

Digital Humanities and the Deutsche Biographie

Matthias Reinert // Dirk Scholz

„Team Deutsche Biographie“ @ IFLA Satellite, Berlin, 2017, Aug 17th

Deutsche Biographie

 NDB/ADB Lexika Quellen Literatur Portraits Objekte

Leaflet | © OpenStreetMap contributors, Imagery GIScience Research Group @ Heidelberg University



Project Responsibilities

- Historical Commission: Data
 - assembling and structuring
 - curation and storage
 - indexing
 - information extraction
- BSB:
 - Supply of GND identifiers
 - Web presentation
 - Long term preservation

The Data

- Digitizations and biographical index data
 - Allgemeine Deutsche Biographie (1875-1912, 55 vols)
 - Neue Deutsche Biographie (1953- [Vocke], so far 26 vols)
- Fulltext of articles (ca 50,000); starting of GND alignment
- „Hauptkartei“ (ca 180,000 adding 45,000 with GND after deduplication)
 - card index of persons not (yet) included in printed volumes
- Partner resources (ca 600,000)
 - Linking to data collections from cultural heritage institutions like archives, museums, academies, digital projects
- Information extraction from articles
 - Geographical names of places of birth / death / activity
 - Personal relations like family / genealogical relations, teacher/pupil relations

Alignment to Authority Data:

GND

- Supply of GND identifiers for
 - ADB/NDB and indexed names (ca 90,000)
 - partner resources (ca 60,000)
 - by means of automated procedures and intellectual work
- Usage of GND identifiers / data
 - addressing of the website: <https://www.deutsche-biographie.de/gnd118540238.html>
 - enhancement of search functionality (e.g. variant names)
 - integration of and linking to external resources via BEACON file mechanisms

Chances and Risks

- Chances by enabling
 - further research by new means of data mining and presentation technologies
 - automated procedures by making data available through interfaces
- Risks
 - Data Management
 - changes to fulltext
 - heterogeneity in additional information extraction
 - Presentation Features
 - integration of foreign technical components
 - other non technical externalities e.g. licence restrictions, privacy protection regulations
 - Reproducibility of content and views at a certain point of time
 - Common project risks

Data & DH

- our approach to DH
 - „transfer“ of methods
 - from information & library science (ILS)
 - adopting authority files (AF), AF-linking , AF-ontologies
 - faceted search
 - from digital philology (TEI)
 - from computational linguistics
 - from historical network analysis / research

Data & Services

WP:Beacon

- 2009/ 2010
- simple vs standardized
 - list of GND identifiers
 - corresponding local URL or
 - search-template
- BSB early adopter

Biografische Lexika/Biogramme ↑

- * Landeskunde Entdecken Online - Baden-Württemberg (LEO-BW) [2003-2014] [↗](#)
- Personendatenbank der Landesbibliographie Baden-Württemberg [2003-2014] [↗](#)
- * Filmportal [2010-] [↗](#)
- Historisches Lexikon der Schweiz (HLS) [2003-2014] Autor/in: Martin Schaub [↗](#)
- Bayerisches Musikerlexikon Online (BMLO) [2005-] [↗](#)
- Menschen aus Bayern [2003-] [↗](#)
- * Neue Deutsche Biographie (NDB) [1969] Autor/in: Fetscher, Iring (1969) [↗](#)
- I. Singer (Hg.): Jewish Encyclopedia. 1901-1906 [1901-1906] [↗](#)
- * Allgemeine Deutsche Biographie (ADB) [1875-1912] Autor/in: Erdmann, Johann [↗](#)

Quellen(nachweise) ↑

- Archivportal-D : 6 [↗](#)
- * Nachlässe der Sächsischen Landesbibliothek – Staats- und Universitätsbibliothek Dresden [↗](#)
- * Kalliope-Verbund [↗](#)
- * Deutsches Literaturarchiv Marbach - Kallias [↗](#)
- * Manuscripta Mediaevalia [↗](#)
- Personendaten-Repository der BBAW [2007-2014] : 3 [↗](#)
- Fröbel-Briefe. Personenindex der Gesamtausgabe [↗](#)
- Alfred Escher-Briefedition [↗](#)
- Behring-Nachlass digital [↗](#)
- Briefwechsel zwischen Eduard Spranger und Käthe Hadlich [↗](#)

Data & Services

- Beacon - Aggregation –

Service <http://data.deutsche-biographie.de/rest/bd/gnd/118540238>

- Open Solr <http://data.deutsche-biographie.de/beta/solr-open/>

- RDF – Mapping / Sparql

Service

<http://138.246.100.12:8888/bigdata/sparql>

- GraphDB – Mapping /

Cypher Service

<http://138.246.100.12...>

Biografische Lexika/Biogramme ↑

* Landeskunde Entdecken Online - Baden-Württemberg (LEO-BW) [2003-2014] [↗](#)

Personendatenbank der Landesbibliographie Baden-Württemberg [2003-2014] [↗](#)

* Filmportal [2010-] [↗](#)

Historisches Lexikon der Schweiz (HLS) [2003-2014] Autor/in: Martin Schaub [↗](#)

Bayerisches Musikerlexikon Online (BMLO) [2005-] [↗](#)

Menschen aus Bayern [2003-] [↗](#)

* Neue Deutsche Biographie (NDB) [1969] Autor/in: Fetscher, Iring (1969) [↗](#)

I. Singer (Hg.): Jewish Encyclopedia. 1901-1906 [1901-1906] [↗](#)

* Allgemeine Deutsche Biographie (ADB) [1875-1912] Autor/in: Erdmann [↗](#)

Quellen(nachweise) ↑

Archivportal-D : 6 [↗](#)

* Nachlässe der Sächsischen Landesbibliothek – Staats- und Universitätsbibliothek Dresden [↗](#)

* Kalliope-Verbund [↗](#)

* Deutsches Literaturarchiv Marbach - Kallias [↗](#)

* Manuscripta Mediaevalia [↗](#)

Personendaten-Repository der BBAW [2007-2014] : 3 [↗](#)

Fröbel-Briefe. Personenindex der Gesamtausgabe [↗](#)

Alfred Escher-Briefedition [↗](#)

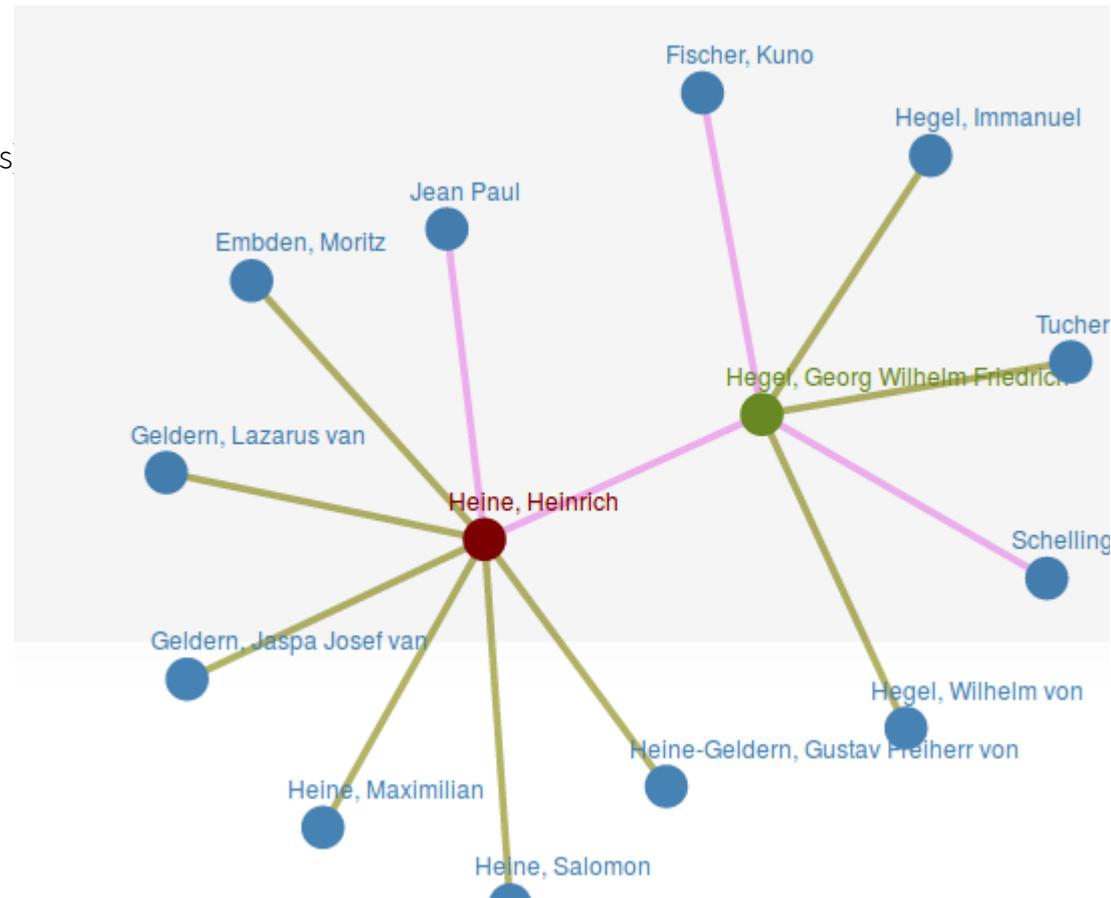
Behring-Nachlass digital [↗](#)

Briefwechsel zwischen Eduard Spranger und Käthe Hadlich [↗](#)

Data & Services

Relation Extraction with Local Grammars (Gross, Guentner, Geierhos)

- initial LG → dictionaries
- elaborated LG → disambiguated entity / relation detection and tagging (Unitex Corpus Proc.)
- entity (person|place) identification with scores / heuristics / context



Problems of Data & Services

- GND Alignment
 - process documentation
 - quality control for each resource
- Services
 - stability vs innovation
 - access control vs equal access
 - documentation / versioning
 - coherence vs ontology (mis)match
- Relations
 - precision vs recall
 - quality accounting vs results
- relying on cooperation
 - Wikimedia / Foto Marburg
 - CL specialists

OpenDtBlo



Deutsche Biographie - Open Solr Index

This API is currently under development. Please do not rely on in production environment. October 2015

API Description

The preliminary API <http://data.deutsche-biographie.de/beta/solr-open.php> connects to a Solr-based Index.

Each entry is identified by an ID. Two pattern are used `sfz[A-Zo]?[0-9-]+` and `gnd[0-9X-]+`

Returned fields parameter `f1` is predefined and fixed (as of 10-2015):

```
f1=id,defnam, defgnd, r_nam, r_bls, r_flr, bdisplay, ddisplay, bmin, bmax, dmin, dmax, r_nai, r_ndb, r_adb
```

field	data type
defnam,defgnd,r_nam,r_bls,r_flr	string (defnam,r_nam follow a "surname, forename" pattern if applicable)

Future Research Lab

ambitions:

- targeted user groups
 - casual searcher
 - DH adept
 - DH expert
 - Institutions
- customizable search
- aggregatable results set
- reproducibility
- citeability

features:

- accounts
- saveable search results
- result set operations
- charts
- complex services
- versioning