



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^{1,2}, Mikko Koho¹, Angel Daza³, and Gregor Pobežin⁴

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/>