Entity Facts

A light-weight authority data service

SWIB14 – Semantic Web in Libraries Bonn, December 2nd, 2014

Dr. Christoph Böhme c.boehme@dnb.de

Michael Büchner m.buechner@dnb.de







Initial requirements from an user's point of view





Deutsche Digitale Bibliothek



Deutsche Digitale Bibliothek (DDB)



- Germany's central **portal** to all digital cultural heritage knowledge
 - sector-comprehensive
 - archive, library, monument protection, research, media, museum and others
 - interdisciplinary
 - multimedia-based
- cooperative **network** of cultural and scientific institutions
 - standardization
 - exchange of experiences
 - services
- central **platform** for applications in the cultural heritage sector
 - Application Programming Interface (API)
 - Hackathons

https://www.deutsche-digitale-bibliothek.de/





On multilingualism, metadata and Aztecs in Arabic – The...

Terms of Use



10 million units of data in the Deutsche Digitale Bibliothek...



Deutsche Digitale Bibliothek expands its social media...





Sitemap Contact



4.6.1/4.3

Copyright © 2014 Deutsche Digitale Bibliothek

https://www.deutsche-digitale-bibliothek.de/

DEUTSCHE DIGITALE BIBL Kultur und Wissen online HOME PAGE ABOUT US V HEL			ed Search
Filter facets Filter results	?	159,477 Results: Objects Persons Institutions Search results in persons Search area	ersons
Time From Day V Month Vear To Day Month Vear		1 2 3 4 5 Next ** Person page Christian Thomasius Philosoph, Jurist, Schriftsteller	
Restrict to ? Extended © Exact Apply Reset Location	>	Person page Joachim Camerarius Theologe, Gräzist, Humanist, Klassischer Philologe, Pädagoge Search results in o	bjects
Person/organisation Keyword Language Legal status	> > >	200-jähriges Bestehen der Schornsteinfegerkreisinnung Leipzig 1908 - 1909, Stadtarchiv Leipzig, 2.2.1 Handwerkerinnungen Es gilt die Archivsatzung des Stadtarchivs LeipzigStadtarchiv Leipzig, Schornsteinfegerinnung C Usability facet	₽
Usability Media type Sector		orted by frequency Page 1 Dn request (124,566) With restrictions (15,617) Without restrictions (11,233) 3, Stadtarchiv Leipzig, 2.2.1	₽ (1)
Data provider	>	Es gilt die Archivsatzung des Stadtarchivs LeipzigStadtarchiv Leipzig, Schornsteinfegerinnung A 2 Leipzig	4
Compare objects	?	Messbildstelle Dresden (Fotograf), 1991 Fotos; LeipzigWaldstr., Nr. 59, Mückenschlößchen, Leipzig Leipzig	٤

https://www.deutsche-digitale-bibliothek.de/searchresults?query=Leipzig



HOME PAGE ABOUT US ♥ HELP DISCOVER ♥

Goethe Advanced Search

𝗞 Link to this page

Johann Wolfgang von Goethe

Schriftsteller, Publizist, Politiker, Jurist, Naturwissenschaftler, Theaterintendant, Maler, Zeichner Born: August 28, 1749, <u>Frankfurt am Main</u> Died: March 22, 1832, <u>Weimar</u>

Objects: ?)











Johann Wolfgang von Goethe

- Gedichte von Johann Wolfgang von Goethe... W
 - Gedichte von Johann Wolfgang von Goethe...

Gedichte von Johann Wolfgang von Goethe...

Gedichte von Johann Wolfgang von Goethe...



Johann Wolfgang von Goethe Source: Wikimedia

Information regarding the license status of embedded media files (e.g., images or videos) may be retrieved by clicking the respective file.

Search:

"Johann Wolfgang von Goethe" in the DDB

External links:

W Wikipedia (Deutsch)

 Bibliothèque nationale de France | National Library of France

All Pictures (59) All Audios (81)

Involved in:

→ Schriften. 2

Goethe, Johann Wolfgang von. - Frankfurt [u.a.]: 1778

→ J. W. von Goethe Herzoglich Sachsen-Weimarischen Geheimraths Versuch die Metamorphose der Pflanzen zu erklären

Goethe, Johann Wolfgang von. - Gotha : Ettinger, 1790

→ Werke. 53

Goethe, Johann Wolfgang von. - Stuttgart [u.a.] : Cotta, (1833)

→ Werke. 7

Goethe, Johann Wolfgang von. - Stuttgart [u.a.] : Cotta, (1828)

All Objects (980)

Subject of:

https://www.deutsche-digitale-bibliothek.de/entity/118540238

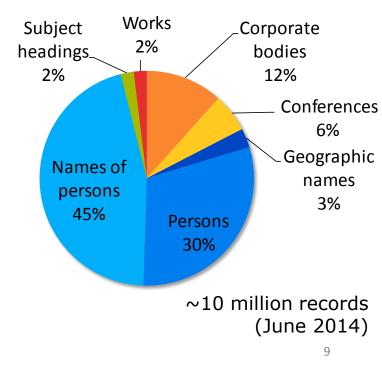
Gemeinsame Normdatei (GND)



Gemeinsame Normdatei (GND)



- Integrated Authority File
- Used by many sectors
 - libraries, archives, museums etc.
 - describing their resources
- Hosted by the German National Library (DNB)
- Run cooperatively
 - library networks in German-speaking countries
 - German Union Catalogue of Serials (ZDB)
 - Swiss National Library
 - numerous other institutions
- Problems
 - very large data dumps
 - domain specific knowledge necessary



Benefits of an authority file



- **Standardization** of access points for the description of resources
- Functional requirements
 - identify, find, represent entities and differentiate from other entities
- All variant names of an entity and attributes for its description are clustered
- Cooperative creation and reuse of records
 - efficiency of the cataloging process

That's what we wanted to have at Deutsche Digitale Bibliothek, too!



Back in early 2013



Back in early 2013

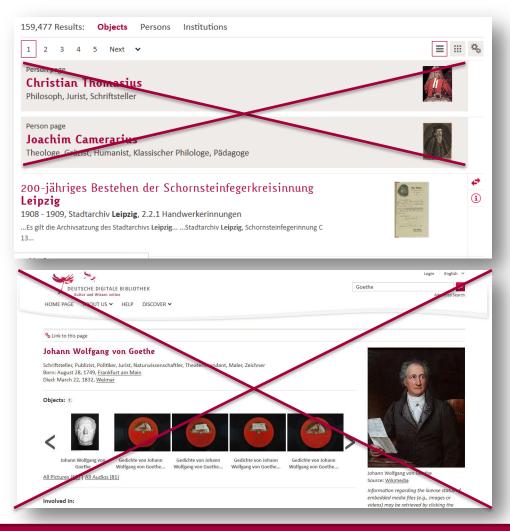


We didn't have...

- ... a search function for specific **authority data**
- ... entity pages (person pages)
 - only for registered (prospective) data providers

We did have ...

- URIs of **GND authority data** in the data of our providers
- URIs of other authority files



And we didn't have any (data) connection to GND!

Requirements



- Coverage (person)
 - names (variant names), dates of birth and death, profession or occupation
- Functional requirements
 - high quality and currentness of data
 - **images** of the entities
 - links to other portals
 - multi-lingual
- Technical requirements
 - light-weight data format
 - high availability

DEUTSCHE DIGITALE BIBLIOTNEK	
	0 1
SUCHE ÜBERUNS HILFE	Erweiterte Suche
on Link auf diese Seite	
Johann Wolfgang von Goethe	
Schriftsteller, Publicist, Publicier, kurist, Neturwissenschaftler, Theaterintendant, Maler, Zeichner Geboren: 28. August 1249, <u>Frankfuut am Main</u> Gestonber: 22. März 1832, <u>Weimar</u>	
Objekte: 1	100
Curche, Judanes Walfgare, Goeten, Judanes Walfgare, Son Star Star Star Star Star Star Star Star	>
Beteiligt an:	Johann Wolfgang von Goethe Quelle: Wikimedi
Die Wahlverwandschaften Gostine, Johann	Quelle: Wikimedia
Goethe, Johanne Li Teil 1	Suche:
	Alphanes us
Course of the second se	"Johann Wolfgang von Goethe" in der DDB
- Amman wid Developer : In mean Gestiger (Amsling, 1919) Goette, Johan Wolfgang ton, Selfer (Wingheit Bichler: Winf. Ges., 1927, [1927] Goette, Jahan Wolfgang ton, Settiger (Amslender Big Binden, g Amman Amman, Settiger (Amslender Big Binden, g	
Horizon, Kalama, Valdigana Kana, Basilgan & Gantha Samalaka, Warka Samalaka, Warka Salahar, Vari L. Gass, 1922, [1927] Gootin, Johan Wolfgang van Stattgart : Cotta, (1566)	Externe Links:
The	w Wikipedia (Deutsch)
Thema in:	
3 Johann Wolfgang von Goethe 1907, Landesarchiv Rote.	National Library of France W Wikipedia (c.,)
* Johan Weigher von Geeta 1001. Lindowski wask Beston Wittmehrer Art. Stattbarchar Signiarlegen, N 1/78 T 1 Nachlars Robert Arnaud (1885-1945) * Gester, Johan Geeta, -Lepier Bochswar 1956 * Statts, and Art. Statt 1 Status, 1956 * Gester, Stattar 1 Status, 1956	W Wikingdia (c
Prutz, Robert Felined	
Etudes sur Goethe	Wikisource
Grant Soler - Paris : Leven 1	VAF - Virtual International Authority File
Rothe, Wolfrage Witchen K.	 Historische Kommission bei der Bayerischen Akademin der
Gongen : Vandenk und Frieden	Bauer Kommission &
Mannos Anno, Santon Annaud (1885-1945) Mannos Anno, Santon Annaud (1885-1945) Roten, Volgue, - Gottiger : Wedenbee Kaptinets und Friedmitiofisung	Bayerischen Akademie der Wissenschaften und der
	Historische Kommission bei der
	Historisches Lexikon des n.e.s.
	Historisches Lexikon der Schweiz (HLS)

An easy access to the authority data of GND!



... so we asked for support ...





... and we replied:

"Well, there's our Linked Data Service"



The DNB-Linked Data Service

- It offers the complete GND
- It's RDF/XML: not domain-specific and easy to process
- It has many links to other data sets
- It's constantly updated



"No, that's not want we need, because ..."



... RDF/XML is not light-weight

- Web-applications prefer JSON over XML
- RDF/XML is expensive to parse
- RDF data is difficult to process: its much easier to work with objects than with statements and blank nodes



... the data is not suitable for presentation

- Format of names
 - The Linked Data Service offers: Goethe, Johann Wolfgang von
 - The user expects: Johann Wolfgang von Goethe
- Dates formats
 - The Linked Data Service offers ISO-formatted dates: 2014-12-02
 - The user expects a date in her current locale: 2. Dezember 2014
- Lots unnecessary information for presentation
 - Old ID numbers
 - Variant names split up in components



... it does not include data from external sources

- Links to other data sources are a good foundation
- But: Aggregating data on-the-fly from different sources is costly
 - It requires multiple requests per resource
 - The data need to be extracted and processed
- A curration process is needed



So we learned

- The Linked Data Service is great for working in a linked data environment
- But Linked Data is too heavy-weight if you just want to display some data from the linked data cloud
- A new service is needed



Entity Facts

http://www.dnb.de/EN/entityfacts



Goals of Entity Facts

A Light-weight data service

- **Easy** and **intuitive** usage \rightarrow "Zero reasons not to use it!"
 - **ready-to-use** data to display for humans \rightarrow "August 28, 1749"
 - JSON-LD over HTTP

- Regular data updates

- on-the-fly from GND database
- BEACON files
- Easy to extend

Multi-lingual

- German & English



Goals of Entity Facts

Enrichment, interlinking und visibility

- Enrichment und interlinking of the GND with ...

- external data sources like ...
 - Wikipedia, VIAF (ISNI, BNF, LoC), IMDb
- links to other resources which link to GND entities like ...
 - bibliographic records in library catalogues
- In order to ...
 - increase the visibility of GND data
 - ease the navigation to other resources



Elements of the data model

- 22 elements

- Single values: preferredName, surname, prefix, forename, academicDegree, titleOfNobility, dateOfBirth, dateOfDeath, dateOfBirthAndDeath, periodOfActivity, biographicalOrHistoricalInformation
- Arrays: variantName
- Single values with links to controlled vocabularies: placeOfBirth, placeOfDeath, placeOfActivity, gender
- Arrays with links to controlled vocabularies: professionOrOccupation, relatedPerson, familialRelationship, affiliation
- Others: depiction, sameAs

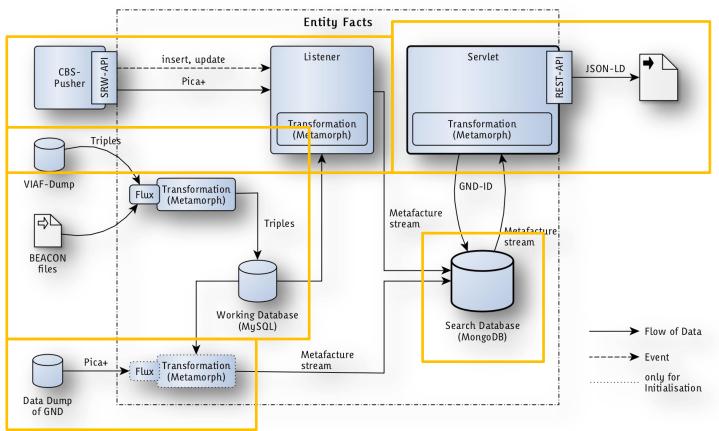


Implementation frameworks

- MongoDB
 - Document-oriented database
- Metafacture
 - Toolkit and Java library for metadata processing
 - Flux: processing metadata
 - Metamorph: transformation of metadata



Architecture





Status quo – entity information for persons

- Basic infrastructure
 - easy integration of data from other sources
 - Workflows are defined
- Images of persons from Wikipedia
- Links to other data sources
 - relations based on **BEACON** files and **data dumps** (e.g. VIAF)
- Redirecting to new records
- multilingual expressions of date values



Future developments

- Integrate with the Linked Data Service: application profiles
- Additional entity types: places and organisations
- Include more data source: as links and aggregate more data
- Extend support for multiple languages
- Refine and enhance the JSON-LD data model



Oh, and finally ...

... that's, what it looks like:

http://hub.culturegraph.org/entityfacts/118540238



Thank you!

31 | Entity Facts | SWIB 2014 – 2. December 2014, Bonn