>Project homepage: <https://seco.cs.aalto.fi/projects/yo-matrikkelit/>

<sup>2</sup>The LOD service is available at <https://ldf.fi/dataset/yoma>

<sup>3</sup>Portal was opened February 2, 2021 at <https://akatemiasampo.fi/en/>

<sup>4</sup><https://www.w3.org/TR/sparql11-query/>

<sup>5</sup>Jupyter Project and Tool: <https://jupyter.org>

<sup>6</sup>Google Colab: <https://colab.research.google.com/notebooks/intro.ipynb#recent=true>

<sup>7</sup>See: <https://seco.cs.aalto.fi/applications/sampo/>

<sup>8</sup>LODI4DH initiative: <https://seco.cs.aalto.fi/projects/lodi4dh/>

<sup>9</sup><https://seco.cs.aalto.fi/projects/fin-clariah/>

<sup>10</sup>Student Registers, University of Helsinki: <https://www.helsinki.fi/fi/yliopisto/ylioppilasmatrikkelit-1640-1907>

career events, relatives, students, references and 3) the date the record was created. If the person mentioned in the register 1640–1852 is found in either of the registers, a HTML link is manually created connecting this mention to the person’s page using the registration number. However, in the register 1853–1899 there are no such links, and the references have been interpreted computationally. In addition, supplementary textual information about a person may be available in other registers. For example, Finnish national poet *Johan Ludvig Runeberg* has further information in the registers of Lagus and Carpelan.

The primary data used in creating ACADEMYSAMPO was therefore mainly text in HTML format without structured metadata, such as places or times of birth, vocation, etc. A major technical challenge in creating the linked data was to unambiguously identify the entities and events mentioned in the text, such as marriages, rewards and promotions, and key concepts, such as vocations. A specific challenge in extracting information was to distinguish between people with the same name, to reason their gender by name, and to infer various relationships, such as little cousin, through other relationships.

The data of the ACADEMYSAMPO was converted into Linked Data<sup>11</sup> [9] by structuring the text descriptions of the Student register 1640–1852 for about 9500 people and the register 1853–1899 for about 18 450 people. This was done by identifying, through regular expressions, basic biographical information about students, their 47 000 relatives, 120 000 interpersonal relationships, 3000 historical places, 10 000 vocations, and 4000 academic teacher-student relationships. The “semantic glue” of the knowledge graph are the events related to the professional and family life of people identified in the texts, which link the people and organizations involved in different roles with places and times according to the CIDOC CRM<sup>12</sup> ontology and ISO standard. The data has been enriched by linkage to external databases, such as the Finnish National Biography and other biographies of the Finnish Literature Society available as LOD in the BiographySampo system [10] and Wikidata<sup>13</sup>, and by inferring relationships between people [8]. The public open data service (CC BY 4.0) is available at the Linked Data Finland for accessing and utilizing the data in research and application development, such as the ACADEMYSAMPO portal.

### 3. Networks based on different criteria

Social networks can be constructed from a biographical LOD publication with various, different criteria. In this chapter four such networks are introduced and analyzed. The examples are 1) Genealogical family relation network, 2) Teacher-student relation network, 3) A network based on similarity of life events, and 4) a reference network imported from Wikipedia. The analyses presented in this chapter were generated in Google Colab notebooks using the data available at the AcademySampo SPARQL endpoint. The SPARQL results were converted into social networks using the Python module NetworkX<sup>14</sup> [11]. The figures were generated using Python modules Matplotlib<sup>15</sup> and Seaborn<sup>16</sup> or using the network visualization software Gephi [12]

---

<sup>11</sup>W3C Linked Data: <https://www.w3.org/standards/semanticweb/data>

<sup>12</sup>CIDOC CRM: <http://cidoc-crm.org>

<sup>13</sup>Wikidata: <https://wikidata.org>

<sup>14</sup><https://networkx.org/>

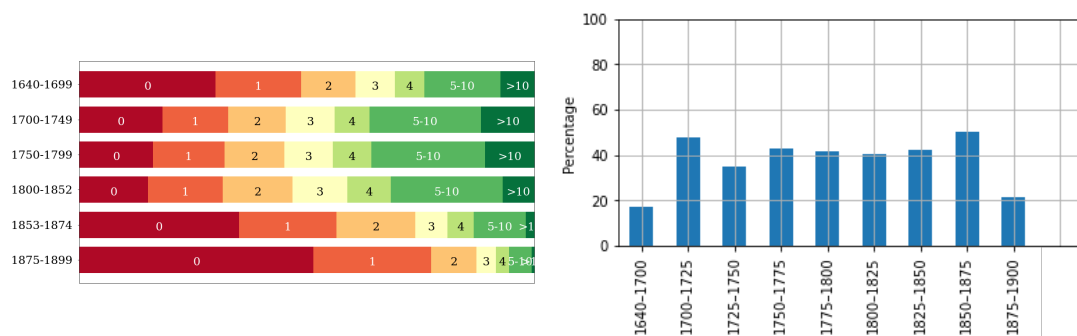
<sup>15</sup><https://matplotlib.org/>

<sup>16</sup><https://seaborn.pydata.org/>

after exporting the network in graphml [13] format.

### 3.1. Family Relations

The ACADEMYSAMPO data is rich with detailed family relationships which were manually added by the authors of the 1640–1852 register data. This linkage was later used as training data for linking the relatives in the 1853–1899 dataset [8]. The students are interconnected with 66 types of relations from close relations like parent or child to more distant ones like stepfather-in-law [14]. Approximately 18 900 students have at least one relative among the students. The diagram on the left in Figure 1 depicts the percentage of students who have other relatives among the students during the centuries, and the time series on the right shows the percentage of students whose parent also studied at the University. Generally, during the later half of 19th century the amount of students without an academic family background starts increasing rapidly.



**Figure 1:** The diagram on the left depicts the number of relatives in the university for each student; the time series on the right depicts the proportion of students whose father also studied at the University.

To further analyze the length of family lines, a genealogical network was created, this time using only the parent-child relations. In this network each kin is represented as a connected component, and the number of generations equals to the length of the longest path in it. In Table 1 the first column is the number of generations in the family line and the second one is the number of kins with that number of generations. For example, there are five kins with a length of eight generations. The largest subgraph has the size 66 nodes. The value 8171 at the bottom row is the number of students without relatives among the students. Notice that a subgraph with a maximum path length 8 contains also subpaths with all shorter lengths (7, 6, 5, ...).

Table 2 lists examples of the student names along two family lines, so that the oldest ancestors are listed first. One could, for example, pay attention to the changes in the spelling of the family names during the centuries. A closer look at the biography of *Nils Abraham Ursin* reveals that in 1845 he was ennobled with the family name *af Ursin*. The history of each family line could be further analyzed by, e.g., looking at the places of birth and death along the family history. Both of the families have their roots in small villages (*Kalvola, Eura, Rantasalmi*) but later they have moved to larger towns (*Turku, Helsinki, Kuopio*) in Finland.

Figure 2 visualizing the changes in the most common places of birth during the years 1650–1900. The places considered are towns and municipalities of Finland and the neighbour countries

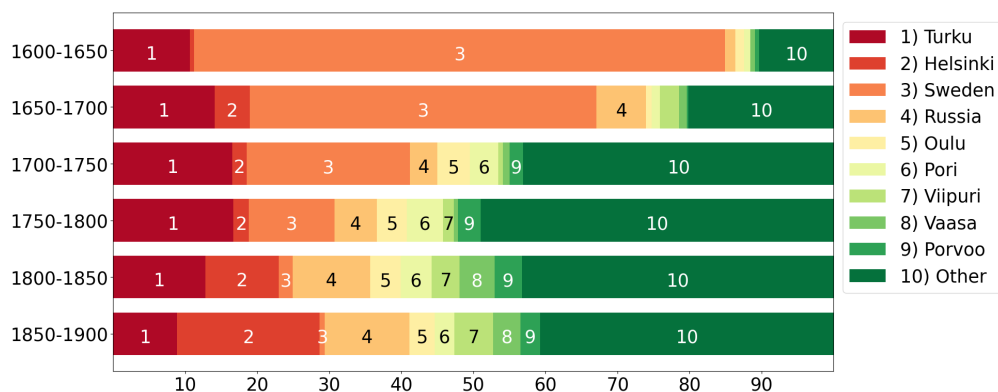
**Table 1**  
Lengths and amounts of family lines

generations	#families
8	5
7	21
6	47
5	81
4	205
3	547
2	1816
1	8171

**Table 2**  
Examples of person names in two family lines  
in format *family name, given names*

Family Line 1	Family Line 2
Homman, Tomas	Ursinus, Jakob
Homeen, Johan	Ursinus, Jakob
Homeen, Johan	Ursinus, Nils
Homén, Johan Jakob	Ursin, Jakob Johan
Homén, Gustaf Vilhelm	af Ursin, Nils Abraham
Homén, Lars	af Ursin, Julius
Homén, Lars Olaf	af Ursin, Nils Robert

Sweden and Russia. Number of student born in other countries was not large enough to have an effect on the results. Turku was the old capital and the largest town in Finland but started to lose its significance during the first half of 19th century when Helsinki first became the new capital and later when the university was moved to Helsinki. The figure also shows that the number of students coming from Sweden was high in the 17th and 18th centuries but decreased significantly in the 19th century when Finland became a part of Russian Empire. Consequently, at that time there is increase in the number of students born in Russia.

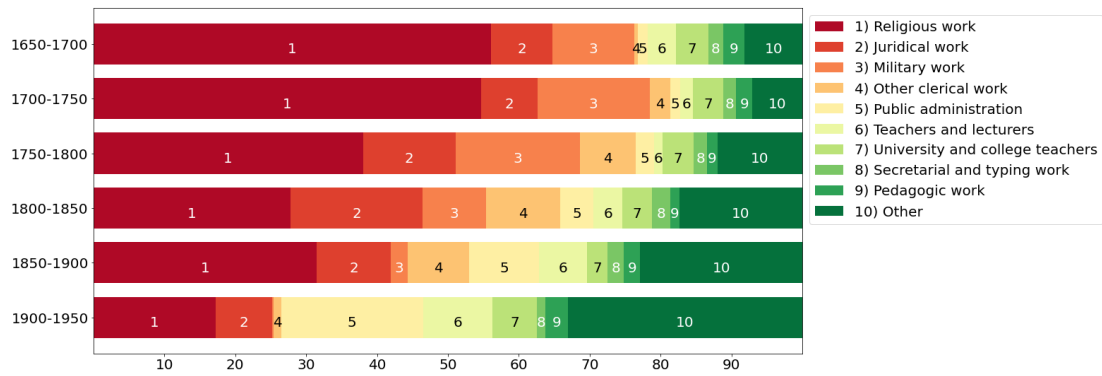


**Figure 2:** The most common places of birth during the years 1650-1900

Figure 3 depicts the distribution of most common vocational groups during the years 1650-1950. In the early years of the time frame, *Religious work* was the dominant category with a portion of over 50 %. However, that significance decreased to a mere 17 % at the 20th century while vocations related to, e.g., *Public administration* gained more importance. Furthermore, there is a notable growth in the proportion for new occupations in the category *Other*.

Figure 4 shows a correlation matrix between the vocational groups of parent-children pairs. The full labels of the vocational groups are shown on the rows; the columns have the same





**Figure 3:** The most common vocational categories during the years 1650–1950

order but only the indices are shown underneath the figures. The values in the cells are the probabilities that a child working in a field represented by the row has had a parent working in the field for the corresponding column. On each row the cell with the largest value has the darkest background color, and all the values on a row sum up 100 %. For example, the uppermost value on the left indicates that 49.1 % of children in the religious work category have had parents in religious work category as well, likewise 53.1 % of the students with parents in the field of agriculture, forestry and fishing (row 6) has chosen a religious work. The values on the matrix diagonal are the probabilities of choosing the same vocational group as one's parent has. However, when looking at these statistics one has to remember the academic context. For instance, on the 6th column all the values are relatively small meaning that independent of the parents' vocation only a few children has chosen a work in agriculture, forestry or fishing, although in 17th–19th centuries Finland was an agricultural country.

Figure 5 depicts the correlations between siblings born in the 19th century. Altogether the dataset contains approximately 8800 such sibling pairs. The numbers in the cells are the number of related pairs. By looking at the matrix one can notice that the four most dominant categories, *Public administration*, *Religious work*, *Juridical work*, and *Teachers and lecturers* have high values of correlation. One also has to remember the biases of our data, like for instance the intercorrelation in the field of *Military work* remains low due to the fact that people who chose a military career may not have studied at the university.

### 3.2. Teacher-Student Relations

The dataset contains a network of teachers and students spanning from 1640's to the year 1853. In a similar way to the family relations, this linkage was manually added by the original dataset authors. There are altogether 4893 links connecting 3159 people. In our work based on the LOD service, network statistics were used to, e.g., locate the most significant individuals. Figure 6 shows how the network spans continuously over the entire time window. On this illustration three most central actors are emphasized, *Henrik Gabriel Porthan*, *Jakob Gadolin*, and *Algot Scarin* who all are famous scholars and professors.

Parent	Religious work (1)	49.1%	12.9%	9.3%	9.5%	4.0%	1.7%	4.5%	1.1%	6.3%	1.8%
	Juridical work (2)	17.6%	29.9%	14.8%	7.5%	7.0%	2.2%	11.7%	1.1%	4.4%	3.7%
	Public administration (3)	11.2%	23.8%	23.6%	7.3%	6.9%	1.9%	11.4%	1.6%	5.8%	6.5%
	Teachers and lecturers (4)	17.8%	15.3%	20.7%	15.8%	3.4%	2.0%	5.1%	0.5%	12.9%	6.6%
	Military work (5)	15.1%	26.3%	12.4%	4.0%	23.7%	2.1%	11.1%	0.7%	2.6%	1.9%
	Managerial work in agriculture, forestry and fishing (6)	53.1%	8.8%	11.5%	10.1%	3.0%	5.3%	1.7%	0.3%	3.0%	3.1%
	Other clerical work (7)	4.8%	34.5%	17.3%	5.2%	7.4%	1.5%	18.1%	0.6%	5.0%	5.6%
	Wholesale and retail dealers (8)	14.6%	16.7%	18.1%	8.3%	3.0%	2.7%	6.5%	16.2%	5.2%	8.7%
	University and college teachers (9)	12.3%	19.7%	16.3%	10.9%	7.2%	1.3%	9.8%	1.0%	17.7%	3.7%
	Administration of private enterprises and organizations (10)	11.3%	9.6%	26.0%	11.1%	4.2%	2.7%	4.7%	5.2%	8.1%	17.2%
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
		Child									

**Figure 4:** Correlations of the vocational categories between parents and children

Public administration (1)	176	226	283	186	138	136	135	101	50	61	
Religious work (2)	226	345	176	237	49	38	55	48	36	35	
Juridical work (3)	283	176	215	133	177	59	86	61	96	90	
Teachers and lecturers (4)	186	237	133	116	46	44	68	81	32	28	
Other clerical work (5)	138	49	177	46	48	16	32	23	60	55	
Administration of private enterprises and organizations (6)	136	38	59	44	16	60	44	33	12	5	
Medical and nursing work (7)	135	55	86	68	32	44	25	34	12	24	
University and college teachers (8)	101	48	61	81	23	33	34	42	8	12	
Military work (9)	50	36	96	32	60	12	12	8	33	42	
Other manufacturing work (10)	61	35	90	28	55	5	24	12	42	22	
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)

**Figure 5:** Correlations of the vocational categories between siblings born in the 19th century

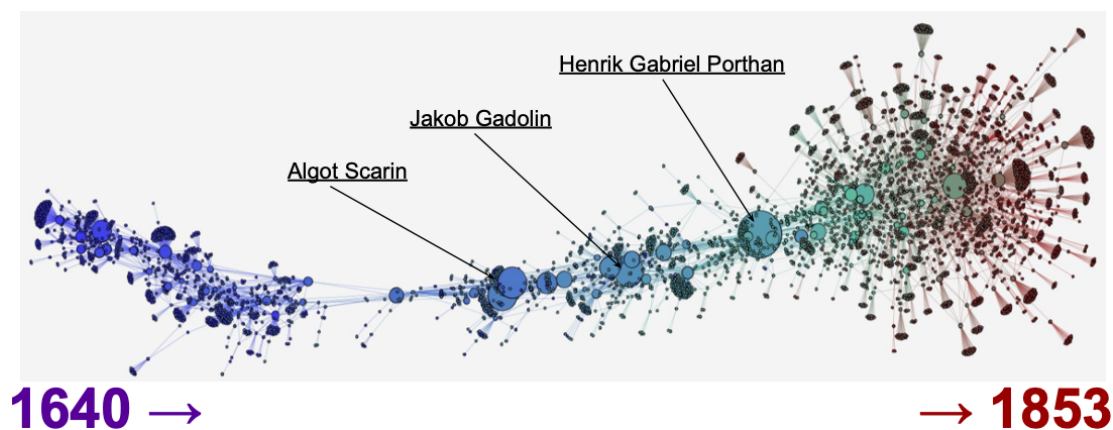


Figure 6: Network based on teacher-student relations

### 3.3. Similarity of Lifetime Events

Similarity between the students was calculated from the RDF data using the lifetime events. Features like having the same vocation, participating in the same event, being in the same places etc. were considered as links connecting the students. The similarity measure was achieved computationally by 1) a breadth-first search querying all the nodes related to each student including, e.g., the hierarchy of places, time spans, and vocations, 2) by filtering out the nodes that are related to very few or to too many students, 3) by constructing a matrix where each related entity is a feature (column) for a student (row), 4) by applying the TF-IDF measure to reduce the weight of most common terms and to emphasize the rarer ones, and by 5) calculating the similarity using cosine similarity. In addition to this method, also *RDF2VEC embeddings* [15] were tested on the data. However, the similarities and recommendations achieved with embeddings did not seem feasible. Furthermore, the approach above allowed to adjust weights by the class of a feature, e.g., to enhance the importance of, e.g., related organizations or inversely reduce the importance of a common time span.

In the data publication the recommended similarities are modeled as RDF resources connecting the two students, indicating the similarity value, and containing links to the database entries having the highest effect on the found similarity. For example, the people similar to chemist and physicist Johan Gadolin are found based on terms like *mineralogist*, *Uppsala University*, and *Royal Swedish Academy of Sciences*. Another example of similarities is a cluster of engineers<sup>17</sup> who worked in Baku, Azerbaijan, for the oil company Branobel<sup>18</sup> that was run by the Nobel brothers. In the ACADEMYSAMPO PORTAL these connections are shown as a network visualization<sup>19</sup>.

<sup>17</sup><https://akatemiasampo.fi/en/people/page/p17642/table>

<sup>18</sup><https://en.wikipedia.org/wiki/Branobel>

<sup>19</sup><https://akatemiasampo.fi/en/people/page/p17642/connections>

**Table 3**

Typical classes of links found in Wikipedia pages

Class of Wikipedia Entity	Count
human	1443
municipality of Finland	236
former municipality of Finland	174
city	102
newspaper	97
town	70
academic discipline	65
position	64
profession	63
organization	59

### 3.4. Enriching Data from an External Databases

Register descriptions of people are often short, and an external database can provide more detailed information about their lifetime. The ACADEMYSAMPO DATA SERVICE contains also a linkage to external data publications, such as the Finnish BiographySampo<sup>20</sup> [10], Members of Finnish Parliament, Ministers of Finland, as well as the international Wikidata. Using the linkage to Wikidata allows to also access the related Wikipedia pages written in various languages. Out of the total of 28 000 students approximately 2700 have an entry in the Finnish Wikipedia. The description text from the Finnish Wikipedia page was queried for each person. The graph is constructed based on the links on the pages so that two entities having the same link get interconnected. The Python module MediaWiki<sup>21</sup> was used for scraping the pages.

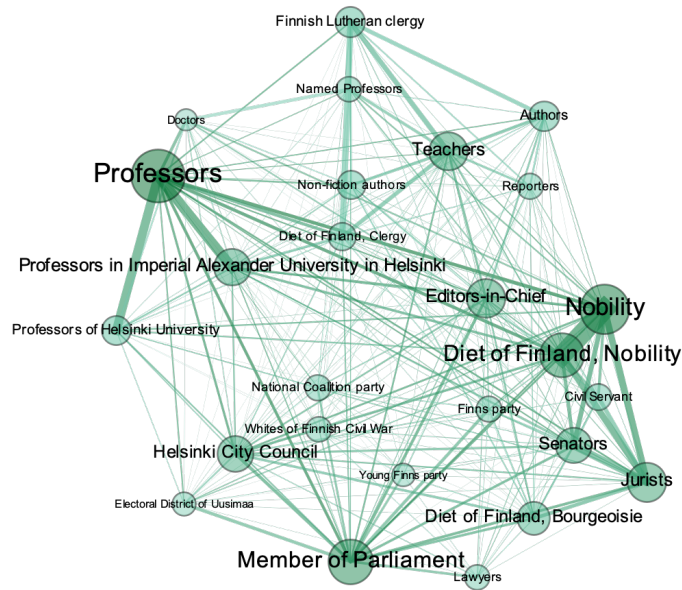
In this graph also the properties of the links can be analyzed, e.g., what are the connections based on related people, places, vocations, or organizations. The most frequent classes of the links are shown in Table 3 indicating that in most cases two people are connected by a mutual reference to a third person. Many of the referenced people are Finnish contemporaries but this information can also reveal clusters of students who became authors and were influenced, e.g., by *Goethe*, *Ovid* or *Aesop*. Besides references to people, in other cases the links are generated by references to places or more rarely by an organization, academic discipline or degree, ideology, historical event, a work of art, or a style of art or literature. Among the organizations is the Finnish Medical Society *Duodecim*<sup>22</sup> which was founded in 1881 by 12 Finnish physicians who are all in the AcademySampo database.

In addition to the links in Wikipedia, the information about the categories can be utilized in analyses. Figure 7 depicts a graph where the weight of an edge connecting two categories equals the number of people belonging to both. Observing the graph shows that the most common categories are *Professors*, *Nobility*, and *Member of (Finnish) Parliament*.

<sup>20</sup><https://biografiasampo.fi>

<sup>21</sup><https://pymediawiki.readthedocs.io/en/latest/code.html#api>

<sup>22</sup>[https://fi.wikipedia.org/wiki/Suomalainen\\_L%C3%A4%C3%A4k%C3%A4riseura\\_Duodecim](https://fi.wikipedia.org/wiki/Suomalainen_L%C3%A4%C3%A4k%C3%A4riseura_Duodecim)



**Figure 7:** Categories based on the Wikipedia linkage

**Table 4**

Top actors in example networks by pagerank centrality

	Student-teacher	Wikipedia
1	Porthan, Henrik G.	Mannerheim, Carl G. E.
2	Gyldenstolpe, Mikael	Paasikivi, Juho K.
3	Scarlin, Algot	Mechelin, Leopold H. S.
4	Hassel, Henrik	Leino, Eino
5	Gadolin, Jakob	Sibelius, Jean

### 3.5. Analyzed Networks

Table 4 depicts the top five actors by their Pagerank centrality in the teacher-student and Wikipedia networks analyzed in the previous section. The central actors in the teacher-student network are the same as in Figure 6 while the central actors in the Wikipedia network are well-known Finnish people of politics and culture. The networks constructed by genealogical relations and similarity values are different in their nature, so applying a social network statistic to them would not reveal useful results.

Table 5 contains general metrics of the four networks, (1) The teacher-student relations, (2) the genealogical network, (3) the network based on actor similarities, and (4) network

**Table 5**

Comparison between the four networks (Similarity, teacher-student, Families, and Wikipedia) in the AcademySampo, *BiographySampo*, and *Email* datasets using network measures

Measure	Similarity	Student-teacher	Families	Wikipedia	<i>Biographies</i>	<i>Email</i>
edges	20000	4893	9380	24988	2741	2396
nodes	20418	3159	12183	2628	-	-
density	0.98	1.55	0.77	9.50	5.48	4.79
average degree	1.01	3.36	1.54	1.52	-	-
HD	17.62	231	10	35.73	323	499
max clique	4	4	2	17	-	-
diameter	103	15	12	8	5	7
GCC	0.04	0.05	0	0.34	0.35	0.54
components	2269	6	2806	3	-	-
giant component	9338	3146	66	2624	-	-
APL	31.70	6.24	6.08	3.09	2.76	1.98
alpha ( $\alpha$ )	1.63	1.38	2.01	1.33	1.43	1.87

constructed based on Wikipedia pages. For comparison also the available measures from the *BiographySampo* reference network (*Biographies*) described in [16] as well as the EU Email Communication Network (*Email*) analyzed by Hashmi et al. [17] are included in the table. This table contains first the numbers of nodes and edges in the network. The Average degree indicates the average amount of links for a single node and highest degree (HD) is the highest node degree in the network. Max clique size is the largest size of a clique. For example, the value 17 indicates that there exists a subgroup of 8 people who all are linked to one another. The table shows the number of separated components in the network, and the size of the largest connected component. The genealogical network is scattered into numerous separated components, while the three reference networks are all more connected having giant components connecting most of the data points. The Diameter is the number of edges along the longest path between any two nodes in the network. The Alpha ( $\alpha$ ) is the constant obtained when a power-law distribution is fitted on the degree distribution of the network [18]. The Global Clustering Coefficient (GCC) is the measure of connected triples; the Average Path Length (APL) is the average number of edges traversed along the shortest paths for all possible pairs of nodes.

The measures provided here are the same as introduced in Hashmi et al., and which were later analyzed in the context of *BiographySampo* data. The four networks are different in their nature: in the network of similarities each node is forced to find a similar pair; the genealogical network is a directed acyclic graph consisting of relatively small connected components; also in the network built by teacher-student relationships the triadic closure (GCC) remains low. However, the measures of the Wikipedia network, specially density, GCC, or diameter, show a small-world behavior of a social network, as measured for *Biographies* and *Email*.



## 4. Discussion

Work on ACADEMYSAMPO is continuation to our earlier biographical LOD systems on Norssit Alumni register [19], the U.S. Congress Prosopographer [20], and BiographySampo [10]. Our earlier articles provide examples of analyses for BiographySampo data [16] and for the Members of the Finnish Parliament in the ParliamentSampo system [21, 22]. Extracting Linked Data from texts has been studied in several works [23]. In [24] language technology was used for extracting entities from biographies and in [25] from news. With epistolary data the social networks can be constructing from the letter exchange information [26, 27, 28, 29].

Representing and analyzing biographical data is a new research and application field. In 2015, the first Biographical Data in Digital World workshop BD2015 was held presenting several works on studying and analyzing biographies as data [30], and the proceedings of BD2017 contain more similar works [31]. In [32], analytic visualizations were created based on U.S. Legislator registry data. The idea of biographical network analysis is related to the Six Degrees of Francis Bacon system<sup>23</sup> [33, 34] that utilizes data of the Oxford Dictionary of National Biography. In earlier research, sociocentric and egocentric networks connecting the actors could be constructed from texts based on, e.g., mentioned names, hypertext links, genealogical relations, or similarities in characteristics, such as lifetime events [2, 35].

This paper presented the idea of creating various networks of academic students based on linkage in Linked Data. It was also shown that networks based on different approaches can reveal different phenomena from the data.

**Acknowledgements** Yrjö Kotivuori and Veli-Matti Autio authored the original data publications used in our work. Our work is related to the EU project InTaVia: In/Tangible European Heritage<sup>24</sup>. CSC – IT Center for Science has provided computational resources for the work.

## References

- [1] P. Leskinen, E. Hyvönen, Linked open data service about historical Finnish academic people in 1640–1899, in: DHN 2020 Digital Humanities in the Nordic Countries. Proceedings of the Digital Humanities in the Nordic Countries 5th Conference, CEUR Workshop Proceedings, Vol. 2612, 2020, pp. 284–292. URL: <http://ceur-ws.org/Vol-2612/short14.pdf>.
- [2] P. Leskinen, H. Rantala, E. Hyvönen, Analyzing the Lives of Finnish Academic People 1640–1899 in Nordic and Baltic Countries: AcademySampo Data Service and Portal, in: DHNB 2022 The 6th Digital Humanities in Nordic and Baltic Countries Conference, CEUR Workshop Proceedings, long papers, Vol. 3232, 2022. URL: <http://ceur-ws.org/Vol-3232/paper07.pdf>.
- [3] E. Hyvönen, J. Tuominen, M. Alonen, E. Mäkelä, Linked Data Finland: A 7-star Model and Platform for Publishing and Re-using Linked Datasets, in: ESWC 2014: The Semantic Web: ESWC 2014 Satellite Events, Springer-Verlag, 2014, pp. 226–230. doi:10.1007/978-3-319-11955-7\_24.

---

<sup>23</sup><http://www.sixdegreesoffrancisbacon.com>

<sup>24</sup><https://intavia.eu/>

- [4] K. Verboven, M. Carlier, J. Dumolyn, A short manual to the art of prosopography, in: *Prosopography approaches and applications. A handbook*, Unit for Prosopographical Research (Linacre College), 2007, pp. 35–70. doi:1854/8212.
- [5] L. Rietveld, R. Hoekstra, The YASGUI family of SPARQL clients, *Semantic Web – Interoperability, Usability, Applicability* 8 (2017) 373–383. doi:10.3233/SW-150197.
- [6] E. Hyvönen, Digital humanities on the semantic web: Sampo model and portal series, *Semantic Web – Interoperability, Usability, Applicability* 14 (2023) 729–744. doi:10.3233/SW-223034.
- [7] E. Hyvönen, Linked open data infrastructure for digital humanities in Finland, in: *DHN 2020 Digital Humanities in the Nordic Countries. Proceedings of the Digital Humanities in the Nordic Countries 5th Conference*, CEUR Workshop Proceedings, vol. 2612, 2020, pp. 254–259. URL: <http://ceur-ws.org/Vol-2612/short10.pdf>.
- [8] P. Leskinen, E. Hyvönen, Reconciling and using historical person registers as linked open data in the AcademySampo knowledge graph, in: *The Semantic Web – ISWC 2021*, Springer-Verlag, 2021, pp. 714–730. doi:10.1007/978-3-030-88361-4\_42.
- [9] T. Heath, C. Bizer, *Linked Data: Evolving the Web into a Global Data Space* (1st edition), *Synthesis Lectures on the Semantic Web: Theory and Technology*, Morgan & Claypool, 2011. URL: <http://linkeddatabook.com/editions/1.0/>.
- [10] E. Hyvönen, P. Leskinen, M. Tamper, H. Rantala, E. Ikkala, J. Tuominen, K. Keravuori, *BiographySampo – Publishing and Enriching Biographies on the Semantic Web for Digital Humanities Research*, in: *The Semantic Web. ESWC 2019*, Springer-Verlag, 2019, pp. 574–589. doi:10.1007/978-3-030-21348-0\_37.
- [11] A. A. Hagberg, D. A. Schult, P. J. Swart, Exploring Network Structure, Dynamics, and Function using NetworkX, in: G. Varoquaux, T. Vaught, J. Millman (Eds.), *Proceedings of the 7th Python in Science Conference*, Pasadena, CA USA, 2008, pp. 11 – 15.
- [12] M. Bastian, S. Heymann, M. Jacomy, Gephi: An Open Source Software for Exploring and Manipulating Networks, in: *Third international AAAI conference on weblogs and social media*, 2009. URL: <https://www.academia.edu/download/3244556/gephi-bastian-feb09.pdf>.
- [13] U. Brandes, M. Eiglsperger, I. Herman, M. Himsolt, M. S. Marshall, GraphML progress report structural layer proposal, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 2265 LNCS (2002) 501–512. doi:10.1007/3-540-45848-4\_59.
- [14] P. Leskinen, E. Hyvönen, Extracting Genealogical Networks of Linked Data from Biographical Texts, in: *The Semantic Web: ESWC 2019 Satellite Events*, Springer, 2019, pp. 121–125. doi:10.1007/978-3-030-32327-1\_24.
- [15] P. Ristoski, J. Rosati, T. Di Noia, R. De Leone, H. Paulheim, RDF2Vec: RDF graph embeddings and their applications, *Semantic Web* 10 (2019) 721–752.
- [16] M. Tamper, P. Leskinen, E. Hyvönen, R. Valjus, K. Keravuori, Analyzing biography collection historiographically as linked data: Case National Biography of Finland, *Semantic Web – Interoperability, Usability, Applicability* 14 (2023) 385–419. doi:10.3233/SW-222887.
- [17] A. Hashmi, F. Zaidi, A. Sallaberry, T. Mehmood, Are all social networks structurally similar?, in: *Advances in Social Networks Analysis and Mining (ASONAM)*, 2012 IEEE/ACM International Conference on, IEEE, 2012, pp. 310–314. doi:10.1109/asonam.2012.59.

- [18] A. Clauset, C. R. Shalizi, M. E. Newman, Power-Law Distributions in Empirical Data, <http://dx.doi.org/10.1137/070710111> 51 (2009) 661–703. URL: <https://epubs.siam.org/doi/abs/10.1137/070710111>. doi:10.1137/070710111.
- [19] E. Hyvönen, P. Leskinen, E. Heino, J. Tuominen, L. Sirola, Reassembling and Enriching the Life Stories in Printed Biographical Registers: Norssi High School Alumni on the Semantic Web, in: *Language, Technology and Knowledge*, Springer-Verlag, 2017, pp. 113–119. doi:10.1007/978-3-319-59888-8\_9.
- [20] G. Miyakita, P. Leskinen, E. Hyvönen, Using Linked Data for Prosopographical Research of Historical Persons: Case U.S. Congress Legislators, in: *7th International Conference, EuroMed 2018, Proc., Part II*, Springer-Verlag, 2018, pp. 150–162. doi:10.1007/978-3-030-01765-1\_18.
- [21] P. Leskinen, E. Hyvönen, J. Tuominen, Members of Parliament in Finland Knowledge Graph and Its Linked Open Data Service, in: *Further with Knowledge Graphs. Proceedings of the 17th International Conference on Semantic Systems, 6-9 September 2021, Amsterdam, The Netherlands*, IOS Press, 2021, pp. 255–269. URL: <https://ebooks.iospress.nl/volumearticle/57420>. doi:10.3233/SSW210049.
- [22] H. Poikkimäki, P. Leskinen, M. Tamper, E. Hyvönen, Analyses of Networks of Politicians Based on Linked Data: Case ParliamentSampo – Parliament of Finland on the Semantic Web, in: *Semantic Web and Ontology Design for Cultural Heritage (SWODCH 2022)*, Turin, Italy, Proceedings, CEUR WS Proceedings, 2022. URL: <https://seco.cs.aalto.fi/publications/2022/poikkimaki-et-al-2022.pdf>, accepted.
- [23] J. L. Martinez-Rodriguez, A. Hogan, I. Lopez-Arevalo, Information Extraction Meets the Semantic Web: A Survey, *Semantic Web – Interoperability, Usability, Applicability* 11 (2020) 255–335.
- [24] A. Fokkens, S. ter Braake, N. Ockeloen, P. Vossen, S. Legêne, G. Schreiber, V. de Boer, BiographyNet: Extracting Relations Between People and Events, in: *Europa baut auf Biographien*, New Academic Press, Wien, 2017, pp. 193–224.
- [25] M. Rospocher, M. van Erp, P. Vossen, A. Fokkens, I. Aldabe, G. Rigau, A. Soroa, T. Ploeger, T. Bogaard, Building event-centric knowledge graphs from news, *Web Semantics: Science, Services and Agents on the World Wide Web* 37 (2016) 132–151.
- [26] W. Ravenek, C. v. d. Heuvel, G. Gerritsen, *The ePistolarium: Origins and Techniques*, JSTOR (2017). URL: <https://www.jstor.org/stable/j.ctv3t5qjk.33>.
- [27] A. Rockenberger, E. Nessheim Wiger, M. Refslund Witting, H. Bøe, E. Irene Thor, O. Wolden, M. Paasche, O. Søndena, *NorKorr – Norwegian Correspondences and Linked Open Data*, in: *Digital Humanities in the Nordic Countries 2019*, 2019. URL: <https://munin.uit.no/handle/10037/15862>, poster paper.
- [28] S. Dumont, *correspSearch -- Connecting Scholarly Editions of Letters*, *Journal of the Text Encoding Initiative* (2016). URL: <https://doi.org/10.4000/jtei.1742>. doi:10.4000/jtei.1742.
- [29] E. Hyvönen, P. Leskinen, J. Tuominen, LetterSampo – Historical Letters on the Semantic Web: A Framework and Its Application to Publishing and Using Epistolary Data, *Journal on Computing and Cultural Heritage* 14 (2023) 1–24. doi:10.1145/3569372.
- [30] S. ter Braake, R. S. Anstke Fokkens, T. Declerck, E. Wandl-Vogt (Eds.), *BD2015, Biographical Data in a Digital World 2015*, CEUR Workshop Proceedings, Vol-1399, 2015. URL: <http://>

[//ceur-ws.org/Vol-1399/](http://ceur-ws.org/Vol-1399/).

- [31] A. Fokkens, S. ter Braake, R. Sluijter, P. Arthur, E. Wandl-Vogt (Eds.), *BD2017 Biographical Data in a Digital World 2015*, CEUR Workshop Proceedings, Vol-2119, 2017. URL: <http://ceur-ws.org/Vol-2119/>.
- [32] R. Larson, *Bringing Lives to Light: Biography in Context*, 2010. Final Project Report, University of Berkeley, [http://metadata.berkeley.edu/Biography\\_Final\\_Report.pdf](http://metadata.berkeley.edu/Biography_Final_Report.pdf).
- [33] C. Warren, D. Shore, J. Otis, L. Wang, M. Finegold, C. Shalizi, *Six degrees of Francis Bacon: A statistical method for reconstructing large historical social networks*, *Digital Humanities Quarterly* 10 (2016).
- [34] A. Langmead, J. Otis, C. Warren, S. Weingart, L. Zilinski, *Towards Interoperable Network Ontologies for the Digital Humanities*, *Int. J. of Humanities and Arts Computing* 10 (2016) 22–35.
- [35] D. K. Elson, K. McKeown, N. J. Dames, *Extracting Social Networks from Literary Fiction*, *aclweb.org* (2010). URL: <https://www.aclweb.org/anthology/P10-1015.pdf>.

# Studying Occupations and Social Measures of Perished Soldiers in WarSampo Linked Open Data

Mikko Koho<sup>1</sup>, Eero Hyvönen<sup>1,2</sup>

<sup>1</sup>*Semantic Computing Research Group (SeCo), Aalto University, Espoo, Finland*

<sup>2</sup>*HELDIG – Helsinki Centre for Digital Humanities, University of Helsinki, Helsinki, Finland*

## Abstract

WarSampo Knowledge Graph contains data about Finland in the Second World War as Linked Open Data, including metadata of more than 100 000 people. A crucial part is the casualty register, containing detailed information about all Finnish soldiers who perished during the war, consisting of 94 676 person records. This register contains occupational labels, which have been manually harmonized into an ontology and linked to occupational classifications such as HISCO. This paper gives an overview of the harmonized occupation ontology and provides an outlook of how the ontology, the occupation classifications, and related social measures could be used for prosopographical study in the future to provide new insights into events of the war or of the Finnish society.

## Keywords

Linked Data, Digital Humanities, Data Analysis, Data Visualization, Prosopography, HISCO

## 1. Introduction

*WarSampo – Finnish Second World War (WW2) on the Semantic Web* [1] collects, integrates, and harmonizes data about Finland in WW2 and publishes the resulting *Knowledge Graph (KG)* as *Linked Open Data (LOD)* [2]. The data is available via a public web portal<sup>1</sup> for searching, browsing, and analyzing the data through nine different perspectives, as well as an open SPARQL endpoint for direct data access. Since its opening in 2015, the WarSampo portal has been used by more than a million end users, corresponding to almost 20% of the population of Finland.

The core dataset of WarSampo is the casualty register [3] of the National Archives of Finland, containing detailed information about all Finnish soldiers who perished during the war, consisting of 94 676 person records. This data is enriched with interlinked datasets about, e.g., military units, war diaries, wartime photographs, historical places, historical maps, and war-time events. WarSampo also contains metadata of all known Finnish prisoners of war (POW) (4200 persons) [4] and notable individuals (5611) from other data sources, such as Wikipedia [5].

The WarSampo KG enables seeking new insights about WW2, the arguably most devastating catastrophe in human history. The ontology and data infrastructure of WarSampo can be further


---

*Biographical Data in a Digital World 2022 (BD 2022) Workshop, co-located with the Digital Humanities 2022 (DH2022) conference on July 25th.*

✉ mikko.koho@aalto.fi (M. Koho); eero.hyvonen@aalto.fi (E. Hyvönen)

🌐 <https://seco.cs.aalto.fi/u/mkoho/> (M. Koho)

🆔 0000-0002-7373-9338 (M. Koho); 0000-0001-7116-9338 (E. Hyvönen)

 © 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

[https://doi.org/10.3986/9789610508120\\_08](https://doi.org/10.3986/9789610508120_08)

<sup>1</sup>WarSampo Portal: <http://sotasampo.fi/en/>



extended with new data to enable digging even deeper into the societal research questions which interest many military history scholars today [6, 7]. Towards this end, considerable effort was put in harmonizing the occupational labels in the person registers [8] to enable prosopographical study of the persons based on their occupational groups and social measures. The devised occupation ontology AMMO [9] is aligned with and linked to the *HISCO* historical international standard classification of occupations [10]. *HISCO* provides the hierarchical backbone of occupational groups in AMMO, as well as two social measures [11]: *HISCLASS* [12] for social classes and *HISCAM* [12] for continuous scale social status. AMMO is also aligned with the Finnish *Classification of occupations 1980 (COO1980)* [13], a occupation classification system which has been in use in Finland.

With the enhanced possibilities for information retrieval and data grouping attained from the harmonization and reconciliation of occupations into rich ontologies, the possibilities for answering humanities-driven research questions are enhanced. Exploiting the new possibilities requires understanding about the data provenance, Semantic Web technologies, and computational data analysis, as well as domain knowledge of military historical research, thus making it an interesting case for interdisciplinary Digital Humanities research [14, 15].

This paper extends our earlier publications on WarSampo and AMMO by giving a statistical overview of the harmonized occupation ontology and providing an outlook of how the ontology, the linked occupation classifications, and related social measures could be used for prosopographical study in the future to provide new insights into events of the war or of the surrounding society. The data harmonization processes of AMMO ontology have been presented in [8] and the ontology structure has been presented in [9]. A previous paper [16] has shown how various prosopographical phenomena can be highlighted and visualized. We extend this line of research by focusing on studying the occupations and social measures of the perished soldiers.

## 2. WarSampo Knowledge Graph and LOD Infrastructure

In WarSampo, Linked Data and the event-based *CIDOC Conceptual Reference Model (CRM)* [17] are used as a basis for harmonizing datasets about Finland in the Second World War into a unified KG [2]. Main entity types in the KG are persons, military units, death records, prisoner records, events, places, photographs, war diaries, articles, and occupations. The death and prisoner records were created from the metadata records of the casualty and POW databases of the National Archives, respectively, and were aligned with the WarSampo KG person entities.

The death record data is of great importance in studying Finland in WW2, as it contains detailed information about all of the known perished soldiers in the Finnish fronts, totalling 94 676 persons. The POW register contains data about all 4200 Finnish POWs. In addition, WarSampo contains information of over 5600 notable persons who survived the war, aggregated from many additional data sources.

The occupations of the person registers have been harmonized into an occupation ontology [9], which is linked to *HISCO* historical international standard classification of occupations [10]. *HISCO* provides the hierarchical backbone of occupational groups in AMMO, as well as social stratification information through several measures like *HISCLASS* [18, 12], a *HISCO*-based 12 level social classification system, and *HISCAM* [19, 12], a social interaction distance measure.



AMMO is also aligned with the Finnish Classification of occupations 1980 [13] (COO1980), a social stratification classification system in use in Finland.

There are also person related documents that are linked to the person instances or their military units, including a large collection of some 164 000 wartime photographs, tens of thousands hand-written digitized war diaries, and thousands war veteran magazine articles. These provide further contextual information for people studying, for example, the war paths of their relatives. The latest version of the WarSampo KG is always available at the Linked Data Finland LDF.fi platform<sup>2</sup> with its SPARQL endpoint and other services [20] and at Zenodo [21].

Using Semantic Web technologies and CIDOC CRM help to create a sustainable and collaborative infrastructure for pursuing historical research [22]. Anyone can link their data to the WarSampo entities, and enrich their data from WarSampo. For example, the domain ontology of people provides a point of access to all of the information about each person contained in WarSampo, making it possible to for anyone to use this information by linking to the person.

Many of the domain ontologies of WarSampo, e.g., occupations, military ranks, and war-time municipalities, are used to provide facet values in the many faceted search-based perspectives of the WarSampo portal. The faceted search user interfaces of the WarSampo Portal provide an easy way for anyone interested in military history to study, explore, and analyze the integrated Knowledge Graph [4]. For example, one can do simple prosopographical data analysis of the person registers. The Casualties perspective of the portal makes use of the harmonized occupational titles of AMMO, but does not currently use the social stratification measures. It would be possible in the future to add facets for occupational groups or associated social classes.

In addition to studying the occupations through the Casualties perspective and Prisoners of War perspective, the portal contains a landing page for each person, as shown in Figure 1, which reconstructs the person's biography based on all of the information relating to a person from various sources [4], including his or her occupations.

The WarSampo infrastructure has recently been employed in the WarMemoirSampo system to provide named entities and contextual information to the things being discussed in war veteran interview videos, such as places, organizations, persons, military units, and events [23].

### 3. Occupations and Social Measures in AMMO Ontology

AMMO<sup>3</sup> contains currently 2258 occupations. These all have HISCO codes attached and 2152 occupations have an COO1980 code, too. The ontology is published on an open SPARQL endpoint<sup>4</sup> and on the ONKI ontology service<sup>5</sup>.

The occupational labels from the source data are grouped to AMMO occupations when there seems to be no semantic difference between several labels. Multiple occupations can also use the same HISCO codes, which means that they would be identical in HISCO terms, unless they make use of the HISCO relation, status, or product codes, which are used to describe the occupation in HISCO.

---

<sup>2</sup>WarSampo Knowledge Graph in LDF.fi: <https://www.ldf.fi/dataset/warsa>

<sup>3</sup>AMMO Ontology: <https://seco.cs.aalto.fi/ontologies/ammo/>

<sup>4</sup>AMMO SPARQL endpoint: <https://ldf.fi/ammo/sparql>

<sup>5</sup>AMMO on the ONKI light ontology service: <https://light.onki.fi/ammo/>

**Persons**

Search for known persons from the past Finnish wars by writing their name in the text input below and/or selecting a person from the list below. Information regarding the person and recommended links will appear on the right. **If you cannot find the person you are looking for, and know in which military unit they have served, you can take a look at the unit's timeline.**

Search by person name

- Aakala, Aarne Eetu (Sergeant)
- Aakko, Juho Vilhelm (Jääkäri)
- Aakkonen, Antti (Sergeant)
- Aakkula, Antero (Major)
- Aakkula, Eino Rikhard (Corporal)
- Aaku, Eero (Captain)
- Aalo, Aulis Antero (Private)
- Aaltio, Erkki
- Aalto, Aarne Ensio (Private)
- Aalto, Aarne Erkki (Private)
- Aalto, Alli Valfrid (Private)
- Aalto, Arvo Aadolf (Lance Corporal)
- Aalto, Arvo Anselm (Private)
- Aalto, Eero August (Lance Corporal)
- Aalto, Einari Johannes (Jääkäri)
- Aalto, Eino Kaarlo (Corporal)
- Aalto, Erkki (Seaman)
- Aalto, Frans Jalmari (Private)
- Aalto, Ilmari (Private)

## Viljo Olavi Einari Salin

Information | Timeline

URI: [http://idf.fi/warsa/actors/person\\_p10098](http://idf.fi/warsa/actors/person_p10098)

### Personal Details

Family name	Salin <sup>[1, 2]</sup>
Given names	Viljo Olavi Einar <sup>[1]</sup> Viljo Olavi Einari <sup>[2]</sup>
Born	10.05.1913 <sup>[1, 2]</sup>
Municipality of birth	Helsinki <sup>[1, 2]</sup>
Municipality of domicile	Helsinki <sup>[1, 2]</sup>
Municipality of residence	Helsinki <sup>[1, 2]</sup>
Citizenship	Finnish <sup>[2]</sup>
Nationality	Finnish <sup>[2]</sup>
Mother tongue	Finnish <sup>[1, 2]</sup>
Marital status	Married <sup>[1, 2]</sup>
Number of children	0 <sup>[2]</sup> 1 <sup>[1]</sup>
Occupation	Shoe merchant <sup>[1]</sup> Carpenter <sup>[1]</sup> Shop keeper <sup>[2]</sup>
Rank	Corporal <sup>[1, 2]</sup>
Military Unit	Pion.K/Rv.Pr. <sup>[1, 2]</sup>

**Figure 1:** The biographical information of a person, including occupations and the information sources, shown on the person's landing page in the WarSampo portal.

### 3.1. HISCO

HISCO contains a hierarchy of 1675 distinct occupational categories, with the purpose to be a classification system that enables comparisons to be made internationally and historically [11]. HISCAM is a continuous historical status scale that is linked to the occupational categories, as is the HISCLASS historical international social class scheme. Several versions of HISCAM exist based on different geographical and temporal datasets. HISCLASS contains 12 social classes, but also simplified versions of 7 and 5 levels are used.

The 12 most commonly used HISCO codes are shown in Table 1, with the total numbers of relation, status, and product codes used for the occupations. The most common occupations seen reflect the variety of occupational labels found in the source person registers relating to the corresponding HISCO occupation codes as well as to the decisions taken in interpreting the labels when creating the occupation ontology and deciding whether several occupational labels can be considered the same occupation in the ontology or whether they are separate occupations but still refer to the same HISCO code. So the HISCO occupation code counts are somehow related to the prevalence of corresponding occupations in the source person registers. However, it is worth noting that for administrative and managerial professions, the exact field of work

**Table 1**

The 12 most common HISCO occupation codes in AMMO with the number of occupations referring to each code, and the numbers of HISCO relation, status, and product codes used for the occupations

HISCO	HISCO label	Occupations	Relations	Statuses	Products
6-11.10	General Farmer	58	26	30	0
9-99.99	<i>Title is too vague</i>	46	13	4	0
-1	<i>Not an occupational title</i>	38	28	8	0
9-99.30	Factory Worker	36	10	0	0
6-21.05	Farm Worker, General	26	5	6	0
8-31.10	Blacksmith, General	24	9	1	0
6-32.20	Forest Supervisor	23	12	2	0
6-11.15	Small Subsistence Farmer (Husbandman)	23	9	13	0
3-93.10	Office Clerk, General	20	6	1	1
4-51.30	Retail Trade Salesperson	19	5	0	5
2-11.10	General Manager	19	6	2	0
8-49.10	Machinery Mechanic, General	19	3	1	1
8-55.10	Electrician, General	19	10	0	0

is often explicitly given in the occupational label, e.g., “bookbindery manager”, “bookprinter manager”, “car paint shop owner”, and “rubber-repairshop’s chief” .

Due to unforeseen issues with the HISCO source data used in AMMO [9], a large amount of HISCO resources missed metadata. For the study of this paper, we have enriched the data with the missing HISCO metadata and additional HISCAM measures provided by the DataLegend project on a SPARQL endpoint [24]. The data in that endpoint does not seem to be perfect either, as it is missing some of the higher level occupational groups, such as the codes “62” and “84”.

### 3.2. Classification of Occupations 1980

The Classification of Occupations (1980) contains 5100 specific Finnish occupational terms arranged hierarchically. It was published in 1980 and does not have a historical dimension but attempts to represent the contemporary occupations in Finland at the time. This classification has roots in the contemporary ISCO classification of the late 20th century, as does HISCO. COO1980 is compatible with several 20th century national censuses. It contains a 5-level social status classification for each occupation, similar to HISCLASS-5.

A look at the most common occupational groups of COO1980 in AMMO are shown in Table 2, which shows the middle-level occupational groups in the three level hierarchy used in COO1980 with the corresponding occupation counts. By surveying the occupations a bit closer, it is evident that there are some errors made in the manual harmonization process. For example, the code 30 “Managerial work in agriculture, forestry and fishing” seems to be the second most common group due to the fact that several occupations of the group with code 31 “Agricultural and horticultural work, animal husbandry” have accidentally ended up in the group 30, such as “livestock caretaker” and “raindeer herder”. The code 91 “Occupation not specified” contains occupational labels that are too vague to fit into any other group, e.g., “workman”.

**Table 2**

The 11 most common COO1980 occupation codes in AMMO with the number of occupations referring to each code

COO1980	Group label	Occupations
65	Iron and metalware work	162
30	Managerial work in agriculture, forestry and fishing	128
91	<i>Occupation not specified</i>	87
67	Wood work	86
11	Administration of private enterprises and organizations	83
72	Food and beverage work	71
31	Agricultural and horticultural work, animal husbandry	69
80	Public safety and protection work	68
63	Smelting, metallurgical and foundry work	65
10	Public administration	59
01	Supervision and executive work in the technical field	59

**Table 3**

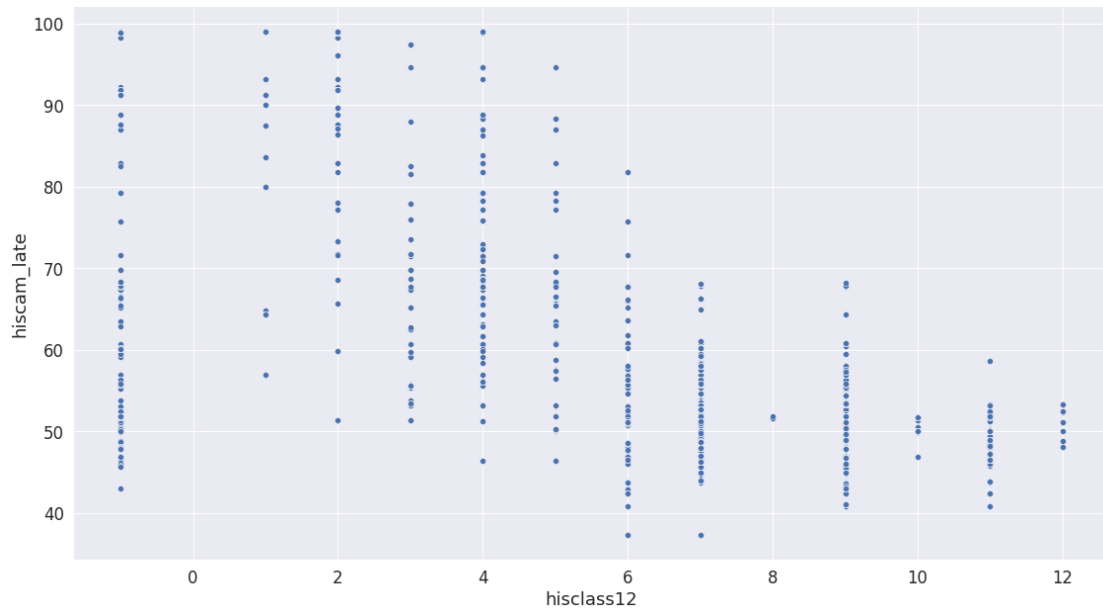
The 10 most common occupations in the Casualties register. The URI namespace is <http://ldf.fi/ammo/>.

URI	Label (fin)	Label (eng)	HISCLASS	Persons
<i>:maanviljelijä</i>	maanviljelijä	farmer	8	22 111
<i>:työmies</i>	työmies	worker	-1	15 319
<i>:sekatyömies</i>	sekatyömies	worker	-1	4761
<i>:maanviljelijän-poika</i>	maanviljelijän poika	farmer's son	8	4602
<i>:maatyömies</i>	maatyömies	farm worker	12	4010
<i>:metsätyömies</i>	metsätyömies	forest worker	12	2181
<i>:autonkuljettaja</i>	autonkuljettaja	driver	9	1523
<i>:talollisen-poika</i>	talollisen poika	farmer's son	8	1447
<i>:opiskelija</i>	opiskelija	student	-1	1364
<i>:kirvesmies</i>	kirvesmies	carpenter	7	1021

## 4. Occupations in the Casualties Register

The casualty register contain 86 069 person records with occupational labels (of 94 676 person records in total). Most records have one occupation, but several records have two or more. The most common occupations in the register are shown Table 3. The HISCLASS column shows the HISCLASS-12 values of the occupations, with level 1 being the elite of the society and 12 being the lowest social class. The -1 (no HISCLASS code) values highlight the issues with mapping occupational labels to pre-existing classifications that rely on knowing the field of work on some level. The two generic worker occupations “työmies” and “sekatyömies” and the general student occupation can not be mapped to any HISCO occupation category. However, it might be possible to map them to 5 or 7 level HISCLASS classification. The very generic worker or manual laborer titles are a known common issue also with census data [11].

Figure 2 shows how the HISCLASS-12 codes and the HISCAM values of the occupations in the casualty register correlate. It is easy to see a correlation with the HISCLASS class, but that correlation is not linear. Many occupations with -1 HISCLASS code have a HISCAM value.



**Figure 2:** The HISCAM and HISCLASS values of all occupations in the Casualties register.

## 5. Discussion

This paper has given an overview of the harmonized occupation ontology and an outlook of how the ontology, the occupation classifications, and related social measures could be used to study the WarSampo person registers. This study helped to pinpoint some issues in the AMMO ontology and develop a solution to fix them and further enrich the ontology with an existing LOD source. Future work will look at how the ontology can be used for prosopographical study using the WarSampo person registers to provide new insights into events of the war or of the surrounding society. HISCO is the most important and widely used standard for historical occupations [11] and the related social measures HISCAM and HISCLASS can be used to study social status and social classes.

In addition to WarSampo, AMMO is used in several other projects to provide occupational groups and information about social stratification, making it an important resource for Digital Humanities research. It is planned to be taken into use in upcoming projects dealing with historical information about Finland. As the current focus is only on early 20th century, one topic of future research is handling occupations changing in time, for example through different interlinked versions of occupations for different time spans.

Wikidata<sup>6</sup> and Finnish KANTO – National Agent Data<sup>7</sup> are becoming relevant LOD sources for occupational labels for historical persons in addition to historical person registers and census data. One aspect of future work is to study and work on compatibility with these sources.

<sup>6</sup>Wikidata: <https://www.wikidata.org/>

<sup>7</sup>Kanto – National Agent Data: <https://www.kiwi.fi/display/Toimijakuvailupalvelu/About+Kanto+in+English>

## Acknowledgments

We wish to acknowledge CSC – IT Center for Science, Finland, for computational resources.

## References

- [1] E. Hyvönen, E. Heino, P. Leskinen, E. Ikkala, M. Koho, M. Tamper, J. Tuominen, E. Mäkelä, WarSampo data service and semantic portal for publishing linked open data about the second world war history, in: H. Sack, E. Blomqvist, M. d'Aquin, C. Ghidini, S. P. Ponzetto, C. Lange (Eds.), *The Semantic Web. Latest Advances and New Domains: 13th International Conference, ESWC 2016*, volume 9678 of *Lecture Notes in Computer Science*, Springer, Cham, 2016, pp. 758–773. doi:10.1007/978-3-319-34129-3\_46.
- [2] M. Koho, E. Ikkala, P. Leskinen, M. Tamper, J. Tuominen, E. Hyvönen, WarSampo knowledge graph: Finland in the second world war as linked open data, *Semantic Web 12 (2021)* 265–278. URL: <https://doi.org/10.3233/SW-200392>. doi:10.3233/SW-200392.
- [3] M. Koho, E. Hyvönen, E. Heino, J. Tuominen, P. Leskinen, E. Mäkelä, Linked death – representing, publishing, and using Second World War death records as linked open data, in: *The Semantic Web: ESWC 2017 Satellite Events*, volume 10577 of *Lecture Notes in Computer Science*, Springer, Cham, 2017, pp. 369–383. doi:10.1007/978-3-319-70407-4\_45.
- [4] M. Koho, E. Ikkala, E. Hyvönen, Reassembling the lives of Finnish prisoners of the Second World War on the Semantic Web, in: *Proceedings of the Third Conference on Biographical Data in a Digital World (BD 2019)*, volume 3152, *CEUR Workshop Proceedings, 2022*, pp. 31–39. URL: [http://ceur-ws.org/Vol-3152/BD2019\\_paper\\_5.pdf](http://ceur-ws.org/Vol-3152/BD2019_paper_5.pdf).
- [5] P. Leskinen, M. Koho, E. Heino, M. Tamper, E. Ikkala, J. Tuominen, E. Mäkelä, E. Hyvönen, Modeling and using an actor ontology of Second World War military units and personnel, in: *The Semantic Web – ISWC 2017: 16th International Semantic Web Conference*, volume 10588 of *Lecture Notes in Computer Science*, Springer, Cham, 2017, pp. 280–296. doi:10.1007/978-3-319-68204-4\_27.
- [6] T. D. Biddle, R. M. Citino, The role of military history in the contemporary academy, *Foreign Policy Research Institute Footnotes (2015)* 1–6. URL: [https://www.fpri.org/docs/society\\_for\\_mil\\_hist\\_whit\\_paper.pdf](https://www.fpri.org/docs/society_for_mil_hist_whit_paper.pdf).
- [7] J. Black, *Rethinking Military History*, Routledge, 2004.
- [8] L. Gasbarra, M. Koho, I. Jokipii, H. Rantala, E. Hyvönen, An ontology of Finnish historical occupations, in: *The Semantic Web: ESWC 2019 Satellite Events*, *Lecture Notes in Computer Science*, Springer, Cham, 2019, pp. 64–68. URL: [https://link.springer.com/chapter/10.1007/978-3-030-32327-1\\_13](https://link.springer.com/chapter/10.1007/978-3-030-32327-1_13).
- [9] M. Koho, L. Gasbarra, J. Tuominen, H. Rantala, I. Jokipii, E. Hyvönen, AMMO ontology of Finnish historical occupations, in: A. Poggi (Ed.), *Proceedings of the First International Workshop on Open Data and Ontologies for Cultural Heritage*, volume 2375 of *CEUR Workshop Proceedings*, 2019, pp. 91–96.
- [10] M. H. D. Van Leeuwen, I. Maas, A. Miles, *HISCO: Historical International Standard Classification of Occupations*, Leuven University Press, 2002.
- [11] M. H. D. Van Leeuwen, Studying long-term changes in the economy and society using the



- HISCO family of occupational measures, in: Oxford Research Encyclopedia of Economics and Finance, Oxford University Press, 2020. doi:10.1093/acrefore/9780190625979.013.541.
- [12] K. Mandemakers, R. J. Mourits, S. Muurling, C. Boter, I. K. van Dijk, I. Maas, B. V. de Putte, R. L. Zijdemans, P. Lambert, M. H. D. Van Leeuwen, F. van Poppel, A. Miles, HSN standardized, HISCO-coded and classified occupational titles, release 2018.01, IISG, Amsterdam, 2018.
- [13] Statistics Finland, Classification of Occupations 1980, Käsikirjoja / Tilastokeskus, Statistics Finland, Helsinki, 1981.
- [14] S. Graham, I. Milligan, S. Weingart, Exploring big historical data. The historian's macro-scope, Imperial College Press, London, UK, 2015. doi:10.1142/p981.
- [15] A. Burdick, J. Drucker, P. Lunenfeld, T. Presner, J. Schnapp, Digital Humanities, The MIT Press, 2012.
- [16] M. Koho, H. Rantala, E. Hyvönen, Digital humanities and military history: Analyzing casualties of the warsampo knowledge graph, in: K. Berglund, M. L. Mela, I. Zwart (Eds.), DHNB 2022 The 6th Digital Humanities in Nordic and Baltic Countries Conference, volume 3232, CEUR Workshop Proceedings, 2022. URL: <http://ceur-ws.org/Vol-3232/paper29.pdf>.
- [17] M. Doerr, The CIDOC conceptual reference module: An ontological approach to semantic interoperability of metadata, AI Magazine 24 (2003) 75–92. doi:10.1609/aimag.v24i3.1720.
- [18] M. H. D. Van Leeuwen, I. Maas, HISCLASS: A historical international social class scheme, Leuven University Press, 2011.
- [19] P. Lambert, R. Zijdemans, M. H. D. Van Leeuwen, I. Maas, K. Prandy, The construction of HISCAM: A stratification scale based on social interactions for historical comparative research, Historical Methods: A Journal of Quantitative and Interdisciplinary History 46 (2013) 77–89.
- [20] E. Hyvönen, J. Tuominen, M. Alonen, E. Mäkelä, Linked Data Finland: A 7-star model and platform for publishing and re-using linked datasets, in: ESWC 2014: The Semantic Web: ESWC 2014 Satellite Events, Springer, Cham, 2014, pp. 226–230. doi:10.1007/978-3-319-11955-7\_24.
- [21] M. Koho, E. Heino, P. Leskinen, E. Ikkala, M. Tamper, K. Apajalahti, J. Tuominen, E. Mäkelä, E. Hyvönen, WarSampo knowledge graph [data set], 2019. URL: <https://doi.org/10.5281/zenodo.3611322>. doi:10.5281/zenodo.3611322.
- [22] D. Oldman, M. Doer, S. Gradmann, Zen and the art of Linked Data: new strategies for a Semantic Web of humanist knowledge, in: S. Schreibman, R. Siemens, J. Unsworth (Eds.), A New Companion to Digital Humanities, John Wiley and Sons, 2016, pp. 251–273. doi:10.1002/9781118680605.ch18.
- [23] R. Leal, H. Rantala, M. Koho, E. Ikkala, M. Merenmies, E. Hyvönen, WarMemoirSampo: A semantic portal for war veteran interview videos, in: DHNB 2022. The 6th Digital Humanities in Nordic and Baltic Countries Conference, volume 3232, CEUR Workshop Proceedings, 2022, pp. 317–325. URL: <http://ceur-ws.org/Vol-3232/paper30.pdf>.
- [24] R. Hoekstra, A. Meroño-Peñuela, A. Rijpma, R. Zijdemans, A. Ashkpour, K. Dentler, I. Zandhuis, L. Rietveld, The datalegend ecosystem for historical statistics, Journal of Web Semantics 50 (2018) 49–61.

# Traveling with Albrecht Dürer - A Case Study for Uncertainty-Aware Biography Visualization

Florian Windhager<sup>1,\*</sup>, Eva Mayr<sup>1</sup>, Johannes Liem<sup>1</sup>, Jakob Kusnick<sup>2</sup>, Stefan Jänicke<sup>2</sup> and Anja Grebe<sup>1</sup>

<sup>1</sup>University for Continuing Education Krems, Austria

<sup>2</sup>University of Southern Denmark, Odense, Denmark

## Abstract

Synoptic accounts on the ‘life and work’ of artists constitute one of the central genres of art history. Especially for well-known historical figures, multiple biographies accumulate over time which motivates succeeding scholars to establish their contributions and argumentative arcs with an assertive style and to rather downplay the interpretive ambiguities and uncertainties which engulf many aspects and sources of every biography. Aside from many other innovations, digital approaches to biography representation can counteract this tendency and make layers of interpretive, historical complexity explicit and thus also map out biographical controversies about lives and works.

With a biography visualization case study we focus on Albrecht Dürer to showcase novel strategies to communicate relevant stations of his life and works in an integrated fashion, including uncertainties and biographical gaps. For that matter, we will have a look at both his overall biographical trajectory and on an episode known as the “Journey to the Netherlands” (1520–1521). Choosing a narrative representation based on a space-time cube perspective, we visualize both Dürer’s biography and a specific time-span known as his Journey to the Netherlands—and we discuss options how to complement such a descriptive perspective with data quality indicators to highlight biographical uncertainty and ambiguity. As such, we also hope to showcase how the often-heard charge of an inherent ‘positivist bias’ of data visualizations could be inverted and utilized for the explication of interpretive ambiguity and plurality.

## Keywords

Visualization, storytelling, biographical data, uncertainty, art history, digital humanities

## 1. Introduction

Albrecht Dürer (1471–1528) counts among the most important figures of Western art history. Already in his lifetime, he gained international renown and was invited by princes and city dignitaries both north and south of the Alps. Thanks to his own travels and his widely sold prints, his works were quickly spread all over the globe and now form the pride of museums and

---

*Biographical Data in a Digital World 2022 (BD 2022) Workshop, July 25, 2022, Tokio, JP*

\*Corresponding author.

✉ [florian.windhager@donau-uni.ac.at](mailto:florian.windhager@donau-uni.ac.at) (F. Windhager); [eva.mayr@donau-uni.ac.at](mailto:eva.mayr@donau-uni.ac.at) (E. Mayr);

[johannes.liem@donau-uni.ac.at](mailto:johannes.liem@donau-uni.ac.at) (J. Liem); [jkusnick@imada.sdu.dk](mailto:jkusnick@imada.sdu.dk) (J. Kusnick); [stjaenicke@imada.sdu.dk](mailto:stjaenicke@imada.sdu.dk)

(S. Jänicke); [anja.grebe@donau-uni.ac.at](mailto:anja.grebe@donau-uni.ac.at) (A. Grebe)

🌐 <https://www.donau-uni.ac.at/en/florian.windhager> (F. Windhager); <https://www.donau-uni.ac.at/en/eva.mayr>

(E. Mayr); <https://www.donau-uni.ac.at/en/johannes.liem> (J. Liem);

<https://portal.findresearcher.sdu.dk/en/persons/kusnick> (J. Kusnick); <https://imada.sdu.dk/~stjaenicke/> (S. Jänicke);

<https://www.donau-uni.ac.at/en/anja.grebe> (A. Grebe)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

[https://doi.org/10.3986/9789610508120\\_09](https://doi.org/10.3986/9789610508120_09)

collections worldwide [1, 2]. Dürer is surely one of the best documented artists from the early modern times with a fairly large amount of autobiographical material and other primary sources. The first biographical accounts on Dürer were already compiled in the early 16th century and since the mid-18th century, his life and work became regularly the subject of monographic studies and catalogues raisonnés. Dürer's biography has been subject to in-depth study by many modern art historians, drawing on both primary sources and later biographies as well as his vast oeuvre.

Looking more closely at these studies, however, a problem of 'traditional' biographical narratives becomes obvious: Biographical writing in art history constitutes a highly developed and sprawling field of study [3] based on diverse models, methods, and strategies. Biographical accounts range from the most complex and fine-grained monographical studies to compact overviews in art historical lexica. However, due to prevailing epistemic cultures and customs, actually developed since the ancient times, the descriptive aspects of art historical "life and work" surveys are often strongly interwoven with larger argumentative arcs, and with the distinct perspectives and positions art historians want to establish as their unique interpretation. For that matter, a certain assertive style is dominant, which aims for a seamless depiction according to the chosen scholarly perspective, and which often tends to diminish uncertainties, ambiguities, controversies, open questions, or simply the lack of historical sources as well as clearly datable and attributable works of art [2, chapter 2].

## 2. Motivation

Against this background of great amounts of 'traditional' biographical knowledge—interlaced with competing claims of validity and certainty—this article aims to do both: Showcasing how digital methods can make complex biographical descriptions newly accessible by methods of data visualization, and exploring how methods of uncertainty visualization can make interpretive controversies explicit. In contrast to the notion of an inherent 'positivist bias' of digital methods, we consider the self-conscious representation of omnipresent uncertainties and interpretive controversies to be a scholarly desideratum that digital methods can support. By leveraging an extended spectrum of visual encoding options we argue for making data provenance—together with parameters of imprecision, ambiguity or absence of sources—explicit in future visualization-based accounts of biography and prosopography representations. This could help to establish visual transparency on all descriptive levels to allow for a more nuanced assessment and representation of critical-historical, interpretive complexity. Instead of regarding uncertainty as an inconvenience and exception, it should be taken as a basic fact and corner stone in biographical writing and visualization.

## 3. Related Work

The digital transformation of arts and humanities methods brings along fascinating new options to deal with (i.e. to search for, create, curate, analyze, and communicate) cultural data, and to go beyond text-based formats by also visualizing arts and humanities data in various constellations [4, 5]. With regard to art-historical knowledge, the last decades have seen a multitude of

initiatives to build large biography data collections—mostly on a national level [6], which store information on the most important life events of thousands of artists and enable new forms of prosopographical research [7] and allow new ways of visual access to them [8, 9, 10, 11, 12]. Regarding artworks and cultural artifacts, cultural information often accumulates in object-oriented archives—whether for individual artists or in bigger databases which aggregate large numbers of subcollections [13]. Collection visualizations allow for more generous overviews and open up more intuitive ways to explore these object collections [14, 15, 16, 17]. Very few approaches have managed to integrate and mediate both aspects so far, e.g. to represent and also visualize both object and biography information in an integrated fashion, and almost none of them addresses the topic of uncertainties so far [18, 19].

With a current European project (<https://intavia.eu>), a consortium has formed to go beyond the current state of siloed cultural information and to bridge and connect data both across national boundaries, as well as across typological boundaries, so that a transnational and balanced (i.e. object- and biography-oriented) interpretive perspective can arise. For that matter, the emerging InTaVia platform allows querying large existing data collections, but also enriching and refining object or biography data manually, e.g. to either analyze or communicate it with integrated visualizations [20, 21]. In the context of this project, various case studies on individual artists—among them Albrecht Dürer—have been set up to investigate options of data curation and the corresponding visualization design space.

Regarding the work of Albrecht Dürer, many of his works are available in digitized form—even though they are dispersed across multiple museums and their respective databases which generate and operate by means of heterogeneous inventory data and data formats. Currently, a relational database on Dürer is under construction, which draws together large parts of his works together with related persons, institutions, primary sources, and secondary literature [22]. While this data collection focuses on Dürer’s works in terms of tangible objects he created, it also offers a few aspects of biographical information (e.g., information on birth, death, as well as selected person relations) based on structured data available via links to different authority files (e.g., in GND, ULAN). However, as with many other artist databases, this project does not provide a detailed biography of Dürer for art-historical analyses, it does not represent information and data uncertainty in a structured fashion, nor does it integrate options of visual data representation or narration for both work and life.

## 4. Data Creation and Curation

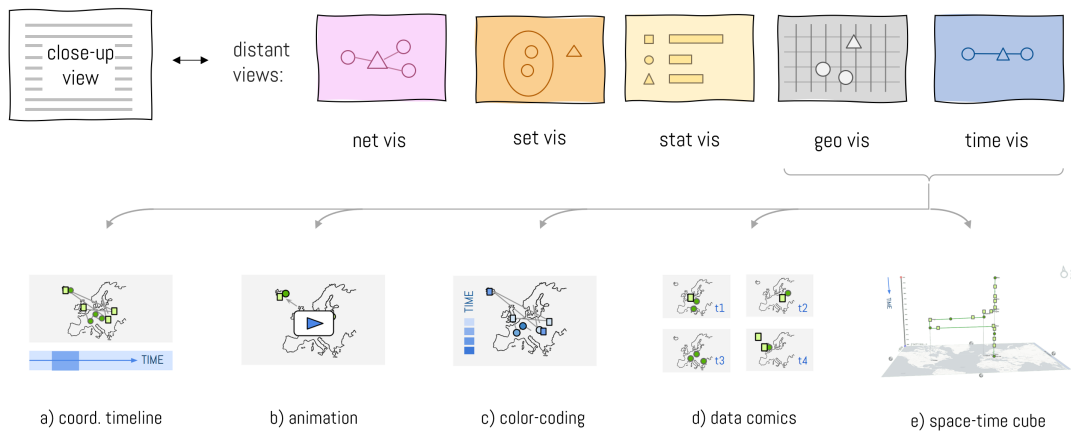
While the InTaVia platform will allow art historians and any other interested person to start their inquiries by querying a transnational knowledge graph, also an open JSON-based data schema is under development that will enable users to import their local data collections on the life and work of specific artists or art schools for a more specific investigation. To simplify the related workflow, a pre-formatted spreadsheet will allow historians or topic experts to fill in biographical data points, consisting of time-stamped biographical events, which are frequently tied to a geographic location, as well as to objects, people, or institutions, and which can be accompanied with any sort of event description or further comments, including remarks on data provenance or data uncertainty. Based on this schema, we manually recreated a slimmed-down

account of Dürer's overall biography (with 29 data points), as well as a fine-grained account of his journey to the Netherlands of 1520–1521, consisting of 158 travel-related events, mostly documented in his so-called travel "diary" as well as by a number of works of art transmitted from this period.

Albrecht Dürer's so-called "Diary of the Journey to the Netherlands" is the most extensive autobiographical travel account by a 16th-century artist. While the original manuscript is lost, its contents have been preserved and transmitted through two handwritten copies from around 1600 considered to be rather close to the original ([23], with a transcription of both copies in the annex). The diary consists of detailed descriptions of places, events, and encounters, but in particular also of the expenses as well as earnings related to the journey. This diary not only allows us to trace Dürer's itinerary very closely, but also to relate places to people, works of art, and various activities, as well as to the duration and costs of those activities. However, due to the predominantly economic purpose of this "diary", Dürer frequently omitted events that were not related to any expenses or earnings. For example, he meticulously lists the costs for food and accommodation as well as all the toll stations during the first part of this journey from Nuremberg to Cologne, which allows to follow his travel route sometimes on an hourly basis. By contrast, for the second part of his outbound trip from Cologne to Antwerp, this sort of information is scarce, which makes it difficult to retrace the entire itinerary in detail. The artist had probably negotiated some sort of fixed rate with the carter, possibly including accommodation in some places, and the toll system was different in this part of the Holy German Empire. In addition, major changes in the topography of this region—now divided between Germany, Belgium, and the Netherlands—make it difficult to identify the ancient place names given in the "diary" in some cases [23, chapter 1].

For more than a century, scholars from various disciplines—i.e. art historians, historians, philologists, geographical historians—have tried in vain to fully reconstruct Dürer's itinerary. Among these, detailed reconstructions of Dürer's itinerary and whereabouts are discussed most frequently and most controversial. Such a reconstruction not only impacts our understanding of Dürer's life but also a better dating and contextualization of his works—and thus his personal and artistic development. Regarding his so-called "first" journey to Italy, contemporary sources do not inform about the day of departure and return, his itinerary, destination(s), or shorter or longer stays during the journey. Up until now, dating of related works of art (mostly watercolor landscapes and architectural views of places in the Trentino region) varies by month and sometimes year. The same is true for Dürer's other travels and especially his journey to the Netherlands. Due to the lacunary information provided by the "diary" and other historical or cartographical sources, many scholars have superposed information with inferences and interpretations. As is often the case with biographical writing, scholarly biographical narratives tend to become more authoritative than the actual "life" of a person—here understood as a sequence of fact-based events. Looking at Dürer's life and his journey to the Netherlands from this more fact-based perspective, the gaps or non-documented periods or phases of his life come into view. Instead of being omitted from the reconstruction of his life, gaps would become equally important and expressive parts of scholarly representations.

In previous monographic studies on Dürer's life and works as well as on his journey to the Netherlands, we have tried to pay specific attention to the issues and challenges posed by biographical gaps and uncertainties [23, 1]. While it is often difficult to fully comply with this



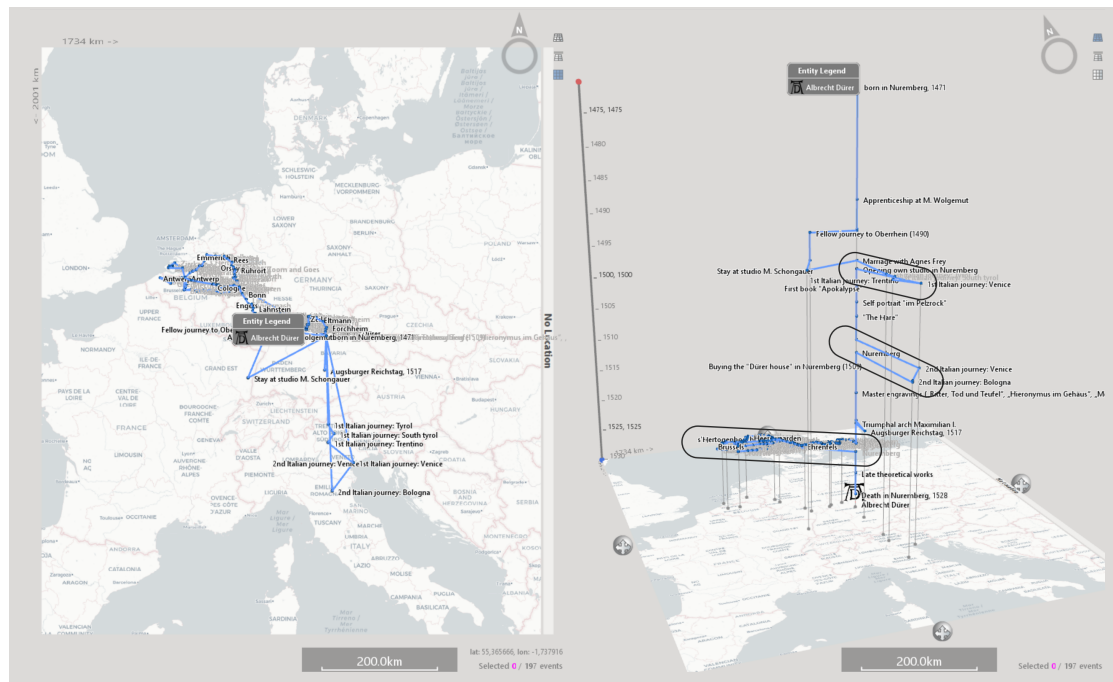
**Figure 1:** Options to visualize the essential time-orientation of biography data (here with focus on geographic space) explored by the InTaVia project, including (from left to right) a coordinated timeline visualization, animation, layer juxtaposition, color-coding, and a space-time-cube perspective.

task in the course of a written reconstruction of an artist’s life or a specific episode from his life, geo-temporal data visualization tools offer novel possibilities to visualize all available and unavailable information by also unveiling biographical lacunae or blank spaces. Ideally, data visualizations thus can help to reduce certain biases in reconstructing an artist’s life and work by treating information and non-information (including uncertainties, ambiguities, and lack of data) as equally important parts of a person’s biography.

## 5. Visualization Case Study

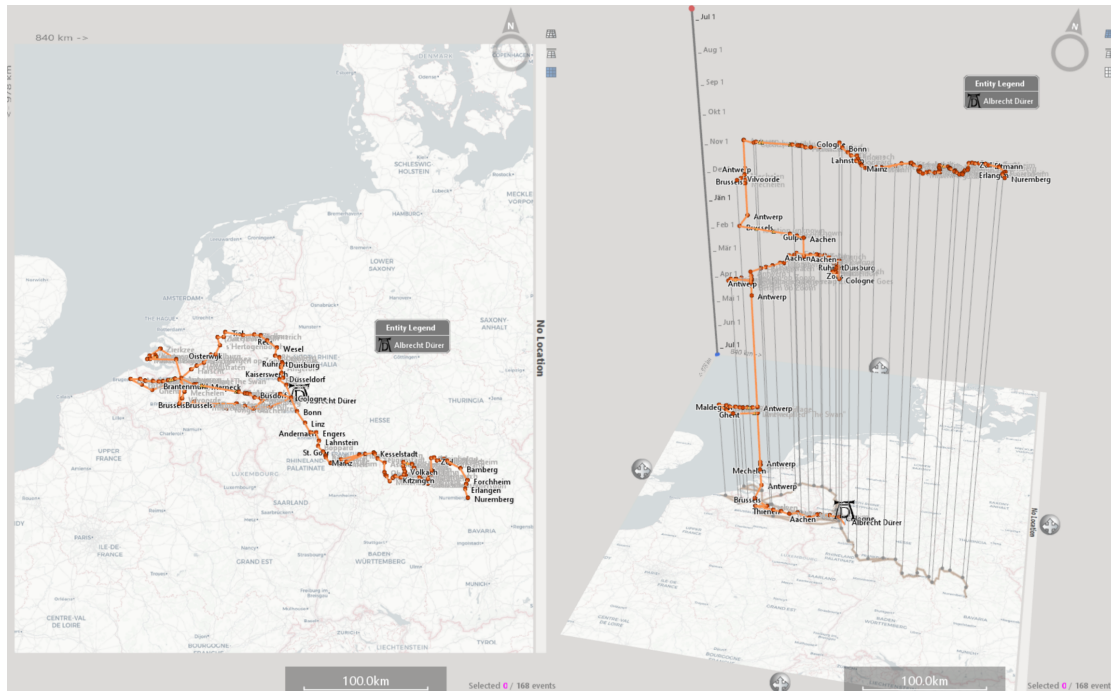
To support the comprehension and reasoning processes of its future users, the InTaVia platform builds up two visualization components: A so-called *Visual Analytics Studio* will enable experts to look at various data selections from multiple data visualization perspectives, including maps, sets, graphs, and faceted timelines (see Figure 1, top). A so-called *Visual Storytelling Suite*, on the other hand, will enable subject matter experts and scholars to use methods of visualization-based storytelling to convey biographical accounts by narrative means to a wide range of non-expert audiences[24]. With the intrinsic time-orientation of biography data playing a central role in all of these various visualizations (i.e., not only from an explicit timeline perspective), the consortium explores multiple options to encode time within other (otherwise non-temporal) types of representations, such as graphs, sets, or maps. For that matter, Figure 1 (bottom) shows various options to incorporate temporal information in maps, which have been discussed and evaluated together with a large group of cultural heritage and (art) history experts, to anchor the project’s design decisions in the principles of user-centered design [21]. From these options, the space-time cube perspective (bottom right) drew a significant amount of interest, which led to the further exploration of this perspective with the case study data of Albrecht Dürer with the GeoTime software package [25].





**Figure 2:** Dürer's biographical trajectory, to be read from top to bottom, from an orthogonal cartographic (left) and a space-time-cube perspective (right).

For the above-mentioned coarse-grained data of Dürer's entire biography, Figure 2 introduces both an orthogonal bird-eye view on his transnational trajectory (left) and a complementary space-time perspective, where the chronological flow of events is mapped from top to bottom. By the means of users' interaction with this view (e.g., by operations of rotating, panning, zooming, as well as drill downs on data selections), the visualization discloses essential temporal information and sequential patterns, which remain hidden from the bird-eye view. Three manual annotations further mark Dürer's three major journeys (two to Italy and one to the Netherlands), which have been deemed an undeniable factor and driver of both the development of Dürer's style and artistic concepts as well as of his transnational reputation [26]. As a matter of fact, Dürer's Italian journeys are largely modern (re-)constructions. This is especially true for his so-called first journey to Italy, which he is thought to have undertaken as a young artist shortly after his marriage in 1494. This journey is only documented by a handful of architectural and landscape views from the Trentino region (which do not necessarily imply a longer stay in Venice) and by a random remark that can be found in one of his letters written a decade later. Only very few scholars have challenged the dominant narrative of Dürer's Italian journeys altogether (i.e. [27]) or put the supposed beginning of the first journey in mid-1494 into question. The latter assumption was first brought up by art historians around 1900 and is now treated as a fixed fact in Dürer's biography which has even made it into his GND entry [28], while it is still part of an active controversy among art historians. When visualizing Dürer's travels and his itinerary it is thus necessary to clearly distinguish documented events from mere suppositions in order not to strengthen any misassumptions.

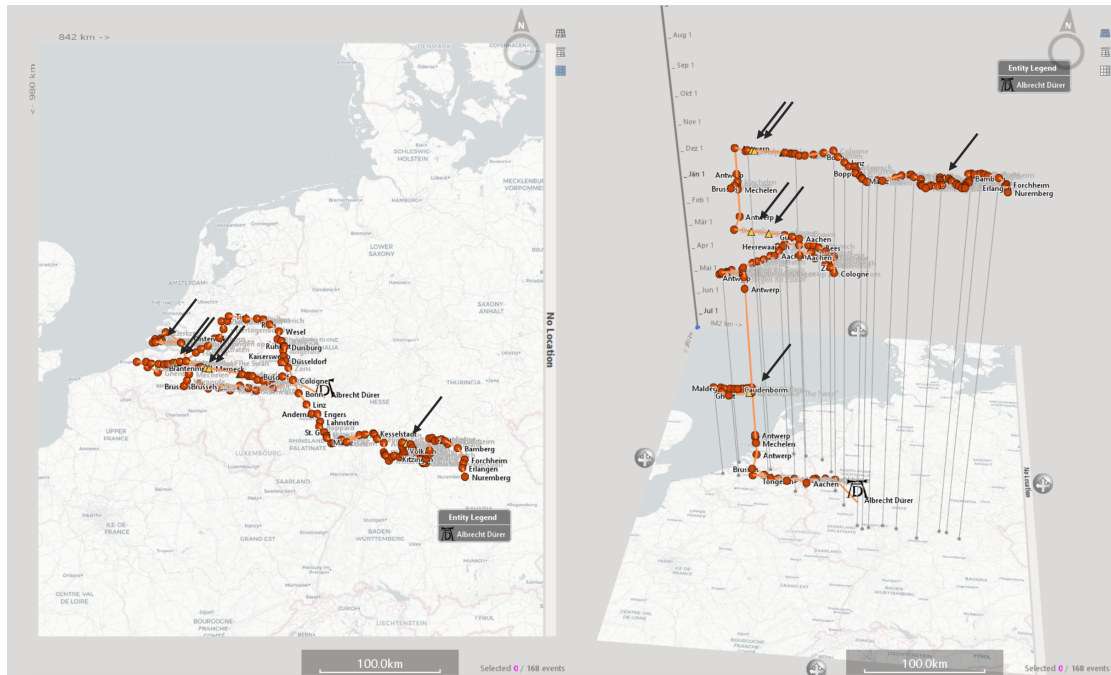


**Figure 3:** The journey to the Netherlands, visualized from an orthogonal (left) and a space-time-cube perspective, the latter to be read from top to bottom.

Dürer’s best documented journey is indeed his journey to the Netherlands thanks to his “Diary” as well as to other contemporary sources. As many of this journey’s events included the production, trading, or strategic donation of artworks to actors of the German, Flemish and Dutch cultural field as well as European princes and members of international merchant houses (from a modern perspective, Dürer aimed to build up his network of clients and customers), the InTaVia project develops methods to also visualize creation events of cultural objects in joint with (previews) of the resulting artifacts (a details on demand-perspective which the GeoTime package is not able to display).

Figure 3 practically zooms in on this third big journey and details the geo-temporal trajectory of the journey to the Netherlands. From top to bottom, it discloses Dürer’s travel patterns, including his main movements and stopover episodes, comprising encounters with princes and dignitaries, creation and exchange of works of art, and notable artworks visited, but also everyday life activities such as the purchase of food or going to the bathhouse.

At the bottom of this representation, we can see the early ending of the documented outbound itinerary of 1520, which results from the above-mentioned lack of diary entries with clearly identifiable locations, leaving room for speculation how the itinerary exactly unfolded from Cologne onwards. The visualization shows the most probable route, based on an average daily distance of ca. 40-60 kilometers.



**Figure 4:** Dürer’s journey to the Netherlands, including indicators for uncertainty about historical place names, marked by yellow triangles on the artist’s trajectory and further emphasized by manually annotated arrows.

## 6. Uncertainty Visualization

Complementing the descriptive perspectives of geo-temporal movements, the InTaVia project will put a specific focus on questions how to make biographical uncertainty, ambiguity, and interpretive controversy explicit—which actually engulf many data points of Dürer’s biography and work, and even more so the biographical accounts of other, less prominent figures of cultural history.

Instead of ignoring or even cutting and ironing out uncertainties, ambiguous or controversial elements will be treated as integral parts of a person’s biography, and will be made visible for experts by superimposing visual data quality indicators on demand. Figure 4 shows how the GeoTime package allows to include different types of icons as data points of the spatio-temporal trajectory. In this specific case, yellow triangles (and manually annotated arrows) mark data points which are only based on “best guesses” about the meaning of historical place names from Dürer’s diary.

Such representations of uncertainty—which remain rather indiscernible in the visual idiom of the GeoTime package—will be further developed by the InTaVia project, so that users can make such ‘negative’ knowledge aspects highly salient on demand. In fact, a whole range of exploratory visualization studies has shown in recent years that a great variety of design options can convey the values and measures of data uncertainty [18]. Figure 5 assembles a variety of these methods which have been introduced in the visualization realm to make varying

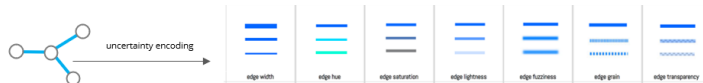
i. Options to visualize **uncertainty of object information** by glyph modifications



ii. Options to visualize **temporal uncertainty**



iii. Options to visualize **relational uncertainty**



**Figure 5:** Options to visualize uncertainty for geographic marks (top), for temporal marks, and for relational marks (bottom).

data quality measures visible for geographic visual marks (top), for temporal marks (e.g., on a timeline view, center), or for relational marks (e.g., encoding the social interactions between persons or the chronological sequences of a trajectory). As a consequence, scholars and users of future biography representations will be able to shift the visual emphasis from a rather seamless depiction of the “best positive guess” about a historical actor’s space-time path to all those structural features which are uncertain, contested, or flat-out missing. From a collective art-historical point of view, this increase of descriptive and visual uncertainty arguably could yield the greatest long-term gains, as representations would not only include all known research gaps, interpretive debates and open questions, but practically also convey other scholars and students with directions for future work.

## 7. Discussion & Outlook

With this exploratory visualization case study on Dürer’s travels, we elaborated on ways and means how to make extensive, existing accounts of artists’ ‘lives and works’ visible to future users of digital, art-historical knowledge bases. While such databases and the development of related visualization methods have seen major progress in recent years, research has also documented a whole spectrum of recurring challenges. Transformed into short descriptive statements, these include:

- *Data is sparse.* Given the omnipresent scarcity and sparsity of preserved information and knowledge about most historical actors, the InTaVia project aims to intertwine scalable options of automated data creation (such as natural-language processing and entity extraction from historical texts) with options for manual data curation. Methods of the latter type will allow to build on the best possible data that automated data creation methods can provide – and to enrich the resulting backbone representations with many

further data points, entities and relations, known to human experts from various other qualitative sources. This is especially important with regard to historical source material which is often lacunary by nature and thus needs data editing and validation by scholars specialized in a certain field of study.

- *Data is uncertain.* Due to the genuine uncertainty and ambiguity that is engulfing many aspects of art-historical and biographical accounts, we do not only see need for corresponding methods of uncertainty visualization (see ch. 6), but also for a related change in academic customs and scholarly writing and working culture. As outlined above, we see the chance of using digital tools to actually foster the collaboration between experts by the means to make uncertainty explicit. Uncertainty, ambiguity, and competing claims of validity are known to have different effects on and relevance for different audiences [29], including the visual and cognitive overload for casual users. By contrast, experts users tend to profit and benefit from the availability of uncertainty indicators, as they do not only strengthen their trust in digital tools [30, 31], but they also allow for more nuanced assessments and historical judgments, while making areas and directions of future work transparent. As such, visual indicators of uncertainty have only started to make their way into biography/prosopography [32] and collection visualization [33, 18], even though their importance for visualization in the digital humanities has been acknowledged for some time [34, 35, 29].
- *Data is autotelic.* Cultural and especially art-historical data includes a significant amount of digital objects with an ‘intrinsic’, aesthetic value and relevance, i.e., of objects which have been collected and preserved over time due to their aesthetic appeal and their self-rewarding value. Among others, this creates the specific requirement to not only complement modern-day distant viewing environments with “generous” close-up views [36]. However, it also creates the necessity for distant viewing tools and their designers to not fall (far) behind the aesthetic qualities of their primary objects when designing ‘scalable’ frames, as the acceptance to work with ‘ugly’ tools is notably limited for aesthetically motivated experts [37], but also for the lion share of casual users or ‘information flaneurs’ which are known to be driven by experiences of emotion, pleasure, and curiosity strongly tied to the unique qualities of objects of art [38].
- *Data often has a complex and convoluted provenance history.* Most data in the digital humanities has its own, significantly complex and convoluted history. To enable and foster a nuanced assessment of related visualizations, the designers of interface to arts and humanities data are well advised to make this kind of provenance information and history explicit, as art history scholars are used to build their interpretation and judgment on these chains of historical references and related inferences.
- *Data has different relevance and affordances for different users.* Art-historical data can be of relevance for a whole variety of users, which have more or less expertise and/or digital skills. While experts are known to ask for analytical and exploratory functions, ‘lay persons’ or casual users are frequently searching for introductory and narrative accounts, with an increased attraction power and a heightened level of engagement and user experience. For that matter, the InTaVia project pairs expert-oriented tools for the visual analysis of art-historical data with tools for storytelling and story consumption [24].



- *Visualizations are non-neutral.* Visualizations and visualization-based tools are late-modern, cultural artifacts themselves, which require not only a basic level of visualization literacy, individual appropriation and interpretation, but also frequently their own share of methodological and epistemological critique. Strategies to support corresponding processes include methods of visualization onboarding [39], tools for individual data curation and annotation of visualizations, as well as means for critical, discursive exchange with both the data, but also with tool developers [40]. The case study presented with this paper was essentially motivated by such question of data and tool critique (including the revision of existing tool critique) and explored ways and means to overcome related challenges with next-generation data and tool development.

Most of these data-related challenges outlined above are also an essential part of ‘traditional’ art-historical research and thus are thoroughly reflected upon in the monographic texts that dominate the scholarly culture of the art-historical field. However, in contrast to the sequential style of ‘traditional’ biographical writing, data visualizations allow for synchronic, multi-level rendering of information. Given a corresponding interest, future representations of “lives and works” of artists thus will be able to draw together manifold biographical interpretations, highlighting both commonalities and interpretive differences. Within or across individual accounts, techniques of uncertainty and provenance visualizations will allow to make ambiguity, interpretive uncertainty, and also the frequent lack of sources explicit. As such, there is hope that the often heard charge of an inherent ‘positivist bias’ of data visualizations [41] could be inverted by deliberately re-appropriating digital tools for the ‘de-positivization’ of traditional, art-historical accounts, and for making the many layers, fractures, gaps, and controversies of bigger art-historical pictures transparent.

## Acknowledgments

This work was funded by the H2020 research and innovation action InTaVia, project No. 101004825.

## References

- [1] A. Grebe, Dürer. Künstler, Werk und Zeit, wbg Academic, 2013.
- [2] A. Grebe, Dürer. Die Geschichte seines Ruhms, Petersberg, 2013.
- [3] K. Hellwig, Von der Vita zur Künstlerbiographie, Akademie Verlag, 2014.
- [4] A. J. Bradley, M. El-Assady, K. Coles, E. Alexander, M. Chen, C. Collins, S. Jänicke, D. J. Wrisley, Visualization and the digital humanities, IEEE computer graphics and applications 38 (2018) 26–38.
- [5] J. Drucker, Visualization and interpretation: Humanistic approaches to display, MIT Press, 2020.
- [6] A. Fokkens, S. Ter Braake, N. Ockeloën, P. Vossen, S. Legêne, G. Schreiber, V. de Boer, Biographynet: Extracting relations between people and events, in: Europa baut auf



Biographien: Aspekte, Bausteine, Normen und Standards für eine europäische Biographik, new academic press, 2018, pp. 193–224.

- [7] M. Kaiser, K. Lejtovicz, M. Schlögl, P. A. Rumpolt, Artist migration through the biographer's lens: A case study based on biographical data retrieved from the austrian biographical dictionary, *Journal of Historical Network Research* 2 (2018) 76–108.
- [8] P. Schmitz, L. Pearce, Humanist-centric tools for big data: berkeley prosopography services, in: *Proceedings of the 2014 ACM symposium on Document engineering*, 2014, pp. 179–188.
- [9] N. Armitage, The biographical network method, *Sociological Research Online* 21 (2016) 165–179.
- [10] E. Hyvönen, P. Leskinen, M. Tamper, H. Rantala, E. Ikkala, J. Tuominen, K. Keravuori, Biographysampo—publishing and enriching biographies on the semantic web for digital humanities research, in: *European Semantic Web Conference*, Springer, 2019, pp. 574–589.
- [11] R. Khulusi, J. Kusnick, J. Focht, S. Jänicke, An interactive chart of biography, in: *2019 IEEE Pacific Visualization Symposium (PacificVis)*, IEEE, 2019, pp. 257–266.
- [12] C. Meinecke, S. Jänicke, Visual analysis of engineers' biographies and engineering branches, in: *LEVIA18: Leipzig Symposium on Visualization in Applications*, 2018.
- [13] B. Valtysson, Europeana: The digital construction of europe's collective memory, *Information, Communication & Society* 15 (2012) 151–170.
- [14] F. Windhager, P. Federico, G. Schreder, K. Glinka, M. Dörk, S. Miksch, E. Mayr, Visualization of cultural heritage collection data: State of the art and future challenges, *IEEE transactions on visualization and computer graphics* 25 (2018) 2311–2330.
- [15] R. Khulusi, J. Kusnick, C. Meinecke, C. Gillmann, J. Focht, S. Jänicke, A survey on visualizations for musical data, *Computer Graphics Forum* 39 (2020) 82–110.
- [16] M. Whitelaw, Generous interfaces for digital cultural collections, *Digital Humanities Quarterly* 9 (2015).
- [17] M. Koolen, J. Kamps, V. de Keijzer, Information retrieval in cultural heritage, *Interdisciplinary Science Reviews* 34 (2009) 268–284.
- [18] F. Windhager, S. Salisu, E. Mayr, Exhibiting uncertainty: Visualizing data quality indicators for cultural collections, *Informatics* 6 (2019) 29.
- [19] F. Windhager, A synoptic visualization framework for artwork collection data and artist biographies, 2020.
- [20] F. Windhager, E. Mayr, M. Schlögl, M. Kaiser, Visuelle analyse und kuratierung von biographiedaten, *Digital History: Konzepte, Methoden und Kritiken Digitaler Geschichtswissenschaft* 6 (2022) 137.
- [21] E. Mayr, F. Windhager, J. Liem, S. Beck, S. Koch, J. Kusnick, S. Jänicke, The multiple faces of cultural heritage: Towards an integrated visualization platform for tangible and intangible cultural assets, in: *Proceedings of the 7th Workshop on Visualization for the Digital Humanities (VIS4DH)*, IEEE Xplore, Oklahoma City, 2022. URL: <https://osf.io/h293m>.
- [22] duerer.online. Virtuelles Forschungsnetzwerk Albrecht Dürer, ??? URL: <https://sempub.ub.uni-heidelberg.de/duerer.online/de>.
- [23] A. Grebe, G. U. Großmann, Albrecht Dürer. Niederländische Reise. "Tagebuch" und Kommentar., Imhof Verlag, 2021.
- [24] J. Kusnick, S. Jänicke, C. Doppler, K. Seirafi, J. Liem, F. Windhager, E. Mayr, Report on narrative visualization techniques for opdb data (d6.1), <https://>

- [//ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5e47d9524&appId=PPGMS](https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5e47d9524&appId=PPGMS), 2021. Deliverable within the H2020 project InTaVia.
- [25] T. Kapler, W. Wright, Geotime information visualization, *Information visualization* 4 (2005) 136–146.
  - [26] S. Foister, P. van den Brink, *Dürer’s Journeys. Travels of a Renaissance Artist.*, New Haven/London, 2021.
  - [27] K. C. Luber, *Albrecht Dürer and the Venetian Renaissance*, Cambridge, 2005.
  - [28] G. U. Großmann, Dürer in Innsbruck. zur Datierung der ersten italienischen Reise, in: G. U. Großmann, F. Sonnenberger (Eds.), *Das Dürer-Haus. Neue Ergebnisse der Forschung.*, Nuremberg, 2007, pp. 227–249.
  - [29] F. Windhager, S. Salisu, G. Schreder, E. Mayr, Uncertainty of what and for whom-and does anyone care? propositions for cultural collection visualization, in: *4th Workshop on Visualization for the Digital Humanities*, 2019.
  - [30] E. Mayr, N. Hynek, S. Salisu, F. Windhager, Trust in information visualization., in: *TrustVis@ EuroVis*, 2019, pp. 25–29.
  - [31] S. Boyd Davis, O. Vane, F. Kräutli, Can i believe what i see? data visualization and trust in the humanities, *Interdisciplinary Science Reviews* 46 (2021) 522–546.
  - [32] W. Zhang, S. Tan, S. Chen, L. Meng, T. Zhang, R. Zhu, W. Chen, Visual reasoning for uncertainty in spatio-temporal events of historical figures, *IEEE Transactions on Visualization and Computer Graphics* 29 (2023) 3009–3023.
  - [33] F. Windhager, V. A. Filipov, S. Salisu, E. Mayr, Visualizing uncertainty in cultural heritage collections., in: *EuroRV<sup>3</sup>@ EuroVis*, 2018, pp. 7–11.
  - [34] R. Therón Sánchez, A. Benito Santos, R. Santamaría Vicente, A. Losada Gómez, Towards an uncertainty-aware visualization in the digital humanities, *Informatics* 6 (2019) 31.
  - [35] G. Panagiotidou, H. Lamqaddam, J. Poblome, K. Brosens, K. Verbert, A. Vande Moere, Communicating uncertainty in digital humanities visualization research, *IEEE Transactions on Visualization and Computer Graphics* 29 (2023) 635–645. doi:10.1109/TVCG.2022.3209436.
  - [36] M. Whitelaw, Generous interfaces for digital cultural collections, *Digital Humanities Quarterly* 9 (2015).
  - [37] H. Lamqaddam, K. Brosens, F. Truyen, R. J. Beerens, I. De Prekel, K. Verbert, When the tech kids are running too fast: Data visualisation through the lens of art history research, *Proceedings of the Workshop on Visualization for the Digital Humanities* (2018).
  - [38] M. Dörk, S. Carpendale, C. Williamson, The information flaneur: A fresh look at information seeking, in: *Proceedings of the SIGCHI conference on human factors in computing systems*, 2011, pp. 1215–1224.
  - [39] C. Stoiber, F. Grassinger, M. Pohl, H. Stitz, M. Streit, W. Aigner, *Visualization onboarding: Learning how to read and use visualizations* (2019).
  - [40] K. Van Es, M. Wieringa, M. T. Schäfer, Tool criticism: From digital methods to digital methodology, in: *Proceedings of the 2nd International Conference on Web Studies*, 2018, pp. 24–27.
  - [41] J. Drucker, Humanities approaches to graphical display, *Digital Humanities Quarterly* 5 (2011) 1–21.