Abstracts

arranged by

Eveline Wandl-Vogt & Katalin Lejtovicz

02.11.2017

Organization Committee

Eveline Wandl-Vogt (Chair of Organization) Austrian Academy of Sciences, AT

Barbara Piringer, Katalin Lejtovicz & Matthias Schlögl, Austrian Academy of Sciences, AT
Paul Arthur, Edith Cowan University, AU
Serge ter Braake, University of Amsterdam, NL
Antske Fokkens, VU Amsterdam, NL
Ronald Sluijter, Huygens ING, NL

Kindly supported by

Paul Arthur
pwlarthur@gmail.com
Edith Cowan University Australia

Integrating Biographical Data: An Australian Case Study

The Humanities Networked Infrastructure Project (HuNI) is a national digital service for the study of Australian history and culture that incorporates rich biographical data. Developed with funding from the NeCTAR (National eResearch Collaborative Tools and Resources) program of the Australian Government, HuNI aggregates data from 30 of Australia's most significant humanities and creative arts datasets and makes them available for use by researchers across the arts and humanities and more widely by the general public. The project objectives are: to make Australia's wealth of cultural resources more accessible and connected; to break down barriers between humanities disciplines and support collaboration and data sharing between researchers, nationally and internationally; to create efficient workflows for researchers working with cultural data centred around enhanced discovery, analysis and sharing; and for the HuNI data aggregate service to lay the foundation for collaborative cross-disciplinary online research capability into the future.

This paper discusses the methods used by HuNI to aggregate data, the conceptual framework which has shaped the design of HuNI's Data Model around six core entity types including person entities, and the social or 'vernacular linking' features that make HuNI a unique resource for advanced scholarship. Vernacular linking allows registered users to make assertions about the relationship of data in the aggregate and have these assertions saved and uploaded to HuNI as an integral part of the data fabric for others to cite and build upon. The ability for researchers to work independently or collaboratively with the data through discovery, analysis and sharing functions is yielding new scholarly outcomes and deepening the world's understanding of Australian culture across space and time.

Yoshiyuki Asahi
yasahi@gmail.com
NINJAL

Detecting and mining biographical data from audio/audio-visual magnetic tapes: A case of the Japanese American collections in the US

This paper reports an attempt to detect and mine the relevant biographical data from audio/audio-visual magnetic tapes. A number of studies especially in history, anthropology, and sociology have conducted in the United States towards various immigrant groups especially since 1960 with an invent of portable audio recording devices. Numerous hours of audio/audio-visual recordings were created although most of them were 'buried' in the carbonate boxes. Finding aids have been created to understand them better. However, the information in the findings aids or inventories has not been enough to have any further access to the recordings especially from inter-disciplinary perspectives.
Therefore, this paper focuses two major collections, one public and the other private, targeting Japanese Americans in Hawai`i (approximately about 200 hours' recordings) and mainland US (approximately about 350 hours' recordings) to discuss how I have been working with finding and collecting the biographical data in the following four steps. (1) Digitizing the analogue sounds from the magnetic tapes, (2) collecting meta-data information from the finding aids/inventories, (3) creating the meta-dataset though annotating the recordings, and (4) mining the biographical data from the recording by listening to them. Through these steps, this paper will discuss the guidelines proposed mainly by oral historians and the applicability to each interview data.

Hans Bauer
hanschristianbauer@gmail.com
Leibniz-Institute for East and Southeast European Studies

Bio-brary The Library as Producer of Biographical Information? – The IOS-Example

The poster refers to two resources, managed by the library of the Leibniz Institute for East and Southeast European Studies Regensburg (IOS), and discusses strategies to open them up to further enhancement:
1. Online-Biographical Dictionary on the History of South-Eastern Europe (BioLex)
2. Amburger database

Ágoston Zéno Bernád
Agoston.Bernad@oeaw.ac.at
Austria Austrian Academy of Sciences

Maximilian Kaiser
Maximilian.Kaiser@oeaw.ac.at
Austria Austrian Academy of Sciences

The biographical formula: Kinds and dimensions of biographical networks

What are the recurrent parts of biographical dictionaries? In our paper we would like to discuss this question by presenting two modes for reading and analyzing biographical articles on a larger scale.

Behind every printed national biography there is a board of editors who is responsible for compiling the next volume and finding qualified authors for the biographies. Therefore they need to stay in contact with established scholars. This contact network, its development and influence, its biographical constructs and narratives will be exemplified in the history of the Austrian National Biographical Dictionary (ÖBL) using as yet uncharted and unedited sources from the archives of the Austrian Academy of Sciences.

A different mode for the interpretation of biographical dictionaries is the analysis of the biographies themselves. Moretti's witty formulation "Human beings employed full time in keeping institutions alive, not vice versa", which resulted from a study of published obituaries in the newspaper The New York Times, can be applied as well to the biographies of the ÖBL. As part of the research project "Mapping historical networks: Building the new Austrian Prosopographical | Biographical Information System (APIS)" biographical data is generated through the annotation of biographies of the ÖBL This data consists of frequently mentioned names of persons, places and institutions which can be subsumed under the term "biographical building blocks". On the basis of this data biographical networks can be built. In the second part of the presentation we would like to show you different dimensions of these networks as well as ways for analyzing this kind of data.

Judith Brouwer
judith.brouwer@huygens.knaw.nl
Huygens ING

Harm Nijboer
harm.nijboer@huygens.knaw.nl
Huygens ING

Golden Agents
A web of linked biographical data for the Dutch Golden Age

The production of art, books, literature and other creative products is covered by many electronic resources like collection databases of museums and libraries, dedicated documentation systems and research databases. Often biographical records are at the heart of these systems, but typically all these resources use their own subset of biographical data. In the Golden Agents program we will connect these resources to each other in a linked data framework. This will result in a sustainable infrastructure to study relations and interactions between producers and consumers of creative goods in the Dutch Golden Age.
In this paper we will discuss our experiences in connecting various and heterogeneous sets of biographical data. We will also provide some examples that highlight the research potential of the interlinked data.
The Golden Agents program is sponsored by The Netherlands Organisation for Scientific Research NWO and is a collaboration between Huygens ING, the Meertens Institute, the University of Amsterdam, Utrecht University, VU University Amsterdam, the Rijksmuseum, KB National Library of the Netherlands, the Amsterdam City Archives, the RKD Netherlands Institute for Art History and Lab1100.

Thierry Declerck
declerck@dfki.de
DFKI GmbH

Considerations about uniqueness and un-alterability for encoding biographical data in ontologies

In past work (Krieger & Declerck, 2015), we have described our considerations on synchronicity and diachronicity and how those aspects can and should be applied for defining properties in a formal ontology about biographical data. The resulting ontology is available at http://www.dfki.de/lt/onto/trendminer/BIO/biography.owl
As a matter of fact, one can soon come to the conclusion that it is very difficult, if not impossible to come up with a descriptor that designs perdurance, immutability, and absolute identity of an element of a biographical ontology. Are there properties of

human beings that are really immutable, so that they can be used as the fixed pillar on the base of which we can describe all other changeable aspects and characteristics of this human being? At the actual state of our current research, this seems not to be the case. Let us take as an example the case of the soldier Manning, introduced in (Wikipedia_Chelsea_Manning, 2017) as “Chelsea Elizabeth Manning (born Bradley Edward Manning, December 17, 1987)”. Is this person the same before and after the change of sex? And also looking at dates of birth or death: this is information that can still be modified, corrected etc. in dependency of new data or appreciations. Also we can have divergent information on this property, depending on the sources. And sometimes one is not able to state which one is the more reliable one.
Our intuition is that a very carefully designed ontology can offer a support here, as we are talking then about one “life” that might have no absolute fixed characteristics, or properties, but which on the basis of the large set of possibly divergent values of descriptors (classes and properties), and their organisation in one ontological space, can be considered as one unique carrier of a life. The need of a carefully designed vocabulary for biographical data is also stressed in (Brown and Simpson, 2013), showing that complying to generic vocabularies in existing ontology frameworks can lead to missing important points in a biography and even to inconsistencies and wrongly derived “facts”. As a first step we need to agree on what constitutes a “biographical unit”. My view on this is that the pseudonym “Michael Field” used by both Katharine Harris Bradley and Edith Emma Cooper for publishing their work is not a biographical unit – and in a sense also not an author. We consider thus a biographical unit as a biological element that has a birth and a death, while concentrating on human beings. In this view, “Michael Field” is just a name (pseudonym) shared by two biological units, and ontological descriptions have to be very cautious that properties for biographical units, like birth and death, are not shared with elements like “name”, as this can lead to inconsistent and wrong inferences.

References

Susan Brown and John Simpson. (2913). The Curious Identity of Michael Field and Its Implications for Humanities Research with the Semantic Web. In 2013 IEEE International Conference on Big Data, 77–85, 2013. doi:10.1109/BigData.2013.6691674.
Hans-Ulrich Krieger and Thierry Declerck. (2017). An OWL Ontology for Biographical Knowledge. Representing Time-Dependent Factual Knowledge. Proceedings of the First Conference on Biographical Data in a Digital World 2015, Amsterdam, Netherlands.
Wikipedia_Chelsea_Manning. (2017). https://en.wikipedia.org/wiki/Chelsea_Manning [last accessed 2017/10/15]

Anita Eichinger
anita.eichinger@wien.gv.at
Vienna City Library

Bernhard Krabina
krabina@kdz.or.at
KDZ - Centre for Public Administration Research

Biographical Information in the Vienna History Wiki

The Vienna History Wiki is a historical knowledge platform of the city of Vienna aiming to combine knowledge from the city administration with those of external experts [1]. It holds more than 37.000 articles about persons, topographical objects (streets, parks, cemeteries, districts, waters, etc.), buildings (houses, churches, bridges, monuments, etc.), organizations (companies, institutions, associations, etc.), events, maps and other entries (such as special Viennese German language terms).
With more than 15.000 entries about Persons, it makes up the largest content part of the wiki and is the most sought after category [3]. The bibliographic information does not only include typical encyclopedic textual information, but also facts about the persons ranging from basic personal data (name, alternative names, title, gender, date and place of birth and death, details on the graveyard, profession, political party, religious belief, estate) to professional functions, honorings, addresses and information about personal relations and relatives.
The open source software Semantic MediaWiki that powers the Vienna History Wiki allows for the data entered in online forms to be used inside the wiki for lists, overview pages or query forms as well as re-use of the data outside of the wiki through several export formats, such as RDF, JSON or CSV. [4] For quality assurance and integration of more data from different sources outside of the wiki, it is also possible to import data into the wiki. Therefore it was possible for the Vienna City Library to discontinue their annually published memorial day index (“Gedenktageindex”) on paper and CD-ROM and replace it with a query form in the wiki that lists all historically relevant persons whose birth or death date have e.g. 100th , 150th, 200th… anniversary in the following year.

[1] Vienna History Wiki. English information. Retrieved July 28, 2017 from Vienna History Wiki http://www.wien.gv.at/wiki/index.php/Vienna_History_Wiki
[2] Vienna History Wiki. Statistik. Retrieved July 28, 2017 from Vienna History Wiki https://www.wien.gv.at/wiki/index.php?title=Statistik
[3] Vienna History Wiki. Beliebteste Seiten. Retrieved July 28, 2017 from Vienna History Wiki https://www.wien.gv.at/wiki/index.php?title=Spezial:Beliebteste_Seiten
[4] Krabina, Bernhard: The Vienna History Wiki - a Collaborative Knowledge Platform for the City of Vienna. Proceedings of the 11th International Symposium on Open Collaboration (OpenSym 2015). ACM, 2015. Retrieved July 28, 2017 from OpenSym: http://www.opensym.org/os2015/proceedings-files/p500-krabina.pdf

Antske Fokkens
antske.fokkens@vu.nl
Vrije Universiteit Amsterdam

Serge Ter Braake
s.ter.braake@vu.nl
University of Amsterdam

A repository for Biographical Data Models

The last decades have brought a significant increase in digitized biographical and prosopographical data such as biographical dictionaries and repositories of prosopographical data. When creating such resources, designers have to think about the structure of the data.
Though standards such as the TEI P5 has been used in some projects, new models are developed independently in many projects.
Naturally, differences in information available or in intended applications may require different structures, but in general, there is much to be gained by basing new biographical and prosopographical datamodels on those used in existing projects. First, by avoiding reinventing the wheel, the process of coming up with a good design is sped up and mistakes may be avoided. Second, connecting information from various databases or building resources that make use of more than one repository is easier when similar data structures are used. However, it is not necessarily easy to find existing datamodels since there is no central overview of relevant projects. Furthermore, several resources consist of copy-righted data, making it impossible to share openly.
In this presentation, we introduce a repository for biographical data models and propose an approach for working towards a shared datamodel. This first version is based on a collection of datamodels used in 13 different projects. The datamodels are illustrated using made-up biographies (accompanied by prosopographical information) from fictional figures written by us.
This approach has two advantages. First, we circumvent the issue of copy-righted data. Second, because we represent the same information in all different datamodels, the repository provides a useful basis for a generic model that connects data from various models. We propose a two-step approach for connecting models, where we first convert all models to RDF staying as closely to the original structure as possible and then work top-down identifying links between RDF relations.

Abdulhamit Kirmizi
abdulhamitkirmizi@sehir.edu.tr
Istanbul Sehir University

We recently completed a three-year-project supported by TÜBITAK (The Scientific and Technological Research Council of Turkey). Our project aimed to identify the non-Muslim officials of the late Ottoman Empire (1803-1914). All the Armenian, Greek, Jewish, Assyrian, Maronite, Chaldean, Vlah, Latin, French, German, Polish officials employed in the civil administration were part of the research spectrum of the project. The main basis of the statistical survey was the personnel registers (Sicill-i Ahval Defterleri) kept in 1879-1914, consisting of 201 large volumes that provided information on 51,698 Ottoman bureaucrats. At the end of our research, we identified 2,946 non-Muslims who were employed in the Ottoman civil bureaucracy in various ministries. This research constitutes the largest prosopographical study completed not only for the Ottoman but for any imperial bureaucracy in 19th-century Europe.
The information obtained from the personnel registers was supplemented and loaded into a database and analysed with the latest SPSS data processing software. We identified 2,826 non-Muslim (1,476 Armenian, 1,036 Greek, and 314 Jewish) officials, plus 120 from smaller denominations. The oldest non-Muslim official was born in 1803 and the youngest one in 1893. The earliest entry of a non-Muslim official into government service according to our list was in 1835 and the latest in 1914. The study stretches over a time-span of 110 years (1803-1914) and includes men from various generations.
The personnel records provide valuable information including each official's name, pseudonym, father's name, father's occupation, birth date and place, educational credentials (such as the schools he attended and the languages he studied), places and ranks of assignments, and the reasoning behind dismissals. In addition, the personnel records provide information on bureaucratic positions held in respective ministries, entry points to offices, and the successive points where an official served during his career. This rich information sheds light not only on the officials' career trajectories, but also on some important aspects of their social and economic backgrounds.

Peter Kraker
pkraker@openknowledgemaps.org
TU Graz, Know-Center GmbH Research Center for Data-Driven Business & Big Data Analytics

Yukiko Sakabe
yukiko.sakabe@oeaw.ac.at
Austrian Academy of Sciences

Linked Cat+:
Exploring half a century of knowledge production at the Austrian Academy of Sciences

The Austrian Academy of Sciences (OeAW) was established in 1847 as a learned society in the tradition of institutions suchas the Royal Society in England and the Academié de Sciences in France. As such the OeAW has been an important forum for the production and communication of scholarly knowledge for over 150 years.
The meetings of the two major divisions, mathematics & natural sciences and humanities & the social sciences, took place in Vienna, with scholars presenting their findings either at the academy or by correspondence, including numerous events that shaped the history of science. It was for example here that Christian Doppler defended his theory against Josef Petzval in a fierce debate.
The main objective of LinkedCat+ is to bring the first half century of this unique forum to life, by making the communication within this forum accessible, discoverable andreusable. Our three pillars are open, linked and visual. In the project, we will be focusing on the written outcome of the meetings: the proceedings (Sitzungsberichte), dating back to 1848. Currently, the proceedings are only available as printed copies in the library of the OeAW. The digitisation of the complete proceedings is therefore a priority for BAS:IS (library, archive, collections -information & service of the OeAW).
For this corpus, LinkedCat+ will not only provide bibliographic catalogue data as linked open data, but also allow for feature-rich visual exploration of this data. The visual cross-linking of catalogue data in combination with open access to the full text does not only represent a new perspective in literature research, it also contributes to data-driven research in the history of science.
“Linked Cat+” complements two current projects “PAAS” and “APIS”. In PAAS, the biographies of the members of the OeAW are gathered in a structured way. In APIS biographies of notable persons with a relation to Austria between 1815-1950 are collected and systematically enriched. By linking these two resources to the metadata of the proceedings, we will create a complex and coherent knowledge network. This network will enable new possibilities for analysis and discovery of the research environment and the knowledge production at Austrian Academy of Sciences.

Almut Leh
almut.leh@fernuni-hagen.de
FernUniversität in Hagen, Institut für Geschichte und Biographie

Doris Tausendfreund
doris.tausendfreund@cedis.fu-berlin.de
Freie Universität Berlin, Center für Digitale Systeme (CeDiS)

Curation and Dissemination of lifestory interviews for the humanities (system demo)

Based on the special nature of audio-and video-interviews with eyewitnesses of National Socialism and survivors of the Holocaust we would like to demonstrate the current efforts to make them accessible online by introducing the Online-Archive “Forced Labor 1939-1945. Memory and History”.

This interview e collection contains 583 comprehensive life story interviews (192 video and 391 audio interviews) with concentration camp survivors, prisoners of war, and “civilian” forced laborers and were conducted in 27 countries in the native languages of the witnesses.
Each interview is accompanied by additional material: a short biography, transcript and translation of the interview, a table of contents showing the structure of the interview, photos and documents, as well as basic biographical information.
Content-based indexing, full-text search and an interactive map application enable a targeted search that leads directly to individual passages of the interviews. An annotation function allows to benefit from the specific knowledge of users to add to the understanding of the interviews. The archive has been designed multilingual and runs in German, English, and Russian in order to accommodate the needs of a greater international audience.
We will discuss considerations involved in designing an online platform to avoid the use of the interviews as a mere quotations quarry and instead supports a comprehensive understanding of the whole testimony in its narrative structure and its biographical meaning.
The presentation shows a powerful tool which enables academics to work effectively with testimonies to answer their own research questions.
Finally, we describe desirable perspectives for future developments, like the consolidation of interview collections in a Meta-Online-Archive which would offer search options for the databases of different archives as well as the option to upload interviews or to create additional content.

Petri Leskinen
petri.leskinen@aalto.fi
Aalto University

Eero Hyvönen
eero.hyvonen@aalto.fi
Aalto University

Jouni Tuominen
jouni.tuominen@aalto.fi
Aalto University

Analyzing and Visualizing Prosopographical Linked Data Based on Short Biographies

In our earlier paper [1] we presented an application case study where data from a printed collection of some 10,0000 short biographies of high school alumni was extracted and transformed into Linked Open Data, enriched by data linking to 10 external data sources, and published in a SPARQL endpoint. On top of the data service, a semantic faceted search engine and browser was developed for searching and filtering persons/biographies. This paper extends this work by showing how faceted search can be utilized as a basis for prosopographical data analysis and research: a new application is presented where various data visualization tools using Google Charts have been integrated with the SPARQL endpoint allowing the end user to filter out subsets of persons/biographies, and then to study them. In addition to providing statistical analyses of person groups, an interesting use case identified here is to compare visualizations based on different subgroups, e.g., famous people with entries in related datasets and those not included there. In this way it is possible, for example, to determine what education, profession, or employer will most likely lead to having an entry in the National Biography of Finland or Wikipedia. The data service is available at the Linked Data Finland platform [2], including some 892,000 triples about 131,000 resources. The extended application [3] is now in use on the Semantic Web.

[1] Eero Hyvönen, Petri Leskinen, Erkki Heino, Jouni Tuominen and Laura Sirola: Reassembling and Enriching the Life Stories in Printed Biographical Registers: Norssi High School Alumni on the Semantic Web. Language, Technology and Knowledge. First International Conference, LDK 2017, Galway, Ireland, June 19-20, 2017, Proceedings. Springer-Verlag,
[2] http://ldf.fi/norssit/sparql
[3] http://www.norssit.fi/semweb

Jelle van Lottum
jelle.van.lottum@huygens.knaw.nl
Huygens ING

Lodewijk Petram
lodewijk.petram@huygens.knaw.nl
Huygens ING

Reconstructing life courses of naval workers for insight into migrants' opportunities in and contribution to the Dutch economy, c.1700-1800

We apply innovative semi-automatic methods for record linkage on biographical data gathered from various large, heterogeneous historical data sets to gain historical insight into a topic that is currently hotly debated in both the political and public arena: what is the economic contribution of migrant workers on a recipient economy?
We present preliminary results from ongoing research into the eighteenth-century Dutch maritime sector. This was a key sector of the economy, characterised by a high level of migrant participation and well-known today for leaving rich sources. Our approach to uncover the link between migration and economic advancement centres on two aspects of what is generally referred to as job mobility: promotion and job switching. By identifying those individuals that during their careers experienced promotion and/or job switching we seek to lay bare the extent to which human capital levels of migrants and native workers differed, and what opportunities for social advancement were available in the eighteenth-century Dutch maritime labour market.
The data on migration and job mobility are obtained through record linkage of biographical data contained in muster rolls of the Dutch East India Company, interrogations of crewmembers of Dutch merchant marine, and birth, marriage and death registers from various Dutch towns. This procedure yields a large collection of mini-biographies of maritime workers, containing information on e.g. place of birth, places of residence, and job positions held at various points in time. From these mini-biographies we extract, among other things, geographical mobility and career events.
We present a research perspective on biographical data in a digital world, which is also meaningful to the point of view of content providers: it gives reason to think about the scope of online biographical data collections and the way in which access to these collections could best be provided.

Rennie Mapp
rcm7e@virginia.edu
University of Virginia

Collective Biographies of Women: A Prosopographical Database with Added Narratological Analysis

Collective Biographies of Women (CBW) is a Digital Humanities project based at the University of Virginia, with Professor Alison Booth as Principal Investigator. This presentation demonstrates current research affordances, briefly details its scholarly and collaborative history, and offers an original research use case.

CBW's database content is drawn from its bibliography of 1271 volumes of collective biographies, each of which includes at least three chapter-length biographies of women. It comprises bibliographic information enabling users to locate publication history and author information; it also collects the life data gleaned from 14,153 individual chapter biographies from within these volumes, and thus offers a wealth of searchable material about the 8709 women whose lives are their subjects. CBW offers many confirmed birth and death dates and alternate names; it also allows search and visualization by person typologies, as identified by the generic groupings that organize many of these books on the basis of profession, nationality or ethnicity, religion, or class. CBW is completing an NEH grant cycle in collaboration with SNAC (Social Networks in Archival Contexts), in which person records between the two databases are being algorithmically matched and then interlinked, thus enriching both prosopographical databases.

For over 300 of CBW's chapter biographies, the project site also displays and visualizes narratological markup that identifies lifestages, events, persona descriptions, discourse modes, and topoi. Captured using an XML schema encoded by hand, these data reveal patterns and narrative modes associated with specific, often historically contingent person typologies.

Using the corpus of chapter biographies of Lola Montez (1821-1861) as a case study, I demonstrate that—visualized together—biographical, bibliographical, and typological metadata alongside narratological markup can produce insights that neither would generate alone.

Websites:

http://cbw.iath.virginia.edu/
http://socialarchive.iath.virginia.edu/snac/search

Manuela Mayer
Manuela.Mayer@oeaw.ac.at
Austria Austrian Academy of Sciences

Biographical data as by-product of editorial projects

When speaking of biographies and biographical data in the field of historical research, one almost immediately (and sometimes exclusively) comes to think of biographical dictionaries or prosopographical studies. These projects have in common that collecting and publishing biographical data is an expected outcome of their work. In contrast to this, many other projects collect biographical data as well but for mere internal use only. In their publications, biographical information ends up as a commentary in a footnote or is cumulated in an index.
Especially when working with personal documents, e.g. correspondence, the dealing with biographical data is inevitable.
The already completed project “Joseph Eckhel and his numismatic network” as well as its currently running follow-up project “The numismatic networks of Eckhels’s Austrian precursors” aim at the publication of the correspondence of the numismatists Joseph Eckhel (1737-1798), Erasmus Fröhlich (1700-1758) and Joseph Khell (1714-1772). Their prime objective is an edition of the letters which will be published in print and online. In order to better understand the discussed topics, people and events mentioned are identified as precisely as possible. Here, we face a challenge that we share with other projects: most of the people mentioned in the letters can be identified by the use of archival material only as they are not mentioned in any of the biographical dictionaries or other publications available.
Therefore, our internal collection of biographies is constantly growing. In collaboration with the ACDH and APIS, a section within the APIS database was established for our project in order to manage our biographical data. This will not only end up in the commentary of our edition or some index. The fact that most of the people mentioned in the letters have never been registered before can be seen as a chance not only to enrich existing repositories, but also to link them to existing or coming authority files. Finally, our collection can also serve as a prosopographical tool e.g. on numismatics in Austria.

Eva Mayr
eva.mayr@donau-uni.ac.at
Danube University Krems

Saminu M. Salisu
saminu.salisu@donau-uni.ac.at
Danube University Krems

Matthias Schlögl
Matthias.Schloegl@oeaw.ac.at
Austrian Academy of Sciences

Maximilian Kaiser
Maximilian.Kaiser@oeaw.ac.at
Austrian Academy of Sciences

Ágoston Zénó Bernád
Agoston.Bernad@oeaw.ac.at
Austrian Academy of Sciences

Florian Windhager
florian.windhager@donau-uni.ac.at
Danube University Krems

Beyond One-dimensional Portraits: A Synoptic Approach to the Visual Analysis of Biography Data

Digital biographical databases are a rich resource for historical research: They provide a huge amount of information, which used to be scattered in different texts or local archives, and makes it possible to reconnect them into bigger pictures on the life patterns of historic individuals and groups. Yet, analyzing, as well as scholarly reasoning and sensemaking with these multidimensional data remains challenging, especially for public audiences or non-experts in digital methods. In this paper we present how a visualization framework developed in the PolyCube project (www.donau-uni.ac.at/en/polycube) addresses these challenges by a synoptic approach to multidimensional data visualization.

Data or information visualizations “use computer-supported, interactive, visual representations of abstract data to amplify cognition” (Card et al., 1999). Some biographical databases already offer such supportive measures in form of basic visualizations like maps, networks, or timelines, which allow analyzing single data-dimensions, i.e. geographical, relational, or temporal aspects of individual lifelines. The PolyCube project develops synoptic information visualizations which bring together multiple data dimensions in one picture.

These visualizations are based on the concept of a “space-time-cube” and add temporal information to visualizations like maps, networks, or various other diagrams (Windhager et al., 2016, fig.1). In this way, movement patterns in multiple analytical spacetimes can be explored, and additional temporal perspectives on the data (like juxtaposition, superimposition, or animation) can be generated with seamless transitions out of the initial space-time arrangement (Bach et al., 2016). Further interaction techniques enable users to explore the data, to zoom into time and space, or to select and examine relevant entries for closer investigation.

We present and discuss a first visualization prototype, and illustrate its potential for biographical research by the means of a case study built on biographical data from the APIS project (https://apis.acdh.oeaw.ac.at/).

References:

[1] S. K. Card, J. D. Mackinlay, and B. Shneiderman, Readings in information visualization: using vision to think . Morgan Kaufmann, 1999.

[2] F. Windhager, E. Mayr, G. Schreder, M. Smuc, P. Federico, and S. Miksch, “Reframing Cultural Heritage Collections in a Visualization Framework of Space-Time Cubes,” in Proceedings of the 3rd HistoInformatics Workshop. http://ceur-ws. org , Krakow, 2016, vol. 1632, pp. 20–24.

[3] B. Bach, P. Dragicevic, D. Archambault, C. Hurter, and S. Carpendale, “A Descriptive Framework for Temporal Data Visualizations Based on Generalized Space-Time Cubes,” in Computer Graphics Forum , 2016.

Barbara Piringer
barbara.piringer@oeaw.ac.at
Austrian Academy of Sciences

Eveline Wandl-Vogt
eveline.wandl-vogt@oeaw.ac.at
Austrian Academy of Sciences

Yalemisew Abgaz
yalemisew.abgaz@adaptcentre.ie
Adapt Centre / Dublin City University

Katalin Lejtovicz
katalin.lejtovicz@oeaw.ac.at
Austrian Academy of Sciences

Exploring and exploiting biographical and prosopographical information as common access layer for heterogeneous data facilitating inclusive, gender-symmetric research

In this paper, the authors offer a cross disciplinary, cross sectoral view on biographical and prosopographical data as access layer and demonstrate it based on the project exploreAT! (2015-2019).
Biographical and prosopographical information is an essential information item in several (humanities) projects. This is discussed on the example of a long term (100+ years) dictionary project (WBÖ, DBÖ) and it´s collection, which is recently opened up for access within the project exploreAT!
The collection (about 5mio paper slips) is compiled by about 11.157 persons, who played different roles and had various professional backgrounds during the century lasting curation and compilation process: authors of books, senior researchers and research managers, student researchers, collectors, informants etc.
The project can be interpreted as an early humanities crowdsourced project, brought into being by the emperor of the late Habsburg monarchy. Said that, within exploreAT! we are mainly dealing with everyday people and/or people of a local / regional importance.
Knowledge about the person/s related to the data collection, their expertise and strengths as well as weakness, is essential to judge the data and derive information as well as create new knowledge based on this.
Due to this, since the beginning of the project in the early 20th century, these data have been documented in the institutes´ archive and since the late 90es, a so called “Mitarbeiterdatenbank” was giving access to documentation of people in a structured way. Since 2010 these data are modelled, published and stored in a MySQL-database, cross-referenced with geo-referenced biographical and bibliographical information as well as interlinked with their collections.
In this presentation the authors discuss how person data could be used as a common

access layer regarding heterogeneous (research) data. They focus on the importance of tracking project related, quality information regarding person data.
They introduce into an enriched data model based on APIS project, mapped with the experiences, learnings and needs from exploreAT! aiming to make visible the unknown mass, contributing to more inclusive acknowledgements and enabling a more gender-symmetric approach within (humanities) research.

Jose Luis Preza
Preza jl@preza.org
Austrian Academy of Sciences

Automatic classification of digital objects in institutional repositories using artificial intelligence

Data Science is simply a discipline that combines data with programming languages, algorithms, statistics, machine learning, artificial intelligence, reporting, and data visualization, all to make sense out of data. Data Science is a very important part of Cognitive Computing that enables Artificial Intelligence.
Public, School and University Libraries are in a very advantageous position: they sit on a lot of data.
The data stored in such libraries are very diverse. There are books, documents, charts, datasets, experiments, software, tables, numbers, images, videos, audio, dissertations, magazines, newspapers, processes, usage, user data, financial data, to mention a few.
The challenge for libraries is not only to digitize all their content (digital objects), but also to classify, organize, link and publish all digital objects. Additionally, Libraries require to measure different indicators of their Repositories, including Storage, Use, Web traffic, Uploads, and Users.
Up until now, most content (digital objects) has been organized and classified manually. However, manual Processes are not sustainable, certainly not when you have to process millions of digital objects in a short period of time and with high accuracy.
Here is where Data Science comes to the rescue.
The techniques and methods used in Data Science allow Libraries to easy the workload and get results faster than with manual processes.

Concrete areas where Data Science can assist Libraries include:

- Digital Object Classification/Semantics/Search: Automatic classification of digital objects (keywords, entities, concepts. See a related document I wrote titled: “Automated Information enrichment for a better search”, Zenodo DOI: https://zenodo.org/record/163933)
- Picture Recognition and Classification: automatic classification and tagging of pictures (also extracted from Video).
- Content Clustering and Segmentation: Automatic clustering and segmentation of digital objects based on content.
- Reporting: Make reports out of your contents
- Predictive Analytics: who is going to read/use what.
- Machine Translation: Automatic translation of digital objects, including Braille.
- Speech to Text: extraction of Audio Speech to convert it into text.
- Text to Speech: convert text into audible speech
- Plagiarism: with machine learning, advanced techniques can be developed to prevent plagiarism
- Visualisation of Data and Statistics regarding Library Repositories.

Matthias Reinert
matthias.reinert@hk.badw.de
Historical Commission

Bernhard Ebneth
bernhard.ebneth@ndb.badw.de
Historical Commission

data.deutsche-biographie.de - interfaces to access biographical data

Biographical information in the internet today has to handle expectations of different user groups. It has to be source-based, reliable and citeable as like as ordinary scientific data and research ever has. To become relevant in the digital field it requires to be used, reused and re-mixed providing adequate interfaces and licenses. Therefore the Deutsche Biographie adopted following principles to provide sustainable biographical resources online. First it is based on dictionaries of scientific biographies who are based on archival research, collected and edited in a peer review like process. These biographies serve as a foundation of knowledge which is enlarged by

bibliographic metadata drawn from selected freely available authority files (GND). In order to provide further relevant and more actual material to its entries the Deutsche Biographie is linking to enduring resources with institutional background.
The Deutsche Biographie offers faceted search capabilities, visual representations of selected interpersonal relationships and individual maps for geographical scope of the person's life.
In the near future the search and presentation modes will be made even more transparent by open search and visualisation interfaces. Users of different levels of expertise shall be equipped with suitable tools to search and visualize as well as analyse and export their results. Citeability will be extended to the visualisations, individual research queries and results and reuseability will be admitted by providing open licenses and new remixing interfaces. DH experts will access the RDF-store and query in Sparql, or use the Neo4J-Graph-Database and query in Cypher.
In order to meet different user requirements a research laboratory will be installed providing access to information, providing individual search, analyses, customizable visualisations for intermediate users, for experts even more possibilities. In addition to individual access the Deutsche Biographie will start to cooperate with other institution aiming to establish an interoperable ontology of biographical information ultimately leading to easier integration of scattered biographical resources worldwide.

Latest Publications

Bernhard Ebneth, Matthias Reinert: Potentiale der Deutschen Biographie (www.deutsche-biographie.de) als historisch-biographisches Informationssystem. in: Christine Gruber, Maximilian Kaiser, Agoston Benad (Hgg.): Tagungsband "Europa baut auf Biographien, Wien Oktober 2015" (in print)
Matthias Reinert, Maximilian Schrott, Bernhard Ebneth, Team Deutsche Biographie, Malte Rehbein: From Biographies to Data Curation – the Making of www.deutsche-biographie.de, in: BD2015 - Biographical Data in a Digital World 2015. Proceedings [...], April 9, 2015, Ed. by S. ter Braake et al., CEUR Workshop Proceedings Vol-1399, p. 13-19 <http://ceur-ws.org/Vol-1399/paper3.pdf>
Sophia Stotz, Valentina Stuß, Matthias Reinert, Maximilian Schrott: Interpersonal Relations in Biographical Dictionaries. A Case Study, in: BD2015 - Biographical Data in a Digital World 2015. Proceedings [...], April 9, 2015, Ed. by S. ter Braake et al., CEUR Workshop Proceedings Vol-1399, p. 74-80 <http://ceur-ws.org/Vol-1399/paper12.pdf>

Matthias Schlögl
Matthias.Schloegl@oeaw.ac.at
Austrian Centre for Digital Humanities

Katalin Lejtovicz
katalin.lejtovicz@oeaw.ac.at
Österreichische Akademie der Wissenschaften

A Prosopographical Information System (APIS)

During recent years a massive amount of biographical datasets have been digitised and partly made available open access. Additionally collaborative efforts such as Wikipedia also contributed to the increased amount of publicly available partly structured prosopographical and biographical data sets.
Since these first endeavours, researchers have worked on extracting structured data from these biographical texts by applying various Natural Language Processing (NLP) techniques such as local grammars, regular expressions, machine learning and deep learning based approaches. However, these approaches have been brought forward by a relatively small scholarly community typically working with non-distributed, tailor made systems, that usually do not have a GUI. These systems are therefore difficult to use for researchers not trained in IT.
The goal of the APIS tool is to create a collaborative research environment for humanists, allowing them to explore digital humanities data sets and provide insight to the information embedded in full text documents by automatically identifying entities and revealing connections between them. The system is developed within the APIS project, but is used in other ACDH projects too and is published on GitHub (MIT).
APIS is a web-based, highly customizable virtual research environment (VRE) that allows researchers to carry out the analysis of their data in a computer-assisted environment, by making use of Semantic Web and NLP techniques. This hybrid approach (i.e. the possibility to annotate documents and edit entities/relations both manually and automatically) allows researchers to "use the best of both worlds" and computer scientists to improve the tools directly on real world data. In our paper we present the tools and methods used for the processing and semantic enrichment of the data set, particularly setting a focus on the data model chosen, the structure of the application and the approach used for entity detection, linking and relation extraction.

Donald Sturgeon
djs@dsturgeon.net
Harvard University

Linking, sharing, merging: sustainable digital infrastructure for complex biographical data

In modeling complex humanities data, projects working within a particular domain often have overlapping but distinct priorities and goals. One common result of this is that separate systems contain overlapping data: some of the objects modeled are common to more than one system, though how they are represented may be very different in each.

While within a particular domain it can be desirable for projects to standardize their data structures and formats in order to allow for more efficient linking and exchange of data between projects, for complex datasets this can be an ambitious task in itself. An alternative approach is to identify a core set of data which it would be most beneficial to be able to query in aggregate across systems, and provide mechanisms for sharing and maintaining this data as a means through which to link between projects.

For biographical data, the clearest example of this is information about the same individual appearing in multiple systems. Focusing on this particular case, this talk presents one approach to creating and sustaining with minimal maintenance a means for establishing machine-actionable links between datasets maintained and developed by different groups, while also promoting more ambitious data sharing.

This model consists of three components: 1) schema maintainers, who define and publish a format for sharing data; 2) data providers, who make data available according to a published schema; and 3) client systems, which aggregate the data from one or more data providers adhering to a common schema. This can be used to implement a sustainable union catalog of the data, in which the catalog provides a means to directly locate information in any of the connected systems, but is not itself responsible for maintenance of data. The model is designed to be general-purpose and to extend naturally to similar use cases.

Serge Ter Braake
sergeterbraake@gmail.com
University of Amsterdam

Antske Fokkens

antske.fokkens@vu.nl
Vrije Universiteit Amsterdam

Famous for only one thing: Concepts, Events and People intertwined

When editors of Biographical Dictionaries decide they want to add a person's biography they do this based on the claim to fame of this person. Some people are mostly famous for being a politician (Winston Churchill) , others for being a scientist (Charles Darwin) and others for only one event or particular feat (Neil Armstrong). This way famous people (Darwin), famous events (publication of the Origin of Species) and certain concepts (Evolution Theory) are intertwined in texts through history.

In this paper we will look at famous people in history in various (Dutch) biographical sources and try to chart their claim to fame by looking at associations of these people with related events and concepts. We start by taking key words or phrases from a sample of texts that mention persons from different categories: science, politics and art. We trace the association scores -in this case the likeliness of people, events and concepts occurring in the same text over time - and map the patterns that emerge. This will allow us to get new insights into why people are remembered and how the canonisation of people, concepts and events are intertwined. This in turn, will shed light on biographer's choices on who to write about, why and in what way.

Jouni Tuominen
jouni.tuominen@aalto.fi
University of Helsinki

Eero Hyvönen
eero.hyvonen@aalto.fi
Aalto University

Petri Leskinen
petri.leskinen@aalto.fi
Aalto University

Bio CRM: A Data Model for Representing Biographical Data for Prosopographical Research

Biographies make a promising application case of Linked Data: they can be used, e.g., as a basis for Digital Humanities research in prosopography and as a key data and linking resource in semantic Cultural Heritage portals. In both use cases, a semantic data model for harmonizing and interlinking heterogeneous data from different sources is needed. We present such a data model, Bio CRM [1], with the following key ideas: 1) The model is a domain specific extension of CIDOC CRM, making it applicable to not only biographical data but to other Cultural Heritage data, too. 2) The model makes a distinction between enduring unary roles of actors, their enduring binary relationships, and perduring events, where the participants can take different roles modeled as a role concept hierarchy. 3) The model can be used as a basis for semantic data validation and enrichment by reasoning. 4) The enriched data conforming to Bio CRM is targeted to be used by SPARQL queries in flexible ways using a hierarchy of roles in which participants can be involved in events.

Bio CRM provides the general data model for biographical datasets. The individual datasets concerning different cultures, time periods, or collected by different researchers may introduce extensions for defining additional event and role types. The Linked Data approach enables connecting the biographies to contextualizing information, such as the space and time of biographical events, related persons, historical events, publications, and paintings. Use cases for data represented using Bio CRM include prosopographical information retrieval, network analysis, knowledge discovery, and dynamic analysis.

The development of Bio CRM was started in the EU COST project "Reassembling the Republic of Letters" [2] and it is being piloted in the case of enriching and publishing the printed register of over 10 000 alumni of the Finnish Norssi high school as Linked Data [3].

[1] http://seco.cs.aalto.fi/projects/biographies/
[2] http://www.republicofletters.net

[3] Eero Hyvönen, Petri Leskinen, Erkki Heino, Jouni Tuominen and Laura Sirola: Reassembling and Enriching the Life Stories in Printed Biographical Registers: Norssi High School Alumni on the Semantic Web. Proceedings, Language, Technology and Knowledge 2017. June 19-20, Galway, Ireland, Springer-Verlag, 2017.

Eveline Wandl-Vogt
eveline.wandl-vogt@oeaw.ac.at
Austrian Academy of Sciences

Liquid infrastructures for complex research questions : The OI-RI Biographical Research Demonstrator in a nutshell

„Was kommt nach dem Brockhaus-denken?“
This question is a quote of Ulrich Weinberg introducing Network thinking [1].
The presentation is based on these theories as well the theory of Peter Spiegel regarding We-culture [2]; furthermore related to relevant political strategies (EC, ministries AT) e.g. open innovation [3][4].
The Open innovation research infrastructure (OI-RI) is introduced. this is a collaborative, international movement, which started in 2016.12 in Vienna.
Against the background of ongoing projects [5] at the research group on methods and innovation, exploration^{space}, the author discusses the implementation of the OI-RI, based on a potential biographical research demonstrator.
The author describes the general architecture of the OI-RI and a certain biographical research demonstrator. The design and architecture of the OI-RI is based on an intense exchange of virtual and physical infrastructures, champions who may take action in liquid experimental labs and close interaction of all societal actors in an environmental setting(quintuple helix model.
In this presentation the author focuses on the new forms of knowledge regarding biographical data [cf. 6,7] and the multidisciplinary collaboration topic: To meet complex research questions beyond disciplinary scope, e.g. value or risks of migration, relationship of technology development, society and culture, impact of emerging technologies and societal evolutions on biographical research, etc., multidisciplinary collaboration is essential [cf. 8]. in this paper the author introduces into open innovation methods and practices as well as other methods applied within the OI-RI design, to implement a user centered approach e.g. Design thinking.
She maps it to other ongoing initatives and focuses on the participatory, collaborative workflows on the example of liquid innovation hubs.
Concluding, the presentation demonstrates on the example of biographical research into strategies, workflows and first results of the new liquid lab at the Austrian Academy of Sciences, founded and coordinated by the author.

[1] Ulrich Weinberg (2015). Network Thinking. Was kommt nach dem Brockhaus-Denken?
[2] Peter Spiegel (2015). We-Q better than IQ.
[3] Chesbrough, Henry (2003). Open Innovation. The New Imperative for Creating and Profiting from Technology.
[4] Österreichische Strategie für Open Innovation (2015). https://www.bmvit.gv.at/innovation/downloads/open_innovation_strategie_oesterreich.pdf (zugriff: 18.5.2017).
[5] APIS. http://www.oeaw.ac.at/acdh/en/apis; https://apis.acdh.oeaw.ac.at/ ; exploreAT! exploring austria´s culture through the language glass: https://exploreat.usal.es/ (zugriff jeweils: 18.5.2017).
[6] Schnapp, Jeffrey (2014): Knowledge Design. Herrenhausen Lectures pamphlet series, Volkswagen Foundation, Hannover, Germany.
[7] Schnapp, Jeffrey: The Intimate Lives of Cultural Objects. forthcoming in Routledge Companion to Media Studies and Digital Humanities, ed. Jentery Sayers, (New York: Routledge, 2017).
[8] Goikhman, Alisa, Theron, Roberto, Wandl-Vogt, Eveline (2016): Designing collaborations: Could design probes contribute to better communication between collaborators? In Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM ’16). ACM, New York, NY, USA, XXX-XXX. DOI: http://dx.doi.org/10.1145/3012430.3012431

Lars Wieneke
lars.wieneke@uni.luLuxembourg
Centre for Contemporary and Digital History

Marten Düring
marten.during@uni.lu
Luxembourg Centre for Contemporary and Digital History

Denis Scuto
denis.scuto@uni.lu
Luxembourg Centre for Contemporary and Digital History

Joël Thill
Joel.thill@an.etat.lu
Luxembourg Archives Nationales Luxembourg

No privacy for the dead.
GDPR legislation and Biographical Data in Contemporary History

Structured biographical data and the emergence of shared repositories for their storage and provision represent significant opportunities for scholars, teachers and cultural heritage institutions. They make it possible to link persons and organisations through time and space on an unprecedented scale; always together with references to the underlying source materials. Such links make it faster and easier to reconstruct individual careers, attitudes, beliefs and social relationships; but may also reveal hitherto unobserved patterns in the data.
Especially for micro-historical research the systematic linking and exploration of structured biographical data becomes a natural extension of traditional methods. Medievalists and classicists can happily apply these approaches for their research since “the dead have no privacy”. Contemporary historians on the other hand come under pressure from new European legislation under the General Data Protection Regulation (GDPR). GDPR is already a reality and has been adopted by the European parliament in April 2016 while coming into force in May 2018. While its main aim – to regulate the use and protection of privacy related data for EU citizens – is more than welcomed by the authors, the concrete implementation on a national level might pose significant threats to already established practices. In this presentation, we want to highlight some of the potential issues that might emerge from GDPR for research in contemporary history, e.g. the reinforced “right to forget” that affects not only published but also internal repositories, the idea of “consent” given by the data subject when dealing with historical files and the idea of anonymization or pseudo-anonymization and their implications for identity disambiguation.

Christophe Verbruggen
christophe.verbruggen@ugent.be
Ghent Centre for Digital Humanities, Ghent University

Jan Vandersmissen
Jan.Vandersmissen@UGent.be
Ghent Centre for Digital Humanities, Ghent University

Sally Chambers
sally.chambers@ugent.be
Ghent Centre for Digital Humanities, Ghent University

Towards a Person-Entities Digital Ecosystem for Europe

For research across a wide range of humanities disciplines, biographies are frequently used as reliable and quality-assured sources of person-related information. Examples include international, regional, national, subject-specific, and period-specific biographies.
Increasingly, online community driven reference works (e.g. Wikipedia) are used for biographical information. There are hundreds of biographical databases and (online) repositories containing biographical data in various digital formats and standards. Conducting research across borders is therefore subject to multiple challenges. In particular, research that targets comparative research into transnational lives and cross-border mobility is often frustrated by dispersed sources. A serious matter of concern is that biographical sources are often trapped in national or subject-specific silos or behind publisher paywalls.
Researchers from a wide range of disciplines urgently need an open, pan-European quality-assured data infrastructure to a) uniquely identify historical persons and b) to trace their life courses across national and disciplinary boundaries. PEDIGrEE intends to be that infrastructure.

The Person-Entities DIGital Ecosystem Europe (PEDIGrEE) is conceived as a pan-European digital ecosystem to cluster, link and exchange data about historical persons from existing biographical data silos. PEDIGrEE is intended as a scalable distributed data infrastructure to allow arts, humanities and social science researchers with an interest in historical persons to link and share their data observations to specific individuals from Europe's past and thus to make cross-boundary and comparative research using digital methods feasible.

PEDIGrEE is being initiated by the Ghent Centre for Digital Humanities at Ghent University and together with an international network of partners from Austria, Belgium, Croatia, France, Germany, Greece, Ireland, Luxembourg, the Netherlands, Serbia, Slovenia, Switzerland and the United Kingdom.

In this paper, we will present the current status of PEDIGrEE, including latest achievements and next steps.