

# Editorial Introduction to Biographical Data in a Digital World 2019

Angel Daza<sup>1</sup>, Antske Fokkens<sup>1,2</sup>, Petya Osenova<sup>3</sup>, Kiril Simov<sup>3</sup>, Alexander Popov<sup>3</sup>, Paul Arthur<sup>4</sup>,  
Thierry Declerck<sup>5</sup>, Ronald Sluijter<sup>6</sup>, Serge ter Braake<sup>7</sup> and Eveline Wandl-Vogt<sup>8</sup>

<sup>1</sup>Vrije Universiteit Amsterdam, <sup>2</sup>Eindhoven University of Technology,

<sup>3</sup>Bulgarian Academy of Sciences, <sup>4</sup>Edith Cowan University,

<sup>5</sup>German Research Center for Artificial Intelligence (DFKI), <sup>6</sup>Huygens ING, <sup>7</sup>Stichting UvO,

<sup>8</sup>Austrian Academy of Sciences, Ars Electronica Research Institute knowledge for humanity

## Abstract

The third edition of the Biographical Data in a Digital World Conference took place on 5 and 6 September 2019 in Varna, Bulgaria. The conference included nine long and four short presentations and a session with work groups on modeling biographical data. These proceedings include nine full papers that were accepted upon single blind review.

## 1 Introduction

The first two editions of *Biographical Data in a Digital World* provided a, for many researchers, first opportunity to connect with other researchers working on digital biographical resources. They brought together a wide variety of perspectives from historians, librarians, literature studies, computer scientists and computational linguistics. Next to differences and new angles, we also found commonalities: a shared interest in the richness and variety of the resources, challenges with gaps in the data and data quality, approaches to data representation and horizons waiting to be explored on several levels.

Projects in various countries are steadily making progress, resources are growing and methods for exploring them are improving. At an international level, we find a continuous wish to connect. As a community, we want to exchange ideas and learn from each other, but in particular, we are interested in identifying connections among our resources. This desire forms the main motivation for continuing to organize events in the BD-conference series, which resulted in the third edition of this conference and (with a long partially Covid-19 related delay) these proceedings.

## 2 Biographical Data in a Digital World 2019

The third edition of *Biographical Data in a Digital World* took place in Varna, Bulgaria in September 2019. Presenters could either submit a full paper or an abstract only. Full papers were reviewed by three or four reviewers in a single-blind review process. Abstracts were verified for relevance. We received ten full papers and five abstracts.

Because interdisciplinary venues bring together different publication cultures, researchers from various domains in digital humanities, may not be familiar with the workshop and conference proceedings publication culture common in computer science. We therefore opted for three modes of acceptance for full papers: regular acceptance (with a chance to make updates and encouragement to incorporate feedback for the camera ready version), acceptance with minor revisions (with a requirement to address the main criticism of reviewers which would be verified by the editors) and major revisions (with an independent review

round of the revised paper which may lead to a decision that the work cannot be included in the proceedings). For all these modes of acceptance, authors could come and present their work. Camera ready versions were collected after the event, so that authors could also make use of feedback from other participants.

Four papersB received a regular accept, three were accepted conditionally based on minor revisions, two papers received a major revision request and one was rejected. All five abstracts were deemed relevant and accepted for presentation (without publication). With one author not being able to make it, there were thirteen presentations around various topics at the event. All accepted papers submitted revised versions. The minor revision papers were checked by the editorial board, which also formed a new reviewing committee for the two remaining papers. One needed an extra round of minor revisions after which all completed papers could be included in the proceedings. An overview of the papers is given below.

In addition to the work presented in this paper, various discussions around connecting resources from different countries took place during the event. This resulted in the proposal of the Horizon2020 project InTaVia<sup>1</sup> that is currently running bringing together researchers from Austria, Denmark, Germany, the Netherlands and Slovenia who met through the Biographical Data conference series.

## 3 Overview of Papers

The papers in these proceedings cover three themes i) Building Digital Biographical Resources ii) applying NLP tools for biographical data mining, and iii) Exploring capabilities of digital resources with case studies.

### Building Digital Biographical Resources

Bhreathnach et al. (2019) identify existing biases, features and omissions in the Irish biographical database *Ainm*.<sup>2</sup> This is an analysis on the distribution of biographies in terms of people's lifespans, gender and birthplace, as well as professions present in the mentioned resource. They

<sup>1</sup><https://intavia.eu>

<sup>2</sup><https://www.ainm.ie/Info.aspx?Topic=resources.en>

found, for example, that even though the resource spans from 15th to 21st centuries, the database has a heavy bias towards people living in the 19th and 20th century (around 75% of the subjects). This work also quantitatively confirms an already acknowledged bias towards people from certain regions and professions.

Hyvönen et al. (2019) argue for shifting towards Linked Data as the paradigm for publishing and using biographical dictionaries. To support their arguments, they describe the data service and semantic portal BiographySampo,<sup>3</sup> where biography texts are enriched with 16 external data sources. Their paper outlines how reasoning through the structured graph can be used to expand and discover serendipitous relations among entities, persons, and places.

Povroznik (2019) studies the collective portraits of deputies of local self-government in Russia in the second half of the XIX century. With this goal in mind, the paper describes the problems of searching, organizing, modelling, analyzing and presenting data during the process for building a useful database for prosopographical research.

Vogeler et al. (2019) discuss the viability of an international prosopographical framework. The authors propose a set of data resources, interfaces and analytical tools to integrate and facilitate the access to prosopographical resources. Their solution is a data model that can be accessed by a RESTful API. The paper illustrates this idea through a concrete example based on resources built for the Austrian Prosopographical database (APIS) by following their proposed methods.

### **NLP Tools for Biographical Data Mining**

Magistry et al. (2019) use Natural Language Processing (NLP) tools to extract information and create a structured data resource based on the biographical Dictionary of Republican China (BDRC). Their approach also uses the constructed graphs as means for exploration on the relationships centered on education and position occupied by the people on the database.

Plum et al. (2019) also apply information extraction methods to identify relevant biography candidates in large databases. The automatic extraction methods are run on two popular data sources: Wikipedia and Wikidata, particularly for the case study of people who had an impact in the Republic of Austria and died between 1951 and 2019. The authors conclude that their NLP pipeline can be helpful to identify suitable candidates and extract relevant information.

### **Exploring Capabilities of Digital Resources with Case Studies**

Filipov et al. (2019) present a visualization tool for interactive analysis on multiple biographical timelines, relating different biographies through a specific set of events and locations. Specifically, they show the example of biographies connected to Austrian music history. Through their visualization technique, new potential narratives can be created and contextualized, providing an extra layer on the process of historical research.

Koho et al. (2019) introduce a database of Finnish prisoners of war in the Soviet Union. The project is targeted to the

researchers of military history. It aims at centralizing all information available from different data sources and linking them to each prisoner. This database facilitates research that explores the details, relations and subgroups of each individual biography.

Mayr et al. (2019) describe research that is part of the Polycube project. Polycube's goal is to visualize multiple data dimensions such as space, categories and relations over time. This paper particularly raises the question of *how visualizations of life and work go together?* They answer it with the case study of Charles W. Cushman showing how these visual-analytical frames of reference could provide a multimodal, narrative framework of biographical knowledge exploration and communication.

We would like to thank the program committee members for their careful and critical reviews, which supported the selection process and helped authors improve their papers. We would also like to thank the other members of the editorial board for their input during discussions and the careful checks and reviews for the conditionally accepted papers. Many special thanks go to the local organizers: Petya Osenova, Kiril Simov and Alexander Popov for an excellent job resulting in an impeccably organized conference. Finally, we would like to thank everyone involved for their patience and understanding for the delays in bringing out these proceedings.

Angel Daza & Antske Fokkens, Chief Editors.

---

<sup>3</sup><https://seco.cs.aalto.fi/projects/biografiasampo/en/>

## **Full Editorial Board**

Angel Daza, Vrije Universiteit Amsterdam  
Antske Fokkens, Vrije Universiteit Amsterdam & Eindhoven University of Technology  
Petya Osenova, Bulgarian Academy of Sciences  
Kiril Simov, Bulgarian Academy of Sciences  
Alexander Popov, Bulgarian Academy of Sciences  
Paul Arthur, Edith Cowan University  
Serge ter Braake, Stichting UvO  
Thierry Declerck, DFKI Saarland  
Ronald Sluijter, Huygens ING  
Eveline Wandl-Vogt, Austrian Academy of Sciences, Ars Electronica Research Institute knowledge for humanity

## **Program Committee**

Peter Bol, Harvard University  
Serge ter Braake, Stichting UvO  
Thierry Declerck, DFKI Saarland  
Lonneke Geerlings, Stichting UvO  
Gernot Howanitz, Universität Innsbruck  
Anders Ingram, University of Oxford  
Bärbel Kröger, Akademie der Wissenschaften zu Göttingen  
Eetu Mäkelä, Aalto University  
Lodewijk Petram, Huygens ING  
Katharina Prager, Ludwig Boltzmann Institute for Digital History  
Matthias Reinert, Neue Deutsche Biographie, Germany  
Matthias Schlägl, Österreichische Akademie der Wissenschaften  
Kiril Simov, Bulgarian Academy of Sciences  
Ronald Sluijter, Huygens ING  
Petra Vide Ogrin, Slovenian Academy of Sciences and Arts  
Georg Vogeler, Österreichische Akademie der Wissenschaften  
Marcos Zampieri, Rochester Institute of Technology  
Kalliopi Zervanou, Leiden University  
Joris van Zundert, Huygens ING

## 4 References

- Úna Bhreathnach, Cathal Burke, Jeaic Mag Fhinn, Gearóid Ó Cleircín, and Brian Ó Raghallaigh. 2019. A quantitative analysis of biographical data from ainm, the irish-language biographical database. In *Proceedings of the Third Conference on Biographical Data in a Digital World 2019*, Varna, Bulgaria. CEUR.
- Velitchko Filipov, Nathalie Soursos, Viktor Schetinger, Susana Zapke, and Silvia Miksch. 2019. Exiled but not forgotten: Investigating commemoration of musicians in vienna after 1945 through visual analytics. In *Proceedings of the Third Conference on Biographical Data in a Digital World 2019*, Varna, Bulgaria. CEUR.
- Eero Hyvönen, Petri Leskinen, Minna Tamper, Heikki Rantala, Esko Ikkala, Jouni Tuominen, and Kirsi Keravuori. 2019. Linked data – a paradigm shift for publishing and using biography collections on the semantic web. In *Proceedings of the Third Conference on Biographical Data in a Digital World 2019*, Varna, Bulgaria. CEUR.
- Mikko Koho, Esko Ikkala, and Eero Hyvönen. 2019. 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 2019*, Varna, Bulgaria. CEUR.
- Pierre Magistry, Cécile Armand, and Christian Henriot. 2019. Mining the biographical dictionary of republican china. from print to network exploration. In *Proceedings of the Third Conference on Biographical Data in a Digital World 2019*, Varna, Bulgaria. CEUR.
- Eva Mayr, Saminu Salisu, Velitchko A. Filipov, Günther Schreder, Roger A. Leite, Silvia Miksch, and Florian Windhager. 2019. Visualizing biographical trajectories by historical artifacts: A case study based on the photography collection of charles w. cushman. In *Proceedings of the Third Conference on Biographical Data in a Digital World 2019*, Varna, Bulgaria. CEUR.
- Alistair Plum, Marcos Zampieri, Constantin Orăsan, Eveline Wandl-Vogt, and Ruslan Mitkov. 2019. Large-scale data harvesting for biographical data. In *Proceedings of the Third Conference on Biographical Data in a Digital World 2019*, Varna, Bulgaria. CEUR.
- Nadezhda Povroznik. 2019. Reconstructing data for modelling collective biography: A case of zemstvo deputies in russia in the second half of xix century. In *Proceedings of the Third Conference on Biographical Data in a Digital World 2019*, Varna, Bulgaria. CEUR.
- Georg Vogeler, Gunter Vasold, and Matthias Schlögl. 2019. Data exchange in practice: Towards a prosopographical api. In *Proceedings of the Third Conference on Biographical Data in a Digital World 2019*, Varna, Bulgaria. CEUR.