

Biobrary.

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

- Poster -

Hans Bauer

Library and Electronic Research Infrastructure
 Leibniz-Institute for East and Southeast
 European Studies (IOS), Regensburg, Germany
<mailto:hbauer@ios-regensburg.de>

Abstract—Biographical and bibliographic data have much in common. However, biographical information within library infrastructures focuses on primary data such as name, date of birth / date of death, etc. Thus, libraries initially provide personal data in a structured form (what differs from biodata). As digitisation and the increase of electronic services are progressing, libraries are gaining access to biographical resources beyond the only metadata. The poster refers to two Open Access resources, managed by the library of the Leibniz Institute for East and Southeast European Studies Regensburg (IOS), and discusses strategies to open them up to further enhancement: the Erik Amburger Database and the BioLexSOE online.

Keywords—biographical dictionary; biographical database; library; authority files; Open Access

I. LIBRARIES, INDEXING, AND THE PERSON

Musil's famous librarian from the Austrian State Library characterises his strategy of knowledge organization as follows: "[...] if you want to know how I know about every book here, I can tell you: Because I never read any of them. [...] With the exception of the catalogue." [1] Albeit this is of course a shortened and pointed statement, librarians like to mention this quote from the *Man without Qualities*. The reason may be that Musil touches an often unseen fact, pivotal for librarian consciousness: even more than books, even more than content libraries organize and manage data referring to them. Indexing, the art of describing entities with distinct terms – metadata – is one of the crucial librarian's tasks. Thus, libraries have developed powerful tools to identify entities in all subjects as unambiguously as possible. Authority files like the Gemeinsame Normdatei (GND) [2] combine different designation variants for named entities. The data sets contain not only preferred names for an informational object but definitions, references, relations, based on the GND-Ontology. [3] Actually, the GND manages more than 10 million

records about persons and person names. However, the way libraries deal with persons differs from the biographer's approach. Biographical information within library infrastructures initially focuses on primary data such as name, name variants, date of birth / date of death, etc. Libraries are primarily interested in identifying individual entities, and not in creating narratives. Therefore thesauri, authority files, and other classification systems are maintained. Librarian descriptions of persons are able to provide personal data, but not biographical data. [4] Personal data, according to Bourdieu and following Kripke's *Naming and Necessity*, are used as *rigid designators*, forcing identity for external (social or institutional) purpose. [5] The availability of this kind of data is required to manage big amounts of information like in libraries or in mass prosopographies like registration databases. They serve to mark up individuals and to reduce problems of redundancy and homonymy, "[...] but these basic facts do not provide a strong foundation for a more complex and interconnected depiction of lives online." [6] In summary library tools are highly appropriate for showing correlations and connections between entities / information objects, and for structuring resources referring to variant names. [7] But in contrast to biographies person data collected by libraries provide rather indications of lives than an insight into their course. Authority files like the GND are first of all indexing tools, they are not an encyclopaedia.¹ Hence, Musil's librarian's strategy not to get into the subject matter seems to reflect a professional conviction: in knowledge organisation it's all about metadata.

¹ As discussed at the Deutsche Biographie-Workshop, Historische Kommission der Bayerischen Akademie der Wissenschaften, München, 2014-12-01.

II. BIOGRAPHICAL RESOURCES AT THE LEIBNIZ-INSTITUTE FOR EAST AND SOUTHEAST EUROPEAN STUDIES (IOS)

The IOS-research library has a long tradition in deep subject indexing and producing meaningful metadata for its community.² Engaged in digitisation activities and projects to set up electronic research infrastructures the library has got the opportunity to run and to manage two (Open Access) digital resources containing certified biographical information. Resources that could not be more different: the first a fully elaborated biographical dictionary from the 1970ties to 1980ties with more than 1,500 authorised articles, the second an electronic personal archive with more than 100,000 records.

A) ONLINE-BIOGRAPHICAL DICTIONARY ON THE HISTORY OF SOUTH-EASTERN EUROPE (BIOLEXSOE ONLINE)[8]

The Biographisches Lexikon zur Geschichte Südosteuropas, printed in four volumes, 1974 to 1981, offers more than 1,500 certified biographies. The dictionary is based on relevance of the recorded persons for the history of the area from the Middle Ages to 1945. In addition to members of the Southeast European nations – from the Bosphorus to Slovakia – it includes persons from the three historic Southeast European empires (Byzantium, Ottoman Empire, Habsburg monarchy). The dictionary, although meanwhile a source document, is the only certified biographical reference work in German language for the area. Its online version was integrated into the register of the transnational Biographie Portal.[9]

Technical details: The dictionary was digitised as part of the DFG-funded project OstDok.[10] It was OCR processed and separated into its structural elements (lemmata). Individual metadata (names, life data, place of birth and death as well as occupation) were extracted, normalised, and converted into a database format.

Functionality: The metadata describing the lemmata form a faceted searchable index, which allows new search entries across the entire database (fig. 1). The metadata are used to increase the information content: e.g. via assigning coordinates to place names and visualising them on a web map (fig. 2). All metadata are freely reusable and downloadable as RDF.

Findings: The enhancement of the BioLexSOE online results from its interoperability. The lemmata were marked up with GND-authority files what improved systematization and retrievability, and solved the problem of ambiguities. The GND-Identifier are used as WWW-anchors realised by the so-called BEACON-files. “By publishing a list of GND identifiers and a concordance to subpages on a website they become interlinkable [...]”[11] with extern applications (fig. 3).

B) ERIK AMBURGER DATABASE ON FOREIGNERS IN PRE-REVOLUTIONARY RUSSIA[12]

² Outstanding examples may be the special catalogues for hidden maps (<http://geoportost.ios-regensburg.de/>) or hidden illustrations, and the database for articles.

The historian Erik Amburger³ collected data about immigrants to Russia and their descendants for more than 50 years. His archive includes over 100,000 records, mostly from urban population: it represents the world's most comprehensive data collection on foreigners in pre-revolutionary Russia, and provides a widespread picture of social status and careers of the described persons.[13]

Technical details: Originally each person entry was noted on a DIN A 7 index card (fig. 5). The cards were entered into a κλειο-database and are by now – after several conversions and migrations – stored in a MySQL database.

Functionality: The database provides a minimalist search entry by ID, name, place and occupation (fig. 6). Although recorded, it is neither possible to query the temporal dimension nor family relationships. The database lacks of normalisation why – in collaboration with the University of Regensburg – the implementation of a correction application via crowdsourcing will be released.

Findings: Amburger's archive represents more than just a migration registry; it is a certified mass prosopography. In addition to personal data like name, birth, baptism, death, and funeral dates, information on education, careers, religion, rank, and assets is also available. Of particular value is the detailed documentation of life-paths which lists places where a person has lived and which profession a person has practised (fig. 7).

III. THE NETWORKED LIBRARY – CONTEXTUALISING BIOGRAPHIES

Thaller claims that a “*digital* library, however, should be more than a *digitized* one.”[14] The introduced resources may illustrate the way to digitality and the paradigm shift libraries are faced with. It is no longer enough to merely make information available.

Both, the Amburger data and the BioLexSOE online, provide a starting point for further evaluation and for more complex queries. But therefore we have to make the information objects really – also from outside – accessible. And this is where librarian tools come on: Libraries should never manipulate or continue biographical narratives, even if they don't match with the today's state of the art. Standards, controlled vocabularies and authority files are the basis for interoperability and platform independence – in short: for contextualisation. By assigning digital information objects like the mentioned biographies to interoperable and machine readable metadata libraries can create new links, new contexts. This way they simplify international and interdisciplinary (data-, information-) exchange and contribute to a scientific quality control – and step towards a transnational comparative biography.

Furthermore, free publishing of the metadata created by libraries would – even for copyright protected resources – open up the information for digital humanities applications (fig. 4).

³ https://de.wikipedia.org/wiki/Erik_Amburger

Continuing this path libraries' digital resources will be data aggregations that form a node in a virtual knowledge network.[15]

- [1] R. Musil, *Der Mann ohne Eigenschaften*. Erstes Buch, Reinbek bei Hamburg: Rowohlt 2006²³, p.462.
- [2] http://www.dnb.de/DE/Standardisierung/GND/gnd_node.html
- [3] <https://d-nb.info/standards/elementset/gnd>
- [4] J. Zedlitz, Biographische Normdaten. Ein Überblick, in: *Archivar* 70/1 (2017), p. 22-25, http://www.archive.nrw.de/archivar/hefte/2017/Ausgabe_1/Archivar_1_2017.pdf
- [5] N.-O. Walkowski, Zur Problematik der Strukturierung und Abbildung von Personendaten in digitalen Systemen. 2009, <http://edoc.bbaw.de/frontdoor/index/index/docId/1>
- [6] P. L. Arthur, Digital Biography: Capturing Lives Online, in: *a/b: Auto/Biography Studies* 24/1 (2009), p. 74-92, here p. 84, DOI: 10.1353/abs.2009.0012
- [7] A. Aurast et al., Big Data und Smart Data in den Geisteswissenschaften, in: *Bibliothek. Forschung und Praxis* 40/2 (2016), p. 200-206, <https://doi.org/10.1515/bfp-2016-0033>
- [8] <http://www.biolex.ios-regensburg.de>
- [9] <http://www.biographie-portal.eu/>
- [10] <https://www.vifaost.de/ostdok/>
- [11] M. Reinert, M. Schrott, B. Ebner, From Biographies to Data Curation – the Making of www.deutsche-biographie.de, in: S. ter Braake et al. (eds.), *BD 2015. Biographical Data in a Digital World 2015. Proceedings of the First Conference on Biographical Data in a Digital World 2015*, Amsterdam, 2015, p. 13-19, <http://ceur-ws.org/Vol-1399/paper3.pdf>
- [12] <https://dokumente.ios-regensburg.de/amburger/>
- [13] H. Beyer-Thoma: Das Material des „Archivs Prof. Dr. Amburger“ am Osteuropa-Institut München über die Deutschen in Rußland, in: I. Zielke, R. Zielke (eds.): *Ratgeber '95 Familienforschung Mittel- und Osteuropa*. Neustadt an der Aisch 1996, p. 137-139.
- [14] M. Thaller, From the Digitized to the Digital Library, in: *D-Lib Magazine* 7/2 (2001), <http://www.dlib.org/dlib/february01/thaller/02thaller.html>
- [15] M. Kindling, e-Research und Bibliotheken, in: K. Umlauf, St. Gradmann (eds.), *Handbuch Bibliothek. Geschichte, Aufgaben, Perspektiven*: Stuttgart: Metzler, 2012, p. 146-152.

Figures

Fig. 1: BioLexSOE online: <http://www.biolex.ios-regensburg.de/BioLexViewview.php?ID=931>

The screenshot shows the BioLexSOE online interface. At the top, it identifies the user as 'Hammer-Purgstall, Josef Freiherr von, österreichischer Orientalist, *Graz 9.06.1774, †Wien 23.11.1856, Sohn eines stierischen Oberamtmannes.' Below this, there is a 'Leben' section with a detailed biography in German, mentioning his education in Graz and Vienna, his work at the Oriental Institute, and his role as a member of the Austrian Academy of Sciences. A map on the left shows the location of Graz in Austria. The interface includes navigation tabs like 'Home', 'Über das Projekt', and 'Volltext und Anmerkungen', and a search bar at the top.

Fig. 2: BioLexSOE online: named places (<http://www.biolex.ios-regensburg.de/BioLexViewview.php?ID=931>)

This screenshot shows the 'Orte' (Places) section of the BioLexSOE online interface. It lists several locations: Istanbul (Konstantinopel), Graz, Wien, Paris, Dresden, and Berlin. Below the list is a map of Europe with green star markers indicating the locations of these places. The map includes labels for countries like Poland, Romania, Italy, Greece, France, and Belgium. The interface also shows zoom controls and a map attribution to 'Leafilet | Map © Thunderforest, Data © OpenStreetMap'.

Fig. 3: BioLexSOE online: See-also-service (<http://www.biolex.ios-regensburg.de/BioLexViewview.php?ID=931>)

This screenshot shows the 'Weiterführende Information (SeeAlso)' section of the BioLexSOE online interface. It provides a list of 20 external resources related to the profile of Hammer-Purgstall, Josef Freiherr von. The list includes Wikipedia, Encyclopædia Britannica, EGO - European History Online, and various academic databases and libraries. The GND number 118545426 is displayed at the top of the section.

Fig. 4: BioLexSOE online: Visualisation birth-/deathplaces (à la Schich) with Palladio

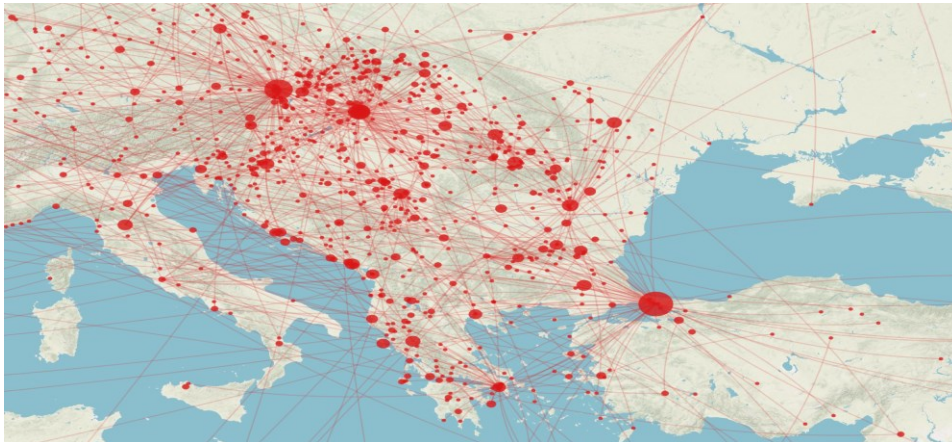


Fig. 5: Amburger database: Index card from the Amburger archive

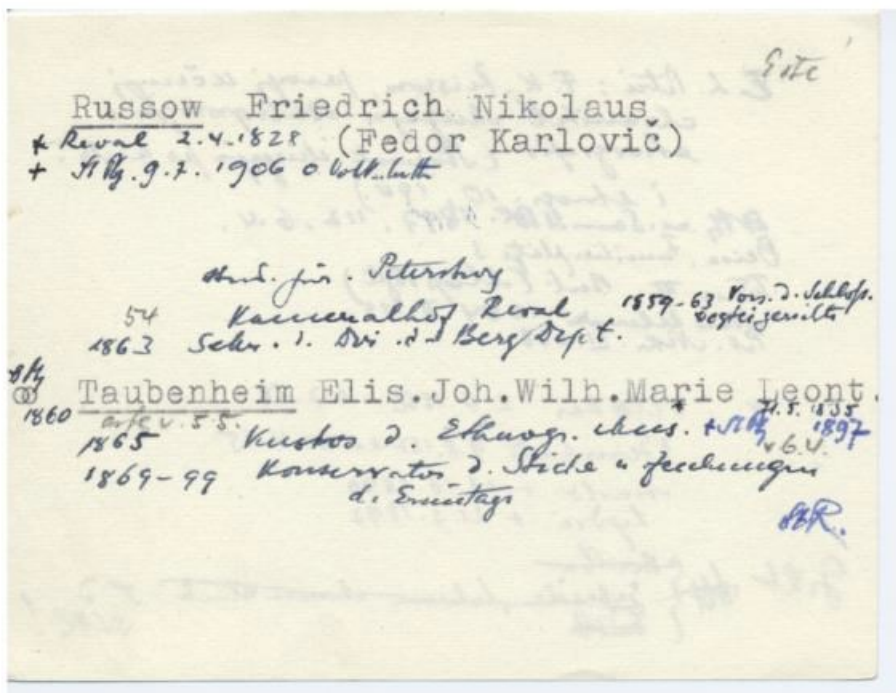


Fig. 6: Amburger database: Search interface

Erik-Amburger-Datenbank
Ausländer im vorrevolutionären Russland

IOS Institut für Ost- und
Südosteuropaforschung
Institute for East and
Southeast European Studies

Suchmaske

Suche	
Name:	<input style="width: 90%;" type="text"/>
Ort / Territorium / Land:	<input style="width: 90%;" type="text"/>
Beruf:	<input style="width: 90%;" type="text"/>
Suche starten... Suchmaske zurücksetzen	

- oder -

ID auswählen	
Datenbank-ID:	<input style="width: 90%;" type="text"/>
alte ID:	<input style="width: 90%;" type="text"/>
Auswählen	

Fig. 7: Amburger database: Paul v. Kugelgen (career)

Lebensweg	
[-]	Lebensweg(1) stammland: Rußland von-ab: 23.1.1870 beruf: an d. Revalschen Ztg. bis: 0.0.1871
[-]	Lebensweg(2) stammland: Rußland von-ab: 0.0.1871 beruf: Ans Kultamt im Senat
[-]	Lebensweg(3) stammland: Rußland beruf: Hauslehrer bei Fst. Dondukov-Korsakov
[-]	Lebensweg(4) stammland: Rußland beruf: Repetitor Alexander Lyzeum
[-]	Lebensweg(5) stammland: Rußland beruf: Lehrer d. dtn. Sprache Patriot. Inst.
[-]	Lebensweg(6) stammland: Rußland von-ab: 12.5.1873 beruf: Chefredakteur d. Nord Presse bis: 31.3.1874
[-]	Lebensweg(7) stammland: Rußland von-ab: 1.6.1874 beruf: Chefredakteur d. St.Pbg.Ztg.
[-]	Lebensweg(8) stammland: Rußland von-ab: 0.0.1877 beruf: Pächter u. Chefredakteur bis: 0.0.1904