

**Biographical Data in a Digital World, Workshop Amsterdam 9 April 2015 -
Abstracts**



Team Deutsche Biographie, Bernhard Ebneht and Matthias Reinert – *From biographies to data curation – the making of www.deutsche-biographie.de*

The website www.deutsche-biographie.de is a joint effort of Historical Commission at the Bavarian Academy of Sciences and the Bavarian State Library to publish online two series of biographical dictionaries, the „Allgemeine Deutsche Biographie“ (ed. 1875-1912) and the „Neue Deutsche Biographie“ (ed. since 1953, 25 vols. A to Stader). Both dictionaries are regarded as the „German National Biography“.

The “making of” the website www.deutsche-biographie.de rests on three pillars: the ongoing edition of biographies in the series of the German National Biography in alphabetical order (Neue Deutsche Biographie/ NDB), the digitisation of two biographical dictionaries, and the preparation for online presentation and linking.

These activities underwent a strong shift towards proper presentation and semantic linking in the past decade.

We will focus on this very effort and describe our objectives: text-encoding, identify individuals and places in authority files (Gemeinsame Normdatei, GND) and aggregate further individual/biographical information online from persistent, scientific and source-based as well as royalty free websites and databases.

Our strength is certified biographical information, namely identified individuals, links to primary and secondary sources, lists of works/objects and portraits and their electronic copies. The website integrates personal information from archives, museums, collections, and libraries on the basis of interlinked biographical characterizations and offers an entry point for historical biographical research.

The potential of our website lies in coordinated biographical data management and integration. We seek funding for a web based workbench for biographical metadata providing access to semantic and graph-based interfaces. We would like to amass and enrich the common database in a collaborative and modular manner together with partners in Germany and Europe. In addition we will compile a consistent corpus of biographical articles for linguistic research which we will enlarge step by step.

The work was funded by grants from the Deutsche Forschungsgemeinschaft (DFG).

Vanessa Hanneschläger and Katharina Prager – Ernst Jandl and Karl Kraus: Two Lives in Bits and Pieces

In the research strand 'Virtual Biography' the 'Ludwig Boltzmann Institute for the History and Theory of Biography' explores the methods and possibilities of arranging biographical data online in 'bits and pieces'. In two pilot projects on the Austrian writers Ernst Jandl (1925-2000) and Karl Kraus (1874-1936), heterogeneous concepts of structuring and linking big biographical data sets are being tested. While the Ernst Jandl platform researches the Bio-Bibliography, the digital approach to Karl Kraus is an Anti-Biography. This paper reflects the different approaches chosen for the two subjects and researches the role of canonization processes for online biography.

Bärbel Kröger and Christian Popp – *Germania Sacra Online – The Research Portal of Clerics and Religious Institutions before 1810*

The research project Germania Sacra provides a comprehensive prosopography database, that makes structured and comparable data of the church of the Holy Roman Empire available for further research. The database already contains approximately 30,000 records. In addition to the own data of the Germania Sacra cross-institutional collaboration and integration of scientific data resources of other research projects are essential.

Shivani Poddar, Venumadhav Kattagoni and Navjyoti Singh – *Personality Mining from Biographical Data with the “AdjectivalMarker” Technique*

The last decade has witnessed significant work in personality mining from lexical cues in social media data. Not much work has yet been undertaken in extracting the same from the biographical data reserves populating social media. Most of this work involves a large crowd of researchers leveraging dictionary-based approaches such as LIWC (which primarily focus on function words). By the means of this paper we intend to introduce a novel method of personality mining from social media data called “Adjectival-marker Technique”. This method involves extracting lexical features from descriptive texts (e.g. biographical data) to train a learning model so as to predict the respective personality traits of the subject. Conceptually, it draws heavily from the last 78 years of work in lexical psychology and the Big Five personality test. However, it is not only a computational variant of the primordial theories of lexical psychology, but is also competent in conferring a substantial accuracy of personality prediction, matching that obtained by psychometric tests. In this study, we hypothesize a variant of the Lexical Hypothesis of psychology. This modified hypothesis is validated by the computational results of personality prediction achieved by the Adjectival Marker Technique discussed below. The paper also discusses some insights illustrating the coherence of people's judgments about the subject's personality (virtual personality) with one another. The average accuracy (i.e. matching that achieved by psychometric tests for Big 5) for prediction approximated to Extraversion - 82.82% Agreeableness - 89.62%, Conscientiousness - 92.48% and Imaginativeness/Intellect - 81.67%.

Firas Dib, Simon Lindberg and Pierre Nugues – *Extraction of Career Profiles from Wikipedia*

In this paper, we describe a system that gathers the work experience of a person from her/his Wikipedia page. We do this by connecting persons to professions through the analysis of parts of speech and dependency relations we extract from text. Setting aside the dates, we computed recall and precision scores on a very limited and preliminary test set for which we could reach a recall of 74\% and a precision of 95\%, showing our approach is promising.

Marie-Charlotte Le Bailly – Biographies of People on the Move and Cultural Heritage in a Digital Era. Participative and biographical collecting at the Red Star Line Museum Antwerp

The Red Star Line Museum opened in Antwerp in 2013. It explains the process of the historical emigration to America through Antwerp to a wide and varied audience within the broader context of migration. Migration stories or migrants' biographies, both past and present, play an important role in the scenography and educational department, as well as in the presentation of the digital collections. So far the museum has collected about 1500 personal stories of migration with related cultural heritage, today and in the past, through research of staff-members and participation of the public. The museum actively invites visitors to participate and share their own migration stories or that of their relatives. The resulting biographical collection comprises a wide range of oral histories, narrative sources, objects and documents.

This poster discusses the process of making our biographical collections accessible and some of the challenges we meet in finding the right balance between our own needs and resources in managing these collections and meeting our public responsibilities and duties as a museum. The focus lies on the following question: Are standards for the exchange of biographical data (A2A, BioDes and GEDCOM) useful in achieving the objective of 'reversed' canonization of people and events in history and presentation of biographies to a varied audience?

Serge ter Braake and Antske Fokkens – *How to Make it in History. Working Towards a Methodology of Canon Research with Digital Methods*

Historians have been intrigued for a long time by the question why some people become famous in history and others do not. Often this question is linked to national identity and memory, which are huge research topics on their own. It is difficult to perform a manual analysis of who becomes famous and why, since you would need to trace the fame of a huge number of people (including the ‘ forgotten’ people) to see what patterns can be discerned. This paper therefore proposes a methodology for the studying of canonization of people in history with the aid of digital methods. The possibilities and limitations of current data and tools are discussed, as well as the eventual steps needed to do such kind of research successfully.

Hans-Ulrich Krieger and Thierry Declerck – *An OWL Ontology for Biographical Knowledge. Representing Time-Dependent Factual Knowledge*

Representing temporally-changing information becomes increasingly important for reasoning and querying services defined on top of RDF and OWL, for practical applications such as modern biographical information systems in particular, and for the Semantic Web/Web 2.0/Social Web in general. Extending binary OWL ABox relation instances or RDF triples with temporal information often translates into a massive proliferation of useless container objects when trying to keep the underlying RDF model.

In this paper, we argue for directly extending RDF triples with further temporal arguments in order to easily represent time-dependent factual knowledge and to allow for practical forms of (temporal) reasoning. We report on a freely-available OWL ontology for representing biographical knowledge that models entities of interest via a tri-partite structure of the pairwise disjoint classes Abstract, Object, and Happening. Even though the ontology was manually developed utilizing the Protege ontology editor, and thus sticking to the triple model of RDF, the meta-modelling facilities allowed us to cross-classify all properties as being either synchronic or diachronic. When viewing the temporal arguments as extra (or annotation) arguments that only apply to ABox relation instances, universal biographical knowledge from the TBox and the RBox of an ontology can still be described as if there is no time. Nevertheless, when it comes to built-in (and custom-based) reasoning, the standard entailment rules for RDFS and OWL need to be extended by a temporal dimension.

Roser Morante and Antske Fokkens – *The expression of uncertainty in Dutch biographies*

In this paper, we present an exploratory study of the expression of uncertainty in biographies. We focus on a corpus of 149 biographies from the Biography Portal of the Netherlands (BP), which contains biographies from the two main sources included in the portal. We aimed at determining to which extent biographical facts are presented as uncertain and what kind of linguistic markers the authors use to express uncertain facts. We found that only 25% of the biographies contain indications of uncertainty in the first place and that very few facts within a biography are presented as uncertain. The linguistic phenomena used to express uncertainty involve mostly modality markers, which we describe in the paper.

Gearóid Ó Cleircín and Brian Ó Raghallaigh – *Ainm.ie: Breathing new life into a canonical collection of Irish-language biographies*

In this paper we present the Ainm.ie online collection of Irish-language biographies. This collection is a product of a project to retro-digitize the 'Beathaisnéis' ('Biography') series, published between 1986 and 2007, as well as ongoing biographical work to expand and enrich the collection. The Beathaisnéis series comprised biographical accounts of 1,650 lives and an additional 520 amendments and supplementary articles, written in Irish. Persons were chosen for inclusion in this series according to their relevance to the Irish-language world. This canonical collection is an invaluable research tool for Irish-language scholars, historians and others but the print volumes risked becoming obsolete and inaccessible. As well as producing a digital version of the original texts, a key aim of the Ainm.ie project has been to ensure the continuity of biographical research in Irish by providing an online platform for publication. This paper introduces the Beathaisnéis collection and explains its context. It goes on to describe the digitization process and the editorial work carried out to enrich the digital version. Finally, it discusses how the project has facilitated contemporary biographical research in Irish.

Irene Russo, Tommaso Caselli and Monica Monachini – *From facts to discoveries: visualising Holocaust deportees' lives*

This work focuses on the development of an NLP module for event and temporal analysis of biographies as available in Wikipedia. At the current level of development, we restricted the extraction to temporally anchored events as they represent salient information which can be further used to extract additional events and facilitate their chronological ordering and the representation of a person's timeline. We also focused on the extraction of spatial information which explicitly expresses where the anchored events took place so as to facilitate the distinction and aggregation of different event mentions.

A visualization tool has been designed to help historical reasoning, by providing in a quick, visual shape basic facts about groups of people but also enabling further discoveries about them by comparing their lives along different dimensions.

Ming-Kin Chu – *Prosopographical survey of lecturers at the Directorate School in early Northern Song China (960-1050)*

This paper discusses how biographical data of the lecturers at the Directorate School in early Northern Song (960-1050) China are mined, extracted, converted, verified, visualized and analyzed. Through this prosopographical survey of lecturers, I aim to address the following questions: to what extent did the scholarly backgrounds of lecturers shape students' learning at the Directorate of Education? To what degree did the geographical background of a lecturer cast an impact on his intellectual inclination? What explains the changes of lecturers' appointment, and how did such changes shed light on the relationship between government education and examination? Answers to the above questions may help us to rethink an important historical question in Chinese history: to what extent were reforms at the metropolitan educational institutions in the Qingli (1041-1048) era revolutionary?

Lonneke Geerlings – *Intersectionality in Social Network Analysis. Visualizing biographical data: a case study of Rosey E. Pool’s correspondence archive*

My research focuses on Rosey E. Pool (1905-1971), a travelling translator that was involved in the Black Poets movement and the civil rights movement. As a part of my research, I’m constructing a relational database out of the extensive correspondence of Pool. This corpus will be analysed with help of computational linguistics, mapping and visualisation techniques (Lima 2011). This will provide a distant reading of these texts and of her network(s) (Boot 2008). The collected data will be entered in a relational database in Access. Various metadata are included in this database (date, address, theme, abstract, mutual friends, acquaintances, and organizations). Then the data will be analysed using database operations and using Gephi, a tool for network analysis and visualization. The results will be visualized in an ego-centred network. Furthermore, persons are registered with multiple features, such as date of birth, gender, sexuality, religion, ethnicity, origins, family bonds, social sector, home city (and others).

A distant reading of the corpus of texts is needed to test my hypothesis that Pool was indeed excluded in the period 1966-1971, an era of race politics. A close reading of key texts is used to interpret letters in a historical context. Hopefully, the dichotomy of (1) empiricism and (2) literary analysis will be overcome by a close reading of key letters in the correspondence archive. The theory of intersectionality will be used to show deeper connections between the different actors (or ‘nodes’) (Lutz et al. 2011:2). This paper includes a case study, focusing on the First World Festival of Negro Arts (Dakar, 1966), Robert Hayden, Rosey E. Pool, and the role of the common faith of these two persons: Bahá’i.

Bram Vannieuwenhuyze – *Who's who in late medieval Brussels*

Without any doubt pre-modern towns were characterised by very dynamic and heterogenic population structures. Yet, historians face many difficulties to plumb the depths of this complex phenomenon, especially because very few records encompass the various layers of the urban population. As a result, historians mainly pay attention to the general demographic evolution of the entire town or study the socio-economic profiles of people belonging to very specific urban groups and communities, of whom the majority has to be situated among the civic elites. There is no doubt that such a 'narrow' viewpoint helps to by-pass the complexity of the subject, but it also leads to a fragmented and static image of urban society as a whole. The question remains if it is possible to obtain a more nuanced view. Therefore it is necessary to collect and unlock larger amounts of data in digital databases, which allow to compose multi-faceted biographies of townspeople and finally pave the way for urban prosopography. That is exactly the aim of the running "Who's who in late medieval Brussels" project, which unlocks biographical data on some 40.000 people who worked, lived or stayed in Brussels before 1600. Of course, such an undertaking requires serious efforts. In this paper I will explain the configuration of the "Who's who in late medieval Brussels" database and take it as a case study for balancing some of the heuristic issues and present some research opportunities.

Sophia Stotz, Valentina Stuss, Matthias Reinert and Maximilian Schrott.- *Interpersonal relations in biographical dictionaries. A case study*

Adopting the concept of “Local Grammars” (M. Gross) which were successfully applied in practice by M. Geierhos to biographical information extraction for English news [1] our project aims to detect, encode, and finally visualize relations between persons.

Our corpus consists of the 2 digitised lexicographical series (Allgemeine Deutsche Biographie/ADB, 55 vols. 1875-1901 and Neue Deutsche Biographie/NDB, 25 vols 1953-2013 ongoing).

We developed local grammars to describe interpersonal relations and applied them to the corpus with Unitex 3.1.

First we had created dictionaries for names, surnames, forenames, occupations, institutions. We adopted two strategies, bootstrapping names and professions using the corpus given as well as we were recurring on additional indexes and authority files (namely the Gemeinsame Normdatei/GND) to add names, surnames, forenames, occupations to dictionaries.

Secondly we were constructing local grammars to describe the syntactical structure of sub-sentence phrases referring to interpersonal relations. The local grammar incorporated existing TEI-XML structures in the texts.

Using the ability of local grammars in Unitex to act as transducers we were able to produce XML-tags and encode semantic information.

So far we described a general set of local grammars for personal names and for place names. Based on these we described the following relations:

“to study” and context (university, teacher, discipline, degree)

predecessor / successor

friends & circles

The third step was to disambiguate the “named entities” and identify persons (as given in the authority file or index).

Finally we displayed relations on our website in an interactive and dynamic way. Utilizing the javascript library D3 we represented named relations between identified individuals as ego centered network graphs.

The work is funded by the Deutsche Forschungsgemeinschaft (DFG) to establish a biographical information system online

Elian Carsenat and Tania Vichnevskaja – *Applying onomastics to scientometrics to evaluate Diaspora engagement in Scientific Networks*

NamSor as a private start-up company has been solicited in 2014 by a European country to help measure the ‘brain drain’ affecting its competitiveness in the BioTech sector and to produce a global map of its scientific Diaspora (who are they, where are they and what are they doing). The objective was to build up the country’s scientific international cooperation and to engage its Diaspora.

Serendipity led analysts to discover interesting patterns in the way scientists names affect co-authorship and citation – not just for this particular country, but globally. So collaboration started between NamSor and bibliometric experts at INSERM –the French National Institute for Health- to evaluate and visualize the effects of migration, Diaspora engagement and possibly cultural biases in Science.

The study involves the analysis of over ten million medical research articles from PubMed (PMC) database and about 50k articles in the field of Cancer Research from Thomson Reuters WebOfScience (WOS) database. The corpus contained 15k articles, 68k authors; cited: 32k articles 168k authors; 17 million author citing-author cited occurrences and covered: China (Fudan University Cancer Hospital), USA (Dana-Farber Cancer Institute), France (French Comprehensive Cancer Centers FCCC), Japan (National Cancer Center), Poland (All applicable institutes), Slovenia (All applicable institutes).

We discovered for example, that scientists in the Chinese corpus, affiliated to a Chinese University, with a Chinese name cite (in rank) scientists geographically affiliated to: the United-States, China, Japan. Analysing cited scientists names, we found that they cite primarily Chinese scientists in the United-States (Diaspora Chinese).

Further research is needed to understand the phenomenon: Cultural Biases? New patterns in migration of highly qualified professionals? Active brain drain? Translation and language matters? Efficient Diaspora engagement? The question remains open but the results have deep implications on topics such as: country rankings in Research, or how a Scientific Diaspora can be engaged for education, scientific or economic development of the home country.

Mark Bell and Sonia Ranade – *Traces Through Time: a case-study of applying statistical methods to refine algorithms for linking biographical data*

The Traces Through Time project, which ran at The National Archives from January 2014 to March 2015 developed algorithms and tools to link historical records referring to the same individual. These tools are of potential value to researchers in a range of disciplines within the digital humanities, but the project was originally conceived to apply ‘big data’ approaches in support of prosopographical research. Our work identifies connections between the people named in the records and reveals the networks that emerge from this analysis. However, the tools have also proved to be of interest to those pursuing biographical research*. Since each graph is built from a large number of underlying pairs of connected occurrences, the resulting networks may as readily be examined to draw out individual threads as to interpret the patterns that emerge when lives are viewed collectively.

The project applied a range of data-comparison techniques to assess the similarity of pairs of occurrences of individuals in the records. Our scoring framework, developed for Traces through Time, gives us flexibility to select and combine different measures according to a defined weighting scheme - which can be tailored for data of different types or from different historical periods. Statistical techniques are applied to calculate a level of confidence for each link. This in-built flexibility allows the tools to process diverse datasets, whilst the robust confidence measures enable the outputs to support different applications and research questions - a necessary consideration for The National Archives, where our collections span 900 years of history and the research interests of our user community are no less diverse.

This paper describes work carried out to further enhance the initial outputs of the Traces Through Time project and to apply a learning approach to extend the methodology for identifying links. The learning approach refines our confidence-scoring metrics as successive runs of the algorithm traverse a series of partially overlapping datasets. Though the intersection between any particular pair of datasets may be small - sometimes only a fraction of a percent - we observe an incremental effect as additional datasets are combined: so that the connections revealed by each set of comparisons accumulate to deliver greater insights into the dataset as a whole.

The paper describes methods for identifying and incorporating common differences in textual information arising from factors such as: handwriting recognition errors, typographical errors and phonetic errors made when names are recorded. A different approach is described for dates of birth, where the algorithm must accommodate inaccuracies in recording such as misrepresentation of age or rounding of declared ages**. The age distribution observed for each dataset is fed back into the algorithm to support a statistical approach to calculating the likelihood that two occurrences of a person with different recorded dates of birth, in fact, relate to the same individual. As each incremental enhancement of the algorithm improves the results of the matching process these, in turn, highlight further discrepancies in the data, from which the algorithm can learn.

The research shows that when matching fuzzy data, a learning approach can significantly improve on the results delivered by the ‘raw’ scoring mechanisms, enabling us to carefully target the type and degree of fuzzy matching to be applied. The approach has helped us strike an improved balance between the poor precision that results from allowing increased ‘fuzziness’

and the poor recall that results from a more restrictive approach. Future work will identify further enhancements, extending the learning approach beyond names and dates of birth, and will embed the enhancements into the Traces Through Time framework and tools.

We have also demonstrated the use of clustering on names to segment a dataset into national/ethnic groups. This not only improves the accuracy of the probabilistic matching method, by adjusting the population figure according to likely place of origin, but also demonstrates how working with data at scale can support inferences about a property of a person that are not explicitly stated in the data.

The automated processing described in this paper offers some benefit to the biographical researcher. It can cover large volumes of data quickly and efficiently, including the ‘fuzzy’ or partial data that is typical of historical sources. It offers excellent scope for tuning the matching and confidence parameters to the researchers’ requirements. Our conversations with researchers working in this field indicate that they require fine-grained control over the operation of the toolset available to them, though there is some debate as to whether this is compatible with a simple interface and a shallow learning-curve. Of course, automatically-generated networks of occurrences cannot be considered to be ‘biographies’ in any meaningful sense - particularly when the metadata on which these are based tends to be terse, may be incomplete and is prone to a range of errors. However, the tools have a role in generating leads, revealing chains of connections, and identifying directions for further investigation. We believe that data-centric approaches such as those explored by Traces Through Time will be important for opening up ‘big data’ for biographical research.

* The Traces through Time project ran a workshop in January 2015 to help us assess the applicability of the tools to a range of digital history research questions. Unexpectedly, the majority of the delegates were engaged in biographical research - although other fields were also represented.

** The ‘age heaping’ effect is observed in datasets which record age (rather than date of birth). The resulting distribution is skewed, typically showing peaks at ‘round’ ages (e.g. 10, 20, 30...). For an illustration, see the 1911 census graphic from the ONS data visualisation centre (<http://www.bbc.co.uk/news/uk-18854073>).

Matvey Kolbasov, Anastasia Bonch-Osmolovskaya – *Tolstoy Digital: mining biographical data in literary heritage editions*

Tolstoy Digital project (<http://tolstoy.ru/projects/tolstoy-digital/>) aims to prepare web-published semantically marked-up version of 90-volume Full Edition of Leo Tolstoy works. The digital version of the 90-volume edition has become easy of access thanks to mass crowdsourcing campaign “All of Tolstoy in one click” (see for example review in Guardian <http://www.theguardian.com/books/2013/oct/16/all-leo-tolstoy-one-click-project-digitisation>).

The next step of Tolstoy digitisation is devoted to semantic tagging of the Tolstoy text and creating a comprehensive database of all the additional reference information that goes along with Tolstoy oeuvres and private archive.

The 90-volume edition comprises exhaustive critical apparatus, which contains relevant information on Tolstoy’s works, life and other people connected with him. Current research, done as part of Tolstoy Digital project, presents on-going work and first results of fact extracting from literary commentaries. The edition contains 21 volume of letters, dated from 1844 till 1910, the year of Tolstoy’s death.

Each letter is followed by a detailed commentary where a biographical reference to the addressees and persons, mentioned in the letter, is provided. Our aim is to analyze unstructured text of a commentary, extract person names and relevant biographical information, and to add TEI semantic annotations. Extracted data will be aggregated and stored in a reference database. The database will be linked with the text of semantic edition.

Nico Randeraad, Christophe Verbruggen and Thomas Dhaeninck – *Visualizing longitudinal data. Rooted cosmopolitans in the Low Countries, 1850-1914*

TIC Collaborative is a Virtual Research Environment (VRE) for the study of 19th and early 20th century international congresses and organizations (www.tic.ugent.be). The VRE applies the principle of scholarly crowdsourcing (as already used by other research units across and beyond Europe), for instance for the disambiguation of persons and the collection of biographical data. It offers access to digitized dispersed sources in a central observatory. The platform includes a Nodegoat powered collaborative relational database (in order to process the data for social network analysis, prosopography, etc.) of participants in international social reform conferences and an Islandora powered source corpus with annotation tools and export functionalities. In this paper we will discuss our possibilities to create longitudinal data visualisations using both the built-in and external exploration and visualisation tools. By looking at multiple memberships of social reformers and congress participation from 1850 until 1914, we will present our first attempts to create a genealogy of causes, issues and collaborations. An accepted way of mapping the evolution of networks over time is through the multiple memberships of participant cohorts as an indicator for transfers of ideas and initiatives. Looking at both local (for now confined to the Low Countries) and transnational milieus (international congresses and organizations linked to transnational exchange), we address questions related to core-periphery relations and cohesive subgroups.

Daniel Shakespear – *Interactive Genealogy Explorer*

An interactive, three-dimensional graphic interface was conceived and developed, to enable visualization and interaction with the movements and important events recorded in a genealogy. The resulting visualization provides a colorful and intuitive view of migrations, as well as plotting the biographical data for individuals in the genealogy. Users can quickly highlight the connections between people, get a perspective on which persons lived in similar eras, or trace the life events of any individual in the genealogy.

The events of individuals' lives are plotted as colored lines above the backdrop of a globe (Google Earth), tracing the life and travels of the person. The date of each event determines the altitude of the line at the event location, and the lines are colored corresponding to the date of birth of each individual. The user can move the view around manually, or cycle through the events in the individual's life. One lifeline or up to hundreds can be shown (for example, chosen by date, name, ancestors, descendants, connection between two individuals, etc.), to give a sense of the connections between people, and provide a visual overview of migrations and movements. The presentation at "Biographical Data in a Digital World" will discuss the background, basics of geometry generation, initial learnings and some areas for additional exploration. Images and videos of the Interactive Genealogy Explorer can be viewed here: <https://www.behance.net/gallery/22589523/Interactive-Genealogy-Explorer> (higher-resolution versions of the videos are available on YouTube via the links).

Andre Blessing, Andrea Glaser and Jonas Kuhn – *Biographical data exploration as a test-bed for a multi-view, multi-method approach in the Digital Humanities*

The present paper has two purposes: the main point is to report on the transfer and extension of an NLP-based biographical data exploration system that was developed for Wikipedia data and is now applied to a broader collection of traditional textual biographies from different sources and an additional set of structured biographical resources, also adding membership in political parties as a new property for exploration. Along with this, we argue that this expansion step has many characteristic properties of a typical methodological challenge in the Digital Humanities: resources and tools of different origin and with different accuracy are combined for use in a multidisciplinary context. Hence, we view the project context as an interesting test-bed for some methodological considerations.