

Introduction: Biographical Data in a Digital World

Serge ter Braake, Antske Fokkens, Ronald Sluijter, Thierry Declerck and Eveline Wandl-Vogt

VU University Amsterdam, VU university Amsterdam, Huygens Institute for History of the Netherlands,
Deutsche Forschungszentrum für Künstliche Intelligenz, Austrian Academy of Sciences

s.ter.braake@vu.nl, antske.fokkens@vu.nl, ronald.sluijter@huygens.knaw.nl, declerck@dfki.de, eveline.wandl-vogt@oeaw.ac.at

Abstract

BD2015 brought together researchers from Computer Science and History working on digitized biographies from all over the world. This introduction to its proceedings provides our motivation for a conference on this topic, a short history of the origin of the event and an outline of the selection process that led to the presentations held during BD2015 and, ultimately, to these proceedings.

Keywords: digital humanities, biographical data, biographies

1. Introduction

‘One human life, indeed, is so insignificant, because there are so many of them’¹. Almost seventy years ago, Dutch historian and biographer Jan Romein nailed with these words one of the main issues academic biographers have to deal with: what exactly does the account of this particular life contribute to our understanding of history or historical processes in general? An individual is, after all, only one out of the billions of people living now (2 billion in Romein’s time) and the many billions who lived before.

Because of their focus on a seemingly limited topic, biographers in academia are asked more frequently than others to justify their work. Often they would say that their individual had a huge impact on the course of history in general, or that the individual offers a lens through which we can look at a past society as a whole. These justifications are valid enough, but it also implies that there could be an almost limitless number of biographies worthy of general interest. Many influential people from the past do not have a full biography and we can look at societies at large through the eyes of practically any individual who left enough traces to reconstruct their life. Regardless of this huge potential of interesting biographical contributions to history, the range of individuals who are studied is very limited. Often people who are already famous and have many biographies dedicated to them attract more scholarly attention than individuals who have been treated relatively marginally in historiography.

The ‘digital turn’ opens new horizons for biographers and historians. There is information about billions of individuals online which could be processed, networks can be researched and visualized in new ways and the printed monograph is not the exclusive mode of presentation anymore. For computer scientists, biographical data are attractive to work with, because they usually are relatively well-structured and because individuals always share certain characteristics (e.g. all individuals have a gender, a date of birth, parents and a place of death, and many individu-

als also have siblings, a career and children). Questions of methodology and technology go hand in hand, often in interdisciplinary research groups, in an attempt to facilitate better and enhanced biographical research.

2. Biographical data in a Digital World

To our knowledge, *Biographical Data in a Digital World/BD2015* was the first event attempting to bring researchers from all over the world together to discuss the possibilities, limitations and opportunities of research with biographical data with digital humanities technology. Looking backwards it is difficult to pinpoint exactly when this idea was brought to life for the first time, but we believe a meeting between Antske Fokkens, Thierry Declerck and Eveline Wandl-Vogt at LREC 2014 in Reijkjavic (Ninth International Conference on Language Resources and Evaluation) was vital for setting up the collaboration between the VU University Amsterdam, Huygens ING, DFKI and the Austrian Academy of Sciences. The managers of the BiographyNet project ² at the VU University gave their support to organize this event, originally planned as a workshop, in Amsterdam. Biographical Data in a Digital World (short: BD2015) would be organized by the previously mentioned three people, all with a computational background, and the historians Serge ter Braake and Ronald Sluijter. The organizers were able to mobilize a group of 23 experts on biographies, history, natural language processing, digital humanities and computer science to be a part of the Program Committee (listed below).

For events on a highly specific topic, it remains to be seen whether enough contributions will be submitted. Instead of the expected (hoped for) 10-15, we received 28 submissions mainly from Western Europe. The papers were reviewed by three independent reviewers from the Program Committee (including the Organizing Committee). Eventually it was decided to be as inclusive as possible and to accept 21 papers. Even though we decided not to look at the final ‘scores’ alone, considering the different reviewing traditions in the diverse fields the program committee members are working in, ultimately the selected papers were

¹‘Een mensenleven, inderdaad, het is zo luttel, omdat er zo veel van zijn’ Jan Romein, *De Biografie. Een inleiding* (Amsterdam 1946)

²<http://www.biographynet.nl>

those with an average positive score (above 0 which indicates a 'borderline' paper).

To our delight prof. dr. Paul Arthur from the University of Western Sydney in Australia, an expert on digital humanities and biographical data, agreed to deliver a keynote speech. Prof. dr. Susan Legêne, who holds a chair in political history at the VU University Amsterdam and who is one of the BiographyNet managers, closed the day by leading the final discussions.

The interest for BD2015 was also apparent from the registrations from people outside the Netherlands or neighboring countries who did not present a paper. Both the *Dictionary of Swedish National Biography* and the *Historical Dictionary of Switzerland*, for example, sent representatives to visit the event. Due to this high interest and the obvious need for a regular event that allows researchers working in this domain to exchange their work and ideas, it was decided to make *Biographical Data in a digital world* the first in a series of conferences taking place every two years, which is why we now speak of a conference rather than a workshop.

The venue, at Rockstart on the Herengracht in Amsterdam,³ ensured an open environment in which discussion and networking was greatly facilitated. A live stream made sure that people around the globe, would be able to follow the conference as well.⁴

3. Proceedings

Most authors of the 21 papers decided to submit a full version of their papers for the proceedings. The final submissions were subjected to a second reviewing and editing round, carried out by the organizers, to ensure that the original reviewers' comments were addressed. This resulted in the 18 papers you find in these proceedings. We grouped these papers in four categories: (i) *Bringing Biographical Data Online* (4 papers), (ii) *Analyzing Biographical Data with Computational Methods* (5 papers), (iii) *Group Portraits and Networks* (5 papers) and (iv) *Visualization and Representation* (4 papers).

Even though there clearly is a wide variety of papers, many of them deal with more than one of the following questions: How do we extract and analyze biographical data (with NLP, or manually added metadata)? How do we link biographical datasets? How do we visualize biographical data online? How does the digital turn change traditional biography? How reliable is the biographical data (content and manipulation of the content) for what kind of research questions? How do we identify and distinguish different individuals within these huge datasets of biographical data (named entity disambiguation)? What 'bigger questions can we answer with all this data?

We hope these proceedings will inspire people who work with biographical data and that it will give fruit to further research.

Amsterdam, 1 July 2015

The Organizing Committee:

Serge ter Braake, VU University Amsterdam
Antske Fokkens, VU university Amsterdam
Ronald Sluijter, Huygens Institute for History of the Netherlands
Thierry Declerck, Deutsches Forschungszentrum für Künstliche Intelligenz
Eveline Wandl-Vogt, Austrian Academy of Sciences

Alphabetic List of Program Committee Members

Paul Arthur, Humanities and Communication Arts, University of Western Sydney
Victor de Boer, Computer Science, Web and Media, VU University Amsterdam
Marc Boone, History, Ghent University
Georgeta Bordea Insight, Centre for Data Analytics, NUI Galway
Matje van de Camp, Taalmonsters.nl
Philip Carter, Oxford Dictionary of National Biography
Gearóid O Cleircín, Dublin City University
Marten Düring, Digital Humanities Lab, Centre Virtuel de la Connaissance sur l'Europe
Marieke van Erp, Computational Linguistics, VU University Amsterdam
Christine Gruber, Österreichische Biographisches Lexikon, Österreichisches Institut für Neuzeitforschung
Pim Huijnen, History, Utrecht University
Eero Hyvnen, Semantic Computing Research Group, Aalto University
Hans-Ulrich Krieger, DFKI GmbH, Language Technology Lab
Jonas Kuhn, Institut für Maschinelle Sprachverarbeitung (IMS), Universität Stuttgart
Stella Markantonatou Institute for Language and Speech Processing, Greece
Malte Rehbein, Digital Humanities, University of Passau
Matthias Reinert Historische Kommission bei der Bayerischen Akademie der Wissenschaften
Anneke Ribberink, Co-founder European Journal of Life Writing
Toma Tasovac, Belgrade Center for Digital Humanities
Jo Tollebeek, History, University of Leuven
Jane Winters, Digital History, Institute of Historical Research, University of London
Kelly Zervanou, Information and Computing Sciences, University Utrecht
Joris van Zundert, Digital and Computational Humanities, Huygens ING, The Hague

³<http://www.rockstart.com/>

⁴<https://www.youtube.com/watch?v=vAypbz0UEw0&app=desktop>