Olaf Berg

„Capturing Displaced Persons' Agency by Modelling Their Life Events: A Mixed Method Digital Humanities Approach"

https://doi.org/10.14765/zzf.dok-2051

Reprint von: Olaf Berg, „Capturing Displaced Persons' Agency by Modelling Their Life Events: A Mixed Method Digital Humanities Approach", in: Historical Social Research 45, Nr. 4, 2020, S. 263-289

Historical Social Research
Historische Sozialforschung

Olaf Berg:

Capturing Displaced Persons' Agency by Modelling Their Life Events: A Mixed Method Digital Humanities Approach.

doi: 10.12759/hsr.45.2020.4.263-289.

Published in:

Historical Social Research 45 (2020) 4

Cite as:

Berg, Olaf. 2020. Capturing Displaced Persons' Agency by Modelling Their Life Events: A Mixed Method Digital Humanities Approach. *Historical Social Research* 45 (4): 263-289. doi: 10.12759/hsr.45.2020.4.263-289.

https://doi.org/10.14765/zzf.dok-2051

Capturing Displaced Persons' Agency by Modelling Their Life Events: A Mixed Method Digital Humanities Approach

Olaf Berg *

Abstract: *»Die Handlungsmacht von ‚Displaced Persons' durch die Modellierung von Lebensereignissen erfassen: Ein Ansatz, der qualitative und quantitative Methoden in den digitalen Geisteswissenschaften zusammenführt«.* The International Tracing Service archives offer process-generated documents from resettlement programs for displaced persons (DP) after World War II. This paper addresses two key challenges to ongoing research based on those archival holdings: the generation of data; and the visual representation of that data in geographic information systems. Digital history offers the opportunity to go beyond case studies and use the wealth of process-generated documents as serial sources for algorithm-based analysis. However, data in that form does not exist as such, and thus needs to be generated – a process that implies interpretative acts such as abstraction, normalization, and trans-coding, which are shaped by the character of digital media. Can modeling a DP's life into a series of events, and digitally processing the resultant data, help to find out more about the agency of DPs negotiating their destiny with the authorities? If the mostly hidden and implicit configurations of digital knowledge production are thoughtfully considered and geostatistical analysis is combined with close readings of selected source documents, hermeneutic and quantitative approaches can be reconciled via digital history. This mixed method approach has implications for research culture and the publication of such data.

Keywords: Mixed methods, digital history, digital humanities, data, modeling life events, displaced persons, geographic information systems.

1. Introduction

At the end of the Nazi Regime and World War II, millions of persons had been displaced from their home towns and countries. Families were torn apart, many persons did not know if their relatives had survived and if so, where they might be. In 1943, the Allies founded the United Nations Relief and Rehabilitation Administration (UNRRA) with the plan to repatriate displaced persons (DPs)

* Olaf Berg, Leibniz Centre for Contemporary History Potsdam, Am Neuen Markt 1, 14467 Potsdam, Germany; olaf.berg@zzf-potsdam.de.

Historical Social Research 45 (2020) 4, 263-289 | published by GESIS
DOI: 10.12759/hsr.45.2020.4.263-289

https://doi.org/10.14765/zzf.dok-2051

and refugees as soon as possible after Victory Day.[1] In 1947, its tasks were taken over by a successor organization, the International Refugee Organization (IRO; see Cohen 2012; Gatrell 2013).[2] However, migration is not simply regulated. Rather, it is a negotiation (Rass and Wolff 2018), and in many cases the DPs had their own plans. For various reasons, many of them did not want to return to their homes. When they registered with the authorities in order to obtain DP status, their hopes and wishes were translated into an administrative process that left traces in documents that now provide important sources via which to research the postwar migration regime. These process-generated documents were produced at the intersection of the institutional environment and the individual decisions and actions of DPs and eligibility officers. They are expressions of the configuration of institutional settings and personal experiences that form part of the negotiation processes concerning the status and opportunities of the DPs. Thus, the constitution of the actors in the postwar migration regime and DPs' agency is inscribed in these documents.

How can historians unveil and understand this agency? Due to the large number of potential documents to be examined, and their diverse and complex structure, these process-generated documents are most often explored as small subsets of documents used to investigate individual cases or small groups of DPs. To create a broader picture of DPs' agency, the resulting insights on a limited group of DPs might be extrapolated to a more general level and combined with other generic sources, such as statistics, administrative acts, and reports. To some degree, this is the usual business of historians, who also have a wide range of scholarly procedures with which to generate knowledge from sources by combining and discussing documents as narrative strategies with the aim of forming credible and meaningful histories out of that knowledge.

However, digital history methodology now offers the possibility to go beyond case studies and use the whole stack of process-generated documents as serial sources for algorithmic analysis – for example, by modeling a DP's life into a series of events, and digitally processing the data. Can this approach help to find out more about DPs' agency in negotiating their destiny with the authorities? If so, how can the relevant information be generated from the sources and modeled into a useful form for research? This paper will discuss these questions in the context of ongoing research in the context of the People on the Move consortium (PotM), focusing on process-generated documents from

[1] UNRRA was created at a 44 nation conference at the White House on November 9, 1943. Its mission was to provide economic assistance to European nations after World War II, and to repatriate and assist the refugees who would come under Allied control (United States Holocaust Memorial Museum 2018).

[2] The IRO constitution was approved by the UN General Assembly on December 15, 1946. In July 1947, the Preparatory Commission of the IRO started operations. On September 15, 1948, the first IRO General Council took place. The new agency inherited responsibility for 643,000 DPs in 1948 (Marrus 1985; United States Holocaust Memorial Museum 2018).

https://doi.org/10.14765/zzf.dok-2051

resettlement programs for DPs after World War II.[3] In his article in this special issue, Henning Borggräfe (2020) points to the potential of the diverse sources in the archives of the International Tracing Service (ITS), and to possible strategies to read the material as expressions of DPs' agency. My article takes up this challenge and focuses on the specific demands inherent to such an approach from the perspective of a historian with IT skills, responsible for data modeling. I discuss the challenges of generating, modeling, and processing information from the documents into a series of events – challenges that are not normally visible in final research articles.

The first part of this paper provides an outline of the research consortium's specific approach and prospects. The second part addresses two key challenges for the consortium: the generation of data; and the visual representation of that data in geographic information systems (GIS). In the final part, I discuss some of the methodological implications of our approach for research culture and the publication of research data.

2. The Vision

The starting point of the People on the Move research consortium are ITS archival holdings of over 30 million documents, many of them generated in the course of applications for assistance in the aftermath of World War II. The holdings bear traces of DPs' agency because they are the product of the negotiations of their applications. Depending on the density of the documents that are inherited by the institutional process of transmission, many of these files contain extensive information on the negotiation process, arguments brought forward, and decisions taken, which can help us to understand the way in which, and how successfully, DPs applied for resettlement, repatriation, or other forms of assistance. Especially in the so-called CM/1 files, we find short autobiographies up to the day of application, reasons given to the authorities, and the results of examinations and decisions by the institutional administrators (for more details, see Borggräfe 2020 in this special issue).[4] The holdings of about 350,000 CM/1 files are immense, although far from complete, and thus also biased by the process of transmission within the institution.

However, the vast amount of material does not allow us to examine the whole stack of documents in depth. Within any given (reasonable) timespan,

[3] Members of the People on the Move consortium are researchers from the International Tracing Service (ITS) in Bad Arolsen, the NIOD in Amsterdam, and the University of Osnabrück (UOS). Between 2015 and 2018, I was a member of the PotM team of Prof. Christoph Rass at UOS.

[4] See the holdings 3.2.1 IRO "Care and Maintenance" Program, ITS online collection <https://digitalcollections.its-arolsen.org/030201?lang=en> (Accessed November 23, 2018).

https://doi.org/10.14765/zzf.dok-2051

researchers can only investigate selected individual biographies or the fate of a small group of selected individuals. Yet how should one select those individuals of interest, and what was the fate of the rest? One possible answer is a mixed method approach that combines hermeneutic research on individual cases and general debates with a more quantitative, prosopographical approach in order to gain knowledge from a large sample of individual cases. While biographical approaches focus on one specific person, prosopographical approaches try to find patterns in the lives of a group of persons with some common features, such as the same profession, place of origin, destination, age, gender, or social status.

Much faster than a thorough examination, it is possible to extract only the main data of time and place for biographical events from the documents. With this data, it is possible to model DPs' itineraries as declared to the authorities. If we are able to do this, and to include some additional data – for example, gender and profession – not only for a small group of individual DPs but also for a large-scale sample, the ITS archival holdings will allow us to detect patterns of movement and their relation to success and failure in the application process. If we add information on the field office and the administrative staff who took the decision, we might also detect patterns in the decision-making process specific to individual field offices and officers. This prosopographical approach allows us to generate knowledge about DPs and forced migration that is built on the accounts and traces DPs left during the process of negotiating their destiny. If we understand DPs and IRO eligibility officers, not as subjects independent of their environment but as actors constituted in the negotiation process, the patterns generated by prosopographical analysis can help us to understand the process of their constitution as actors.

Initial sample checks within the PotM context have shown that this mixed-methods approach and the material archived at the ITS have the potential to extend and challenge knowledge generated by classic approaches, specifically regarding the individual agency of the persons involved – for example, the widespread assumption that in order for an application to be successful, DPs needed to have, or fake, a profession that fitted resettlement programs' criteria (see Sebastian Huhn's [2020] article in this special issue). The statistical knowledge generated from serialized life-event data allows us not only to review the relevance of such assumptions but also to point to statistically irrelevant phenomena and groups of DPs – for example, a small group of Buddhist DPs that became visible in the course of encoding by PotM. Once identified as a group, they can be investigated further by examining the existing material in detail.

To allow such interaction between quantitative and hermeneutic levels of investigation, it is important that the information from the case files used for the generation of patterns and statistics is still connected in the data model to the documents from which it was extracted. Detail and context appear on a

https://doi.org/10.14765/zzf.dok-2051

continuum of data, and it is this continuum that allows us to go back and forth between general and more detailed views of the archival material. Individual cases can be seen in the context of general patterns, and general patterns can lead to new questions that can be further investigated by looking into individual cases connected to that pattern. Thus, hermeneutic and quantitative approaches stimulate and enrich each other in this mixed method approach.

A specific group of data abstractions upon which we focus are geospatial abstractions provided by GIS. GIS can be used to generate visualizations of data encoded with time and place coordinates. From a given dataset, for any given moment or period, GIS can model the distribution of individuals and objects in space, and the pathways of those persons and objects through time and space. Although this data shows movements (or persistence) and not agency itself, by processing large amounts of data from case files, patterns of movements appear that can reveal the conditions constituting the migration regime, specific negotiation "hot spots," and DPs' strategies. Thus, within this mixed method approach, quantitative analysis helps to identify cases and sources of interest relevant to a detailed hermeneutic investigation.

Summarizing, the vision of the People on the Move research consortium consists of: (1) generating prosopographical data of life events from the process-generated sources about DPs in the ITS archival holdings; (2) processing this data via GIS in order to generate a context and questions for each individual source that need to be investigated further via hermeneutic research; and (3) creating an infrastructure to present sources and research results in an integrated and extensible context of growing knowledge about DPs.

3. The Challenges of Producing Data and Modelling DPs' Trajectories

The above mentioned prospects for a GIS-based access to, and analysis of, sources about negotiations between DPs and migration regulating institutions such as UNRRA and the IRO, and their life itineraries, all build on a foundation common to any computer-based analysis. Information from the sources has to be identified, extracted, and separated into discrete units that can be represented and stored in a database and processed by algorithms. This poses various challenges for the transformation of the vision sketched out in the previous subsection into a feasible and fruitful project on at least two intermingled levels: On a technical level, we need an appropriate markup and generation process for the relevant information, appropriate solutions for the representation and storage of the generated data and information, and finally, appropriate processing instructions for the modeling of the data into meaningful representations of DPs' lives and agency. On an epistemological level, we need to be

https://doi.org/10.14765/zzf.dok-2051

aware of the limits and implications for knowledge production of such a data-driven research approach based on algorithms.

3.1 Data Production in Times of New Media

At first glance, extracting prosopographical data from life events might appear to be a straightforward process of finding and collecting data. However, as Geoffrey Bowker (2005) suggests, "raw data is both an oxymoron and a bad idea; to the contrary, data should be cooked with care" (184). For example, date of birth might appear to be a simple matter of fact. However, in the way we define and register it, we inscribe upon that notation our worldview of a linear time, starting with an event relevant to Christian culture, and the administrative purposes of identifying a person by his or her name and birthday. In other cultures, the event of birth is represented in different data. For example, the Akan people of Ghana frequently name their children after the day of the week they were born, and the order in which they were born (Agyekum 2006). In ancient Mayan cultures, the day of birth is identified by a system of various circular calendars, and determines the opportunities and character of a child. Like Gregorian calendars, Mayan calendars are expressions of deep-seated structural, cultural, and historical identities (Rice 2007). Thus, the way in which we represent a certain event influences the knowledge we can gain from it. Data is not something pre-existing, waiting to be found and collected, but rather something that is produced by the very process of collecting it. "Data [does] not just exist," Lev Manovich (2001, 224) explains, it has to be "generated." Text encoding is a kind of modeling (Ciotti and Tomasi 2016, 5) and at a certain level, the collection and management of data may be said to presuppose interpretation. "Data need to be imagined as data to exist and function as such, and the imagination of data entails an interpretive base" (Gitelman and Jackson 2013, 3).

Like data, historical facts are not a "given" – something inherited from the past that we can find and pick up somewhere "out there" in the world. They are products of the transformation of social interaction into symbolic representation and the transformation of this representation in the course of its tradition, as well as the research process that places one fact relative to another (Goertz 2001). Both historical facts and the history that is built upon them are coined by the medium of symbolic representation – namely, language (White 1973). This is why extracting data from the sources should be considered a productive process of selection and transformation of representation by its transposition from one medium (written text, spoken word, etc.) into a digital medium – a process that constitutes something new.

This perspective raises questions regarding the specificity of the digital medium, and the relation between medium and knowledge. Some theories situate the medium as the determining factor for knowledge, like in Marshall McLu-

https://doi.org/10.14765/zzf.dok-2051

han's (1964) defining phrase: "The medium is the message." From this perspective, technology determines the social. In a similar fashion, Friedrich Kittler (2003) describes power and knowledge formations as discursively produced by mediums of storage, transmission, and processing. Other theories are based on the idea that a medium is only the carrier of pure information. Following this view, meaning is constituted by a system of signs independent of the medium. Niklas Luhmann (1997) – an exponent of this idea – distinguishes between medium and form: while the medium only loosely links elements, the form coins the medium via a rigid coupling of these elements. However, both positions deny social interaction with the medium as constitutive of the production of meaning and knowledge. Actor-network theory (ANT) has shown that knowledge emerges from negotiation processes between observing human actors and observed non-human actors (Latour and Woolgar 1986; Latour 2005). Thus, media have agency, but their agency is not detached from their interaction with human actors. Thus, meaning created by media cannot be a pure object of knowledge; rather, it needs someone who perceives, and thereby constitutes it, although only becoming perceptible through the medium (Bächle 2016). The medium is thus not simply the message. Rather, the message carries the trace of interaction between the medium and human actors. Moreover, what is expressed through one medium might also be expressed through another medium, but surely not without any medium. Thus, meaning cannot be detached completely from a medium, and the idea of pure information is false (Krämer 1998). The medium is what paradoxically disappears in its own manifestation, and is thus present via its absence (Mersch 2008, 2009).

Specifically regarding digital media, we must first note the tension between the persistence of traditional conventions for knowledge production on the one hand, and the projections of new technologies on the other (Bächle 2016). In their critical introduction to new media, Martin Lister et al. (2009) emphasize the continuity between "old" and "new" media. Despite technological changes, for instance from photochemical to digital image production, the social practices of photography have not changed that radically. As the key characteristics of new media, they identify their "digital, interactive, hyper-textual, virtual, networked and simulated" characteristics (Lister et al. 2009, 13). With much more emphasis on technology-induced changes, Lev Manovich (2001) identifies five key differences that make new media "new": numerical representation; modularity; automation; variability; and cultural transcoding. Digital representation slices information into separated data units that can be processed by algorithms independent of human interference. Media become programmable, and transcoded information is organized in databases that structure knowledge in a non-linear manner. Manovich (2001) emphasizes the impact of technology on the social, while Lister et al. (2009) highlight the idea that technological developments reflect general social and epistemological trends.

https://doi.org/10.14765/zzf.dok-2051

Bowker (2013) is also concerned with the relation between the organization of data and the social. His notion of "cooked data" (mentioned earlier) alludes to Claude Lévi-Strauss's (1970) work *The Raw and the Cooked*. Lévi-Strauss describes a set of mythological beliefs and practices in accordance with which members of the Brazilian Bororo tribe organize their lives. In the majority of the cases, the myths are based on dichotomies that can be described as variants of the nature/society divide, where the natural was raw (honey), and the social was cooked (ashes). This binary structure, between the naturally given and the socially constructed, is also characteristic of the modern worldview (Bächle 2016), and reflected in the distinction between raw data and processed data (Bowker 2013). Moreover, as our understandings of nature project our views of ourselves (Latour 1993), the way in which databases are organized reflect social and organizational developments: early databases were hierarchically organized, followed by relational databases built on fixed structures, and more recently, object-oriented or object-relational databases and linked data (Bowker 2013).

Following these changes in both the structure and storage of information, classic linear accounts of life stories become fragmented by encoding and storing them into relational or object-oriented databases. In the course of data analysis, these fragmented data are reassembled into life stories. This time, however, life events are not necessarily ordered into an individual, linear story, but become distributed through time and space, clustered by abstract criteria and the data of others into a generic life story or into transversal views of specific aspects of various lives. The resulting diversity of non-linear references to past events constitutes both a challenge and an opportunity for historical research and storytelling.

3.2 Transcoding Written Sources into Digital Representations

In the endeavor to know and understand past events, historical research is dependent on sources, often in the form of written documents. Although the object of interest is the past event, in a strict sense the research object is inevitably only a representation or a bundle of representations of that event (and in some cases the event itself might never have taken place). In the case of the process-generated documents that we find in the archival holdings of the ITS, these are only traces of such events. Reading historical sources as sources of data or information is itself an act of interpretation and restructuring. Using these documents for an analysis supported by GIS can thus be described as a process of translating or transcoding the written representation in the source documents into a digital representation thereof.

One of the first decisions in this translation process is how to encode the data generated from the sources – a decision that has far-reaching consequences for the entire research project. In general terms, a choice has to be made be-

https://doi.org/10.14765/zzf.dok-2051

tween a *source-oriented* or a *goal-oriented* approach. While the source-oriented approach tries to keep the representation in data as close as possible to the source, the goal-oriented approach restructures the data taken from the source according to the research goal (Meroño-Peñuela et al. 2014). If we define information as data in context, in its representation the source-oriented approach maintains information by keeping the source as the context of the extracted data, while the goal-oriented approach creates information by generating from the project's goal a new context for the extracted data. The main advantage of the source-oriented approach is that it allows for more flexibility and different uses of the data in a later stage of the historical information "life cycle" (Boonstra, Breure, and Doorn 2004).[5] The main disadvantage of this approach is that great effort is required generating digital information from the sources that might not be needed for the project's specific goals. Conversely, a strict focus on a specific goal at an early stage of the information life cycle allows for more efficient data production, but restricts the possibilities, not only for redesigning the research project at a later stage but also for re-using the data in different contexts. This restriction might seem less important in the context of a research project with clear aims. However, in historical research projects, this has to be seriously considered, not least because important research questions frequently emerge from the close reading and rethinking of the sources during the research process.

Let us assume for the moment that the project would have sufficient resources to implement the source-oriented approach. A reasonable workflow would start from the archival documents – the process-generated representation of complex social interaction. It would then produce scanned images of these documents – a visual digital representation of those documents.[6] In the context of the People on the Move research consortium, the ITS has already gone to great efforts to scan its archival holdings in order to provide us with visual digital reproductions of most of the documents at hand. While the text encoded in those images can be read by human beings, it is not accessible to computers without further processing.

Thanks to recent developments in handwriting and character recognition technologies, as well as the semi-structured character of the documents, it would be possible to process these images semi-automatically into character-encoded text (for example, see Rass and Bondzio [2019] on their experience with the Gestapo-Kartei of Osnabrück). Nevertheless, such an approach would require much work in order to adequately reconstruct the complex structure of the documents. Although the IRO administration worked with forms and maintained a certain formal structure in its writings, these semi-structured texts are

[5] The historical information life cycle proposed by Boonstra, Breure, and Doorn (2004, 17–19) consist of six phases: creation; enrichment; editing; retrieval; analysis; and presentation.

[6] More precisely, it is a digital representation that is made to be represented visually.

https://doi.org/10.14765/zzf.dok-2051

far from standardized and sometimes contain handwritten notes on the side, stamps, and other remarks that contain important information. Thus, we cannot reconstruct the diverse range of documents fully automatically, and any text generated by automated processes must therefore be checked and corrected by traditional (human) means.

A common encoding approach to the textual representation of the content of the image would be the use of Extensible Markup Language (XML) according to the guidelines of the Text Encoding Initiative (TEI 2018). The TEI guidelines are a widely used standard for text encoding and provide a wide range of elements for text structure and content. The advantage of XML markup is that it is readable by humans, as it keeps the original text, adding additional information in the form of tags. As such, it resembles closely the widespread practice of highlighting portions of a printed text. In general, XML implies a data model in the form of an ordered tree that structures the document by hierarchy (A contains B), ordered adjacency (A followed by B), and co-occurrence (if A, then B). For this clearly source-oriented approach, the TEI-XML encoding of the documents would include – in addition to general structural markup – tags for persons, places, dates, and other useful information.

Despite the considerable effort it takes to transcribe the scanned image into TEI-XML, it would still be insufficient for our goal to model the itineraries of DPs. The structural markup that TEI provides – the position of a name or date in a given text – is of secondary importance when modeling itineraries, as it is the semantic connection between a specific person and a specific place on a specific date that is of the most importance. Thus, despite the huge effort of the encoding process, the relevant information would be highlighted by tags, but still not machine-readable with respect to its semantic content. For this reason, within the TEI community, various authors discuss the need to extend TEI-XML markup with Linked Open Data (LOD) encoding (Ciotti, Lana, and Tomasi 2014; Lana, Ciotti, and Tomasi 2018). XML is a powerful formalism with which to define the syntactic aspects of the markup language, but it does not provide a computational semantics and thus owes its semantic value and consistency almost entirely to human interpretation and control. The benefits of a formal description of the semantics of a markup language include "automatically deriving facts from documents, and feeding them into various inferencing or reasoning systems" (Ciotti and Tomasi 2016, 2-3).

LOD is one of the main technologies of the Semantic Web.[7] The Semantic Web was envisioned in 2001 by Berners-Lee, Hendler, and Lassila (2001) as an evolution of the existing web (based on the paradigm of the document) into a Semantic Web (based on the paradigm of structured data and meaning). In the Semantic Web, "information is given well-defined meaning, better enabling

[7] For an application of Semantic Web technology to historical data, see also Edwin Klijn's (2020) contribution in this special issue.

https://doi.org/10.14765/zzf.dok-2051

computers and people to work in cooperation" (Berners-Lee, Hendler, and Lassila 2001, 37). The terms "well-defined meaning" and "reasoning systems" obscure the fact that while computers can apply logical operations to data, they cannot understand information or meaning in the way human beings generate meaning from understanding information. This is why computers need a predefined rule conform formalized and classified meaning (i.e., "well-defined meaning") attached as data to an information item in order to apply "reasoning systems" – that is, to process data with algorithms.

Semantic Web technologies such as LOD are based on the open-world assumption (Robertson 2009; Doan, Halevy, and Ives 2012). This refers to a basic distinction made in data science between the open-world assumption and its opposite, the closed-world assumption. While the latter assumes the completeness of data – that is, a statement that cannot be verified is considered false – the former assumes the incompleteness of data – meaning that a statement that cannot be verified, nor explicitly shown to be false, is simply considered unknown. Following this logic, it is neither possible nor necessary to know everything about a subject, but desirable to discover unexpected information by connecting one's own database to information stored elsewhere. This logic is useful for historical research, as often one does not know beforehand what information will be generated from one's sources.

In general terms, LOD represents information in a graph that is built on a set of one or more "triples." These triples consist of subject, object, and predicate. This simple structure enforces the explicit encoding of semantic relations (Kräutli and Valleriani 2017, 6).[8] In combination with ontologies that standardize and define the structure of knowledge, it facilitates flexibility and interoperability between different data sources (Robertson 2009; Meroño-Peñuela et al. 2014). Ontologies in information science define the nouns representing classes of objects, and the verbs representing relations between objects. Using an established LOD model with one or more ontologies has the advantage of ensuring the re-usability of data and connecting one's own data to other existing data sources. However, this comes at the cost of increased complexity compared to the table-oriented data structures used in relational databases. Instead of interrelated tables, we have one or more graphs consisting of subject-predicate-object triples. Rather than using more or less intuitive column names to identify the content that we need to ingest into the database table, for the LOD approach, we first need to identify the relevant classes for the object and predicate in the LOD model, and then annotate the semantic information in a formalized triple format.

In comparison, relational databases are mainly a set of interlinked tables. This implies that each class of information that will be inserted into the data-

[8] "The data model enables us, but also forces us, to be specific about what we know and cannot know from our sources" (Kräutli and Valleriani 2017, 6).

https://doi.org/10.14765/zzf.dok-2051

base needs to be anticipated in the design of the database by assigning a specific column in a specific table for that piece of information. This makes it difficult to integrate new information that does not fit into the pre-configured database schema, and every change to the database schema to encompass new information will affect the whole database system. Nevertheless, a relational database system is open to new information and does not enforce a closed-world assumption of the completeness of data. Although the representation of information as LOD is also structured by predefined ontologies, in contrast to relational databases, an extension of the set of elements used from an ontology or the integration of a new ontology into the system does not affect the rest of the data triples.

The Semantic Web approach has gained momentum in the last years. Various projects published collections of information as LOD. For example, in the historical domain, since 2015, the WarSampo project offers a LOD service and web portal based on approximately nine million triples about Finland during World War II (Hyvönen et al. 2016; Koho et al. 2018c) and continuously extends its scope (Koho et al. 2018a). The portal merges datasets from many institutions (National Archives, Defense Forces, Association for Military History, Finnish Literature Society, etc.) that each contain different classes of sources, such as war diaries, history books, Wikipedia entries, photographs, and magazine articles. In addition to a LOD service for data re-use, the project includes a web portal that offers different views on the available data, such as an event-based perspective, a person-based perspective, an army unit perspective, or a historical places perspective (Hyvönen et al. 2016).

Various LOD-based research projects are related to prosopographical research into the history of the American Congress (Larson 2010; Miyakita, Leskinen, and Hyvönen 2018), national biographies (Hyvönen et al. 2018; Tamper et al. 2018), or the tracing of transnational mobility in national biographies.[9] Similarly, census data can be addressed by LOD approaches, such as that employed by the CEDAR project, which transformed Dutch historical censuses into LOD (Meroño-Peñuela et al. 2016). Other projects aim to contextualize individual biographies, such as the diary of the Canadian artist John Hammond, describing his journey from Montreal to the West Coast in 1871 (Robertson 2009), or Emma Goldman's travels through North America between 1910 and 1916 (Larson 2010). Other projects map networks between people, such as the Reassembling the Republic of Letters project (Tuominen et

[9] See the Cosmobilities project at IEG - Leibniz Institute for European History, Mainz, <https://www.ieg-mainz.de/forschungsprojekte/cosmobilities_grenzueberschreitende_lebenslaeufe_in_europaeischen_nationalbiographien> (Accessed November 12, 2018). The "cosmobilities" tool (<http://search.de.dariah.eu/cosmotool/search> [Accessed November 12, 2018]) is currently not available online.

https://doi.org/10.14765/zzf.dok-2051

al. 2018),[10] or between early modern scientific books, such as the CorpusTracer project (Kräutli and Valleriani 2017).

Most of these projects link their data to Wikipedia or DBpedia (the LOD version extracted from Wikipedia) and Wikidata. These linking hubs provide a good example via which to explain the prospects and limits of the LOD promise to discover information by linking to the Semantic Web. For example, another project in the Finnish context created biographical LOD from a register of over 10,000 alumni of Norssi, the prominent Finnish high school. An examination of the matching of the resulting data to other LOD sources reveals the scope of linking open data: From 10,137 alumni, 894 (8.8%) matched to the family tree data service Geni, followed by 609 (6%) matches with Wikipedia and Wikidata. To the Virtual International Authority File (VIAF) joint service of national libraries' authority files, 135 (1.3%) entries matched, and to the Getty Institutes Union List of Artist Names (ULAN), 16 (0.16%) entries matched (Leskinen et al. 2017). This example shows that even from an elite high school, the great majority of alumni have no entry in common databases.

Florian Kräutli and Matteo Valleriani (2017) from the CorpusTracer project report that "roughly half of the people we find in our sources do not exist in Wikidata" (7). However, due to the community structure of Wikidata, they can add new entities to the database as they identify them (Kräutli and Valleriani 2017, 7). This is true for their research domain, namely early modern scientific literature. However, like all community projects, Wikipedia and Wikidata have rules that determine which person qualifies for an entry. Most of the DPs about whom we find information in the ITS sources are ordinary people who do not fit the catalog of criteria for entries in such databases. Unsurprisingly, for our test sample of about 1,100 case files with around 3,600 names, we could not find a single match for any name and date of birth with a corresponding entry in Wikipedia or ULAN, while most of the geographical names could be matched to the *Getty Thesaurus of Geographical Names*.

However, Yad Vashem's central database of Shoah victims' names is one of the few examples of data collections of ordinary people in the historical domain.[11] This database has only been made possible because of the dedication of a large number of people, substantial resources, and both political and institutional support; however, it still has no interface via which to access the information as LOD. Thus, for the People on the Move consortium, the possibility to link our project's data about persons to other relevant datasets is currently not an option. Nevertheless, as discussed, openness and flexibility to integrate new and unforeseen information into the structure of the data collection and the

[10] See also the Mapping the Republic of Letters project at Stanford University, which uses different technologies <republicofletters.stanford.edu> (Accessed January 30, 2019).

[11] See <https://yvng.yadvashem.org/> (Accessed November 12, 2018).

https://doi.org/10.14765/zzf.dok-2051

ability to present the same data from different perspectives still makes LOD an attractive choice for data management.

In the context of our project, an appropriate approach to DPs' and in general migrants' agency is to model their lives as a series of events in time and space. This event-based approach allows us to identify common patterns of movement via agglomerated geostatistical data, as well as individual moments in a migrant's life when decisive decisions worthy of more detailed investigation were made. If we understand a person's life essentially as a sequence of events, it is a logical choice to represent such lives via an event-based LOD model. Among the many approaches for modeling events (Rovera 2016), the CIDOC Conceptual Reference Model (CRM; Doerr 2003)[12] is a good choice because it is an ISO standard and maintained by the International Council of Museums, and thus supported and applied by a worldwide community of cultural heritage institutions. While an event-based model is helpful if one wishes to represent the movement of persons and objects in space and time, this approach can also make data storage more complex. For example, a photograph would be represented both as an image and as the event that created it (Koho et al. 2018b). Fortunately, like XML, LOD allows the use of different models in parallel by using prefixes to identify the respective ontology. This facilitates pragmatic solutions for the encoding of semantic information in different schemes.

This brief overview of the different layers of representation that a diligent source-oriented approach would imply, illustrates the effort and amount of resources that such an approach necessitates. A goal-oriented approach can bypass most of these layers and generate the data needed for the project from the sources directly into a database tailored to the project's purposes. Therefore, because of the very restricted resources available, the People on the Move consortium opted for a more pragmatic, goal-oriented approach that generates only the data relevant for the envisioned GIS representation and stores them in a relational database model. Despite the advantages of LOD, for pragmatic reasons, in the context of the current People on the Move project, we decided to use a relational database to record the information generated from the sources. The use of relational databases has a long, established history and features extremely robust implementations. Furthermore, the staff in charge of data ingest were already accustomed to working with this type of database. Although GIS software prefers to ingest data in table form, our database model is differentiated enough to allow us to, in a future step, transform the representation of the collected data and the semantic meaning implicit in the database structure into an explicit LOD structure. Indeed, many web services that offer LOD still run a relational database in the background, rather than a native triple store.

[12] See <http://cidoc-crm.org> (Accessed November 12, 2018).

https://doi.org/10.14765/zzf.dok-2051

While this is clearly a goal-oriented reorganization of information taken from the source, we preserve the information necessary for back linking to the original source. We also take care to maintain the literal expression for places, persons, religious affiliations, and other data as written in the source. In addition, we keep contradicting information regarding occupation, nationality, religion, etc. as is. The decision to leave the task of data control and normalization to a later step in the workflow has various benefits: It maintains a certain level of accuracy with respect to the source; in addition, it separates the process of transcoding the source from the process of interpreting the resulting data. The staff that generate data from the sources do not need to decide about the validity of information or interpret the different spellings of names, while further interpretation of the data can be made transparent and reversible by qualified historians. The preserved contradictions in the sources provide starting points for questions about the reasons why, for example, a person would state a different age, profession, religion, or nationality in different contexts. This can provide insights into the strategies used by DPs in the negotiation process.

Keeping contradicting data is alien to common database systems. Most databases are optimized for maintaining the integrity of data and expect unambiguous entries. However, in the context of historical research, a special challenge is presented by the fuzziness and inconsistency of data in the sources. We do not control the social constructions behind process-generated sources, and thus have to handle terminological and conceptual inconsistencies. As Manfred Thaller (2017, 291) points out, even a single linguistic term for a certain profession can have very different meanings – e.g., a person labeled as a blacksmith might be an artisan, or an industrial worker.[13] To handle this problem, Thaller develops a more complex formula by calculating the probabilities of certain qualities of such datasets. In his conclusions, he proposes that in such cases, "we have [to] consciously [...] avoid *general* quantitative statements" (emphasis in original), but can apply statistical methods to "narrowly defined groups" and test to what extent the results for those groups can be generalized to other groups (Thaller 2017, 302).

The "real problem" for historians is that, when looking at their sources, what they can observe are always disputable facts derived from the actual events:

> Even on the most extreme level of abstraction, the most simple statement in the historical disciplines is a statement about the consistency of three hypotheses, two related to the tradition of information, one to the relationship between the actual events. (Thaller 2017, 304)

[13] Van Leeuwen, Maas, and Miles (2004) describe a Historical International Standard Classification of Occupations, a LOD ontology to connect and normalize different historical and current terms of professions in an international context. However for obvious reasons, the ontology cannot solve the fundamental problems raised by Thaller (2017).

https://doi.org/10.14765/zzf.dok-2051

This means that the statistical abstractions we can generate from our encoded source material do not discover or prove existing cause-effect relations, but create one or more functions as representations of our hypotheses on observable co-occurrences. Nevertheless, if these abstractions lead us to new perspectives on the material that can be researched further via a close reading of the sources, they are still a valuable contribution to the research process.

3.3 Challenges of Modeling Meaningful Representations in GIS

When talking about modeling life events, we need first of all to think about what modeling means and what kind of knowledge it can produce. Arianna Ciula et al. (2018) understand modeling as a way of thinking in practice. In their opinion, models are tools for interactive inquiry that help to operationalize research questions, whereas modeling is an iterative and creative process of reasoning in which meaning is made and negotiated. In her essay on the use of integrated assessment models in climate and sustainability studies, Saskia Ellenbeck (2017, 113) distinguishes between three main approaches to models: (1) a positivist understanding of models as tools to map real-world data onto a true representation of the research object; (2) a social-constructionist understanding of models as representations of the epistemological and normative worldview of the scientists who use them; and (3) a discourse analytical understanding of models as dispositif[14] – an infrastructure of discourse that by itself produces knowledge (Ellenbeck 2017, 113). She argues that these perspectives on models are mutually exclusive, because the different epistemological foundations do not allow for the formulation of a critique from a single perspective that could be recognized from any of the other perspectives (Ellenbeck 2017, 124).

Behind the conflict between these different approaches to modeling lies a contested notion of reality. Is reality something external that can be researched and modeled as closely as possible, or is it the construct of our access to the external world? Or, is reality the contested product of our social interaction with people and things around us? Among historians, the claim for the true representation of "reality" has been contested for a long time, because the past "reality" as a possible research object is gone forever (Jenkins 1995; Goertz 2001). Instead, we have the intermediate notion of "history": a construct that somehow connects the fluid, always already past "reality" to our current discourse, and thus allows us to integrate different approaches to "reality" into one or multiple histories via specific narrative strategies. In this sense, history – although not a transmission or storage technology – has a lot in common with the characteristics of media.

[14] For a definition of the term dispositive, see, e.g., Foucault (1980).

https://doi.org/10.14765/zzf.dok-2051

From a historical perspective, we already know that the past as such is inaccessible to us. Inevitably, historians build on sources – representations of their research object – in their attempt to (re)construct historical events and their context into a meaningful history. Thus, the tension between a positivist representation of past events, a social-constructionist representation of worldviews, and the idea of representation as an actor itself in current discourse, is inherent to historians' critical evaluation of their sources. Each object inherited from the past gains this threefold character when it becomes a source. It is both a product and a representation of a past event. It is also the product of a specific past perspective on that event, as well as of the processes of transmission and preservation of the source (and the selection decisions made about what should be preserved and what should be vanished). Finally, the existence or the absence of a source, the classification of what is a source and what is not, and the choice and use of the source by the researcher make the source not only a representation of the researcher's interests and worldview, but also a dispositif that allows us to ask certain questions while foreclosing others. Historians know that the models they build from sources are, like the sources themselves, not substitutes for, but representations of, past events. Historians know that their models are entangled with their own research interests and questions. They know that such sources and models are infrastructures of discourse that delimit what knowledge can be gained by interacting with them. Nevertheless, it is the only material we can build our history on, and what makes this task a science, is the fact that historians document their decisions and reflect on the process of producing history.

As discussed in section 2.2, the data that sustains our model of DPs' itineraries is generated via a process of abstraction. The social practices of movement and sojourn, embedded in a complex web of reasons, intentions, and interactions, was first reduced to administrative documents – our sources – and is subsequently further reduced to observable, quantifiable, and calculable data within a matrix of space and time coordinates. More precisely, the documented DPs' claims of actually being at a certain place and time before the UNRRA or IRO officials are transformed into algorithmically computable data.

The process of data preparation for GIS processing can be described as one of loss: "the loss of details in reducing sources to limited data fields, the loss of accuracy in accepting approximate locations, the loss of exemplars in trimming outliers from the dataset" (Hornbake 2015, 179). However, at the same time, it is inclusive, "incorporating information from more sources than one could possibly reference in a purely textual form" (ibid.), and equals the most interesting sources to lesser sources by extracting the same set of data.[15] In effect, we have a large amount of data, but we should not forget, as Luc Boltanski and

[15] For more information on information reduction by visualization and strategies to avoid reduction, see Manovich (2011).

https://doi.org/10.14765/zzf.dok-2051

Arnaud Esquerre (2018, 22) remark, that, by reducing research to the analysis of big data, we always find a socially pre-configured item, making it impossible to include social change and the reflexivity of actors that have not already been the object of taxonomic inquiry and technical and institutional recording (ibid.).

Unlike Boltanski and Esquerre, in historical research we cannot resort to ethnographic research in order to sidestep pre-constructed datasets and generate fresh views on the subject through self-reflected observation (except for the realm of oral history). Investigating past events means relying on sources that already exist, and therefore carry with them the inscription of the terms of their production and transmission. Moreover, to generate data from such sources implies the use of encoding schemas that are themselves social constructs. Historical research can only try to keep these conditions of knowledge production present in its representation of the source and investigate on both levels: the rules of discourse as well as that which is expressed within these rules. Our approach to generating and encoding data as closely to the source as feasible – including presumed misspellings, inconsistencies, and fuzziness – is aimed towards keeping the inscriptions of knowledge production visible. Thus, the traces left by the reflexivity of agents and social change in the source can be integrated into the analytical abstraction process. In this workflow, the data normalization that is necessary for further processing and interconnecting data is kept visible and reversible as an interpretative act. Our intent is to enable the dynamic generation of visual abstractions in GIS that allow researchers to react and interact with findings resulting from the hermeneutic analysis of the source material.

This points to the double character of data visualization, as defined by Stephen Few (2018): "Data visualization is the graphical display of abstract information for two purposes: sense-making (also called data analysis) and communication" (n.p.). Data is mobilized graphically. It typically requires graphical representation in order to be used as part of an explanation, or as a basis for argumentation (Gitelman and Jackson 2013, 12). To paraphrase Marshall McLuhan (1962), data visualizations are said to amplify our senses and our abilities to make sense of the world around us. Data visualizations, however, are not a neutral tool. They "come with particular 'ways of seeing,' particular analytical, mediation and narrative regimes regarding which we ought to be attentive as we use them to do research and tell stories" (Gray et al. 2016, 291). Similar to supposedly neutral data, mapping and tagging are persuasive and tend to create an understanding of constructed reality as self-evident (Bächle 2016, 150). We should keep this in mind while working with GIS-produced visualizations. Against the backdrop of various experiences with GIS-based abstraction and analysis in projects of the Social-GIS workgroup at the Univer-

https://doi.org/10.14765/zzf.dok-2051

sity of Osnabrück, I want to highlight three productive aspects of GIS-supported analysis.[16]

(1) GIS-based analysis can help to separate observable facts from the meaning commonly attributed to them, enabling new perspectives on concepts that are ideologically coagulated. The subjects of migration and refugees have been discussed at length in recent years. The definition of who is recognized as a migrant or refugee seems self-evident in these discussions and tends to be connected to further social and legal attributions. GIS model human mobility in the first instance in a technical context as movement in space. It detaches movement from its interpretation as migration. In this sense, computing is semantically blind, as Silke Schwandt (2016, 337) put it in the context of semantical historical analysis of text corpora. Semantic blindness should not be understood as objectivity. We can only see what is encoded in our system and hence, we will only find what we are searching for. However, the process of encoding and processing builds an intermediate step between hypothesis and analysis. It generates discrete fragments of observable facts that are separated from the meaning attributed to them within their original social context. Reducing their meaning to a technical function opens the mind to new encounters with the source material and allows the production of new, meaningful arrangements. For example, an ongoing project of the Social-GIS workgroup collects the birthplaces of parents and grandparents in school classes and enters this data into a GIS. The result shows that almost every family has an experience of mobility (Kaim, Panagiotidis, and Rass 2018). This observation presents an ideal starting point from which to discuss the production of migration as a phenomenon of perception and attribution. Similarly, the processing of data from the ITS archives carries with it the potential to rethink accepted knowledge on the strategies, social structures, origins, and destinations of DPs, as well as to identify new groups of DPs to be investigated. It also allows us to contextualize experiences of forced migration with previous and later experiences of different kinds of mobility.

(2) GIS-based analysis can render visible the co-occurrence of temporally and spatially dispersed events – co-occurrences that lead to questions about the interrelation of these events. In an earlier project within the Social-GIS workgroup, Sebastian Bondzio (2018) mapped the death toll of soldiers at the front during World War I with the arrival of the notification of death at the soldiers' home. The visual abstraction shows that battles known in war history due to their high mortality as main events – such as the Battle of Verdun and the Battle of the Somme in 1916 – on a local level might have far less impact on the social perception of the war than the sum of much smaller battles. In the

[16] For an overview of the Social-GIS workgroup and its projects, see <http://www.socialgis.uni-osnabrueck.de/> (Accessed November 23, 2018). Between 2015 and 2018, I was part of the Social-GIS workgroup at University of Osnabrück.

https://doi.org/10.14765/zzf.dok-2051

case of soldiers from Osnabrück, over the course of the war, at least 50% of soldiers' deaths did not occur in prominent battles and 1916 was the year with the lowest death toll (Bondzio 2018, 169). Similarly, by processing DPs' data in GIS, it is likely that otherwise imperceptible patterns in itineraries, preferences, and the success or failure of applications for resettlement programs will appear. These patterns might point to the agency of DPs – for example to negotiate their further progress, even if they did not meet the demands of receiving countries. Such patterns could thus help to reshape and understand the landscape of important hubs for refugees before, during, and after the war.

(3) From a given set of sources, GIS-based analysis generates a context for individual sources. One specific capability of GIS is to map a set of life events into a pathway through time and space, but also to agglomerate co-occurrences and show the dispersion of events within geospatial and temporal coordinates. Thus, for each representation of a life event, GIS can create a context of relations to other data representations. It can relate the representation of a single life event to representations of other events in the life of the same person, as well as to general classifications, or to representations of the life events of others in the dataset. This creates contextualized perspectives on, and access to, the individual source. From an ANT perspective, this context of patterns can be taken as modeling the work and translation of hybrid subject-object actants.

The persuasive and analytical power of these contextualized perspectives on sources and the resulting prospects for the presentation of research outcomes have been shown in various projects of the Social-GIS workgroup. For example, a project in cooperation with the Center for Persecuted Arts rendered data from artists fleeing Nazi Germany.[17] On an individual level, the GIS model could show the range before, during, and after persecution, and offers an explanation concerning the relation between the work and life of an artist as a linear story of action and reaction to his or her surroundings. Agglomerating the data from various artists, a broader context emerges, pointing towards general trends, individual deviations from that trend, probable options for decisions made, and the general as well as individual consequences. Moreover, in this context, artworks were produced before, during, and after Nazi rule. As with the artists themselves, some of these works survived, while others vanished. Mapping the movements of artists and their work, even beyond a particular artist's death, provides important insights into the artist as he or she is created by the perception of his or her work. In a similar fashion, from the dataset of various DPs, we expect to create a context for each individual case and source that can be explored through GIS-generated dynamic maps, which are linked back to digital representations of the source documents.

[17] See <https://www.geschichte.uni-osnabrueck.de/abteilungen/neueste_geschichte_und_historische_migrationsforschung/arbeitsgruppen/social_gis/follow_people_trace_art.html> (Accessed November 23, 2018).

https://doi.org/10.14765/zzf.dok-2051

The three abovementioned productive uses of GIS do not generate representations that are more truthful than written representations. The evidence they produce has to be questioned and handled with care. However, if done well, GIS can provide a point of reference for further investigation and support different views on well-known topics. A fundamental quality of digital media is the way in which it fragments knowledge into bits of information that are subsequently dynamically recombined into new insights. In this respect, GIS can help to multiply possible visions of the past and thus enrich the field of historical research.

4. Conclusion: Reconciling Hermeneutic and Quantitative Approaches through Digital History

How can DPs' agency be better understood by modeling their life events into GIS? Indeed, is it actually possible to do so? In a strict sense, the answer is no. Agency is exactly the kind of social interaction that cannot be transformed into calculable bits of information. It is the adaptation and response to the unforeseen – something that algorithms cannot understand, but only simulate by detecting patterns and calculating sophisticated statistical probabilities. Nevertheless, exactly such expectations are raised in current discourse surrounding big data and artificial intelligence. They reveal the underlying algorithmic paradigm of knowledge production inherent to the project of the Enlightenment: the subjugation of nature through the scientific description of its functional rules. This paradigm results in the idea that human behavior can also be calculated, if only we would have enough data (Bächle 2015).

Conversely, the intent to look through the source documents at the social interactions, beliefs, and reasons behind them is a challenge always present in historical research. It is the point at which the researcher's curiosity meets the sources and relates to the past with empathy, creating a historical narrative that sounds credible to the contemporary reader. This interaction transforms both the researcher and the source. Following this line of thought, history can be understood as a specific form of appropriation of the present (and not the past; Berg 2006). Paradoxically, the use of computational technologies that reduce complex social interactions to a calculable constellation of data ideal for algorithmic processing can stimulate reflection upon the interrelation between knowledge production, the medium of that knowledge production, and its representation.

By combining hermeneutic approaches with quantitative and graphical abstractions, GIS can be a useful means via which to understand DPs' agency. GIS can model traces of DPs' agency into graphical representations and discover patterns of movement and sojourn that help pose the right questions about agency, subject constitution, and frames of opportunity. Taking up

https://doi.org/10.14765/zzf.dok-2051

ANT's critique of the subject-object divide, the detected patterns can be understood as the work or agency of the institutional settings that co-constitute the DPs and eligibility officers at the IRO. Lukas Hennies, Sebastian Huhn, and Christoph Rass (2018) described this process of interaction between DPs, eligibility officers and institutional settings based on PotM data about Osnabrück. Even with small datasets, GIS representation can help to shed new light on sources and create new hypotheses with which to better understand past events (Hornbake 2015). One of the main opportunities and challenges that the digital humanities offer is the unexpected. What we think we will build and what we actually build, are often quite different (Thomas 2011). GIS can help to produce a broader context for the always limited findings generated by detailed hermeneutic research, while hermeneutic understanding can help to improve data modeling in GIS. The mixed method approach encourages historians to acquire the basic skills necessary to apply new digital technologies, or at least to understand what they are doing when working together with IT experts.

The new media of digitized sources, LOD, and interactive GIS representations also call for a rethinking of traditional habits of scholarly publication. Digital representation tends to fragment information, but also has the ability to recombine these fragments of information into new perspectives on the sources and their interrelation. Fragmentation and recombination are both interpretative acts. New digital media offer new ways of storytelling, including visual and/or non-linear stories that follow complex graphs of knowledge. History has always been a construct; a narrative constellation; the singularization of facts as well as the production of their context. Perhaps it is time for a change in scholarly culture – time to better value those new forms of publishing knowledge, provided that they are performed carefully and comprehensibly. Our mixed method approach favors the openness for new perspectives generated by algorithmic processing, combined with a critical appraisal that does not take generated representations as more truthful than any other representation. In this sense, digital history represents an opportunity to reconcile hermeneutic and quantitative approaches.

References

Agyekum, Kofi. 2006. The Sociolinguistic of Akan Personal Names. *Nordic Journal of African Studies* 15 (2): 206-35 <http://www.njas.helsinki.fi/pdf-files/vol15num2/agyekum.pdf>.

Bächle, Thomas Christian. 2015. *Mythos Algorithmus: Die Fabrikation des computerisierbaren Menschen*. Wiesbaden: Springer VS.

Bächle, Thomas Christian. 2016. *Digitales Wissen, Daten und Überwachung. Zur Einführung*. Hamburg: Junius.

Berg, Olaf. 2006. Benjamin und Deleuze: Ansätze für eine kritische Geschichtswissenschaft in Filmbildern. *Zeitschrift für kritische Theorie* Jg. 12 (22/23): 68-98 <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-210763>.

Berners-Lee, Tim, James Hendler, and Ora Lassila. 2001. The Semantic Web. *Scientific American* 284 (5): 34-43. doi: 10.1038/scientificamerican0501-34.

Boltanski, Luc, and Arnaud Esquerre. 2018. *Bereicherung: Eine Kritik der Ware*. Berlin: Suhrkamp.

Bondzio, Sebastian. 2018. Soldatentod und Kriegsgesellschaft: Eine Niedersächsische Stadt zwischen "Augusterlebnis" und totaler Niederlage (Osnabrück 1914-1918). *Niedersächsisches Jahrbuch für Landesgeschichte* 90: 159–200.

Boonstra, Onno, Leen Breure, and Peter Doorn. 2004. Past, Present and Future of Historical Information Science. *Historical Social Research* 29 (2): 4-132. doi: 10.12759/hsr.29.2004.2.4-132.

Borggräfe, Henning. 2020. Exploring Pathways of (Forced) Migration, Resettlement Structures, and Displaced Persons' Agency: Document Holdings and Research Potentials of the Arolsen Archives. *Historical Social Research* 45 (4): 45-68. doi: 10.12759/hsr.45.2020.4.45-68.

Bowker, Geoffrey C. 2005. *Memory Practices in the Sciences*. Cambridge, MA: MIT Press.

Bowker, Geoffrey C. 2013. Data Flakes: An Afterword to "Raw Data" Is an Oxymoron. In "*Raw Data*" *Is an Oxymoron*, ed. Lisa Gitelman, 167-71. Infrastructures Series. Cambridge, MA: MIT Press.

Ciotti, Fabio, Maurizio Lana, and Francesca Tomasi. 2014. TEI, Ontologies, Linked Open Data: Geolat and Beyond. *Journal of the Text Encoding Initiative [Online]* 8 (December): n. p. doi: 10.4000/jtei.1365.

Ciotti, Fabio, and Francesca Tomasi. 2016. Formal Ontologies, Linked Data, and TEI Semantics. *Journal of the Text Encoding Initiative [Online]* (9): n.p. doi: 10.4000/jtei.1480.

Ciula, Arianna, Øyvind Eide, Cristina Marras, and Patrick Sahle. 2018. Modelling: Thinking in Practice. An Introduction. *Historical Social Research Supplement* 31: 7-29. doi: 10.12759/hsr.suppl.31.2018.7-29.

Cohen, Gerard Daniel. 2012. *In War's Wake: Europe's Displaced Persons in the Postwar Order*. Oxford, UK: Oxford University Press.

Doan, AnHai, Alon Halevy, and Zachary G. Ives. 2012. *Principles of Data Integration*. Amsterdam et.al.: Morgan Kaufmann.

Doerr, Martin. 2003. The CIDOC Conceptual Reference Module: An Ontological Approach to Semantic Interoperability of Metadata. *AI Magazine* 24 (3): 75. doi: 10.1609/aimag.v24i3.1720.

Ellenbeck, Saskia. 2017. Modelle in der Klimaökonomik: Instrument, Bild oder Dispositiv? Eine wissenssoziologische Annäherung. *Leviathan* 45 (1): 111-30. doi: 10.5771/0340-0425-2017-1-111.

Few, Stephen. 2018. Data Visualization for Human Perception. In *The Encyclopedia of Human-Computer Interaction*, 2nd ed., ed. Interaction Design Foundation<https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/data-visualization-for-human-perception> (Accessed November 7, 2018).

Foucault, Michel 1980. The Confession of the Flesh (1977) interview. In *Power/Knowledge Selected Interviews and Other Writings,* ed. Colin Gordon, 194-228. New York: Vintage Books.

Gatrell, Peter. 2013. *The Making of the Modern Refugee*, Oxford, UK: Oxford University Press.

Gitelman, Lisa, and Virginia Jackson. 2013. Introduction. In *"Raw Data" Is an Oxymoron*, ed. Lisa Gitelman, 1-14. Infrastructures Series. Cambridge, MA: MIT Press.

Goertz, H.-J. 2001. *Unsichere Geschichte: Zur Theorie historischer Referentialität*. Stuttgart: Philipp Reclam jun.

Gray, Jonathan, Liliana Bounegru, Stefania Milan, and Paolo Ciuccarelli. 2016. *Ways of Seeing Data: Towards a Critical Literacy for Data Visualizations as Research Objects and Research Devices*. SSRN Scholarly Paper. Rochester, NY <https://papers.ssrn.com/abstract=2846398>.

Hennies, Lukas, Sebastian Huhn, and Christoph Rass. 2018. Gewaltinduzierte Mobilität und ihre Folgen: "Displaced Persons" in Osnabrück und die Flüchtlingskrise nach dem Zweiten Weltkrieg. *Niedersächsisches Jahrbuch Für Landesgeschichte* 90: 183-232.

Hornbake, Laure. 2015. Reduction, Approximation, and Omission: Preparing a Dataset for Visualisation. *Geschichte Und Informatik* 18/19: 169-81. doi: 10.5169/seals-685433.

Huhn, Sebastian. 2020. Negotiating Resettlement in Venezuela after World War II: An Exploration. *Historical Social Research* 45 (4): 203-225. doi: 10.12759/hsr.45.2020.4.203-225.

Hyvönen, Eero, Erkki Heino, Petri Leskinen, Esko Ikkala, Mikko Koho, Minna Tamper, Jouni Tuominen, and Eetu Mäkelä. 2016. WarSampo Data Service and Semantic Portal for Publishing Linked Open Data About the Second World War History. In *The Semantic Web: Latest Advances and New Domains*, ed. Harald Sack, Eva Blomqvist, Mathieu d'Aquin, Chiara Ghidini, Simone Paolo Ponzetto and Christoph Lange, 758-73. Lecture Notes in Computer Science 9678. Cham, Germany: Springer International Publishing. doi: 10.1007/978-3-319-34129-3_46.

Hyvönen, Eero, Petri Leskinen, Minna Tamper, Jouni Tuominen, and Kirsi Keravuori. 2018. Semantic National Biography of Finland. In *Proceedings of the Digital Humanities in the Nordic Countries 3rd Conference (DHN 2018)*, 372-85. Helsinki, Finland: CEUR Workshop Proceedings, Vol-2084 <http://www.ceur-ws.org/Vol-2084/short12.pdf>.

Jenkins, Keith. 1995. *Re-Thinking History*, reprinted. London: Routledge.

Kaim, Anna, Jannis Panagiotidis, and Christoph Rass. 2018. Familiengeschichte Als Migrationsgeschichte. Neue Perspektiven Auf Migration Und Gesellschaftlichen Wandel Anhand von Familiengeschichten Im Regionalen Kontex. *Osnabrücker Mitteilungen* 123: 275-90.

Kittler, Friedrich A. 2003. *Aufschreibesysteme 1800-1900*. 4th revised edition. München: Fink.

Klijn, Edwin. 2020. From Paper to Digital Trail: Collections on the Semantic Web. *Historical Social Research* 45 (4): 244-62. doi: 10.12759/hsr.45.2020.4.244-262.

Koho, Mikko, Erkki Heino, Esko Ikkala, Eero Hyvönen, Reijo Nikkilä, Tiia Moilanen, Katri Miettinen, and Pertti Suominen. 2018a. Integrating Prisoners of

War Dataset into the WarSampo Linked Data Infrastructure. In *Proceedings of the Digital Humanities in the Nordic Countries 3rd Conference (DHN 2018)*, Helsinki, Finland. Aachen: CEUR Workshop Proceedings <http://www.ceur-ws.org/Vol-2084>.

Koho, Mikko, Erkki Heino, Petri Leskinen, Minna Tamper, Esko Ikkala, Eetu Mäkelä, Jouni Tuominen, and Eero Hyvönen. 2018b. *Linked Open Data and Ontology Infrastructure for Second World War History*. <https://seco.cs.aalto.fi/publications/submitted/linked-open-data.pdf> (Accessed June 21, 2018).

Koho, Mikko, Esko Ikkala, Erkki Heino, and Eero Hyvönen. 2018c. Maintaining a Linked Data Cloud and Data Service for Second World War History. In *Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection. 7th International Conference, EuroMed 2018, Nicosia, Cyprus*, 138–49. Lecture Notes in Computer Science 11196. Cham, Germany: Springer-Verlag <https://doi.org/10.1007/978-3-030-01762-0_12>.

Krämer, Sybille, ed. 1998. *Medien, Computer, Realität: Wirklichkeitsvorstellungen Und Neue Medien*, Frankfurt am Main: Suhrkamp.

Kräutli, Florian, and Matteo Valleriani. 2017. CorpusTracer: A CIDOC Database for Tracing Knowledge Networks. *Digital Scholarship in the Humanities* <doi: 10.1093/llc/fqx047> (Accessed September 7, 2017).

Lana, Maurizio, Fabio Ciotti, and Francesca Tomasi. 2018. *Dialogue and Linking between TEI and Other Semantic Models* <https://www.academia.edu/27306714/Dialogue_and_linking_between_TEI_and_other_semantic_models> (Accessed May 2, 2018).

Larson, Ray R. 2010. *Bringing Lives to Light: Biography in Context* (Final project report). Berkeley: University of Berkeley <http://metadata.berkeley.edu/Biography_Final_Report.pdf> (Accessed July 21, 2018).

Latour, Bruno. 1993. *We Have Never Been Modern*, trans. Catherine Porter. New York: Harvester Wheatsheaf.

Latour, Bruno. 2005. *Reassembling the Social: An Introduction to Actor-Network-Theory*. Clarendon Lectures in Management Studies. Oxford, UK: Oxford University Press.

Latour, Bruno, and Steve Woolgar. 1986. *Laboratory Life: The Construction of Scientific Facts. With a New Postscript and Index by the Authors*. Princeton, NJ: Princeton University Press.

van Leeuwen, Marco H. D., Ineke Maas, and Andrew Miles. 2004. Creating a Historical International Standard Classification of Occupations. *Historical Methods* 37 (4): 186-97. doi: 10.3200/HMTS.37.4.186-197.

Leskinen, Petri, Jouni Tuominen, Erkki Heino, and Eero Hyvönen. 2017. An Ontology and Data Infrastructure for Publishing and Using Biographical Linked Data. In *Proceedings of the Workshop on Humanities in the Semantic Web (WHiSe II)*, 12. Vienna, Austria: CEUR Workshop Proceedings <https://seco.cs.'aalto.fi/publications/2017/leskinen-et-al-biographies-2017.pdf>.

Lévi-Strauss, Claude. 1970. *The Raw and the Cooked*, trans. John Weightman. New York: Harper & Row.

Lister, Martin, Jon Dovey, Seth Giddings, Iain Grant, and Kieran Kelly. 2009. *New Media: A Critical Introduction*, 2nd ed. London: Routledge.

Luhmann, Niklas. 1997. *Die Gesellschaft Der Gesellschaft*. Frankfurt am Main: Suhrkamp.

https://doi.org/10.14765/zzf.dok-2051

Manovich, Lev. 2001. *The Language of New Media*. Cambridge, MA: MIT Press.
Manovich, Lev. 2011. What Is Visualisation? *Visual Studies* 26 (1): 36-49. doi: 10.1080/1472586X.2011.548488.
Marrus, Michael Robert. 1985. *The Unwanted: European Refugees in the Twentieth Century*. New York, NY and Oxford, UK: Oxford University Press.
McLuhan, Marshall. 1962. *The Gutenberg Galaxy: The Making of Typographic Man*. London: Routledge & Paul.
McLuhan, Marshall. 1964. *Understanding Media: The Extensions of Man*. New York: McGraw-Hill.
Meroño-Peñuela, Albert, Ashkan Ashkpour, Marieke van Erp, Kees Mandemakers, Leen Breure, Andrea Scharnhorst, Stefan Schlobach, and Frank van Harmelen. 2014. Semantic Technologies for Historical Research: A Survey, ed. Aldo Gangemi. *Semantic Web* 6 (6): 539-64. doi: 10.3233/SW-140158.
Meroño-Peñuela, Albert, Ashkan Ashkpour, Christophe Guéret, and Stefan Schlobach. 2016. CEDAR: The Dutch Historical Censuses as Linked Open Data, ed. Pascal Hitzler. *Semantic Web* 8 (2): 297-310. doi: 10.3233/SW-160233.
Mersch, Dieter. 2008. Tertium Datur: Einleitung in Eine Negative Medientheorie. In *Was Ist Ein Medium?,* ed. Stefan Münker and Alexander Roesler. Frankfurt am Main: Suhrkamp.
Mersch, Dieter. 2009. *Medientheorien Zur Einführung*, 2nd ed. Hamburg: Junius.
Miyakita, Goki, Petri Leskinen, and Eero Hyvönen. 2018. Using Linked Data for Prosopographical Research of Historical Persons: Case U.S. Congress Legislators. In *Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection*, ed. Marinos Ioannides, Eleanor Fink, Raffaella Brumana, Petros Patias, Anastasios Doulamis, João Martins and Manolis Wallace, 11197: 150-62. Cham: Springer International Publishing. doi: 10.1007/978-3-030-01765-1_18.
Rass, Christoph, and Sebastian Bondzio. 2019. Data Driven History: Methodische Überlegungen zur Osnabrücker Gestapo-Kartei als Quelle zur Erforschung datenbasierter Herrschaft. *Archiv-Nachrichten Niedersachsen* 22:124-138.
Rass, Christoph, and Frank Wolff. 2018. What Is in a Migration Regime? Geneaological Approach and Methodological Proposal. In *Was Ist Ein Migrationsregime? What Is a Migration Regime?*, ed. Andreas Pott, Frank Wolff and Christoph Rass, 19-64. Wiesbaden: Springer VS.
Rice, Prudence M. 2007. *Maya Calendar Origins: Monuments, Mythistory, and the Materialization of Time*, William and Bettye Nowlin Series. Austin, TX: University of Texas Press.
Robertson, Bruce G. 2009. "Fawcett": A Toolkit to Begin an Historical Semantic Web. *Digital Studies / Le Champ Numérique* 1 (2): n.p. <http://www.digitalstudies.org/ojs/index.php/digital_studies/article/view/175>.
Rovera, Marco. 2016. A Knowledge-Based Framework for Events Representation and Reuse from Historical Archives. In *The Semantic Web: Latest Advances and New Domains*, 845-52. Lecture Notes in Computer Science. Cham, Germany: Springer. doi: 10.1007/978-3-319-34129-3_53.
Schwandt, Silke. 2016. Digitale Objektivität in der Geschichtswissenschaft? Oder: Kann man finden, was man nicht sucht? *Rechtsgeschichte - Legal History* 2016 (24): 337–38. doi: 10.12946/rg24/337-338.

Tamper, Minna, Petri Leskinen, Kasper Apajalahti, and Eero Hyvönen. 2018. Using Biographical Texts as Linked Data for Prosopographical Research and Applications. In *Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection. EuroMed 2018*, vol. 11196: 125-137, ed. Marinos Ioannides, Eleanor Fink, Raffaella Brumana, Petros Patias, Anastasios Doulamis, João Martins and Manolis Wallace. Lecture Notes in Computer Science. Cham, Germany: Springer-Verlag. doi: 10.1007/978-3-030-01762-0_11.

Text Encoding Initiative [TEI]. 2018. P5 Guidelines – TEI: Text Encoding Initiative <https://tei-c.org/guidelines/p5/> (Accessed October 18, 2018).

Thaller, Manfred. 2017. Source Oriented Data Processing and Quantification: Distrustful Brothers [1995]. *Historical Social Research, Supplement* 29: 287-306. doi: 10.12759/hsr.suppl.29.2017.287-306.

Thomas, William G. 2011. What We Think We Will Build and What We Build in Digital Humanities. *Journal of Digital Humanities* 1 (1): n.p. <http://journalofdigitalhumanities.org/1-1/what-we-think-we-will-build-and-what-we-build-in-digital-humanities-by-will-thomas/>.

Tuominen, Jouni, Eetu Mäkelä, Eero Hyvönen, Arno Bosse, Miranda Lewis, and Howard Hotson. 2018. Reassembling the Republic of Letters: A Linked Data Approach. In *Proceedings of the Digital Humanities in the Nordic Countries 3rd Conference (DHN 2018)*, 76-88. Helsinki, Finland: CEUR Workshop Proceedings <http://www.ceur-ws.org/Vol-2084/paper6.pdf>.

United States Holocaust Memorial Museum. 2018. United Nations Relief and Rehabilitation Administration. *Holocaust Encyclopedia* <https://encyclopedia.ushmm.org/content/en/article/united-nations-relief-and-rehabilitation-administration> (Accessed October 31, 2018).

White, H. 1973. *Metahistory: The Historical Imagination in Nineteenth-Century Europe*, vol. 1. Baltimore: Johns Hopkins University Press.

https://doi.org/10.14765/zzf.dok-2051

Historical Social Research
Historische Sozialforschung

All articles published in this Special Issue: