

# 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



tomopteris [CC-BY-NC-SA 2.0], via flickr.com

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/>



Leipzig



Advanced Search



Search field

## Filter facets

**Filter results** ?

**Time**

From

Day ▼ Month ▼ Year

To

Day ▼ Month ▼ Year

Restrict to ?

☒ Extended ☐ Exact

Apply Reset

Location >

Person/organisation >

Keyword >

Language >

Legal status >

**Usability**

Media type >

Sector >

Data provider >

☐ Only results with thumbnails

159,477 Results: **Objects** Persons Institutions

1 2 3 4 5 Next ▼

Person page

**Christian Thomasius**

Philosoph, Jurist, Schriftsteller



Person page

**Joachim Camerarius**

Theologe, Gräzist, Humanist, Klassischer Philologe, Pädagoge



**200-jähriges Bestehen der Schornsteinfegerkreissinnung Leipzig**

1908 - 1909, Stadtarchiv Leipzig, 2.2.1 Handwerkerinnungen

...Es gilt die Archivsatzung des Stadtarchivs Leipzig... ...Stadtarchiv Leipzig, Schornsteinfegerinnung C



## Usability facet

Sorted by frequency Page 1

On request (124,566)

With restrictions (15,617)

Without restrictions (11,233)

...Es gilt die Archivsatzung des Stadtarchivs Leipzig... ...Stadtarchiv Leipzig, Schornsteinfegerinnung A 2...

**Leipzig**

Messbildstelle Dresden (Fotograf), 1991

...Fotos; Leipzig... ...Waldstr., Nr. 59, Mückenschlößchen, Leipzig Leipzig...



Search results in persons

Search area

Search results in objects



Link to this page

## Johann Wolfgang von Goethe

Schriftsteller, Publizist, Politiker, Jurist, Naturwissenschaftler, Theaterintendant, Maler, Zeichner

Born: August 28, 1749, [Frankfurt am Main](#)

Died: March 22, 1832, [Weimar](#)

Objects: ?



Johann Wolfgang von Goethe



Gedichte von Johann Wolfgang von Goethe...



Gedichte von Johann Wolfgang von Goethe...



Gedichte von Johann Wolfgang von Goethe...



Gedichte von Johann Wolfgang von Goethe...

[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:



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:

Wikipedia (Deutsch)

Bibliothèque nationale de France | National Library of France

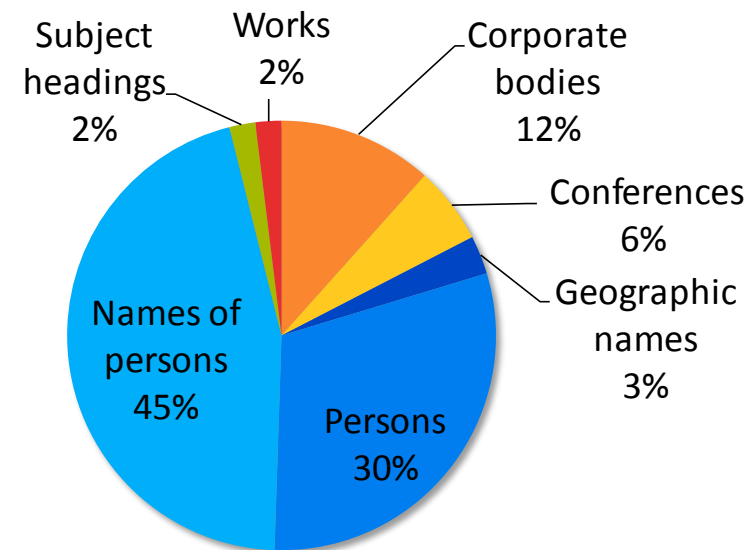
# Gemeinsame Normdatei (GND)





# 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



~10 million records  
(June 2014)

# 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



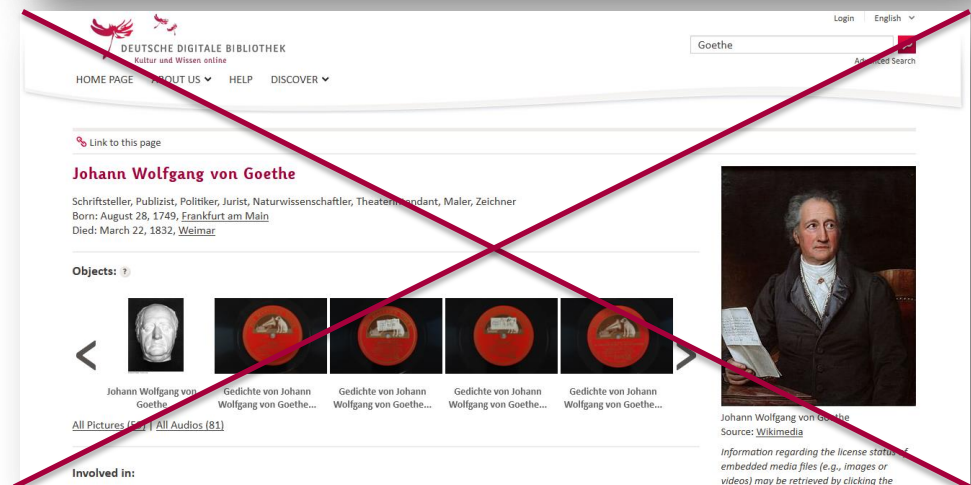
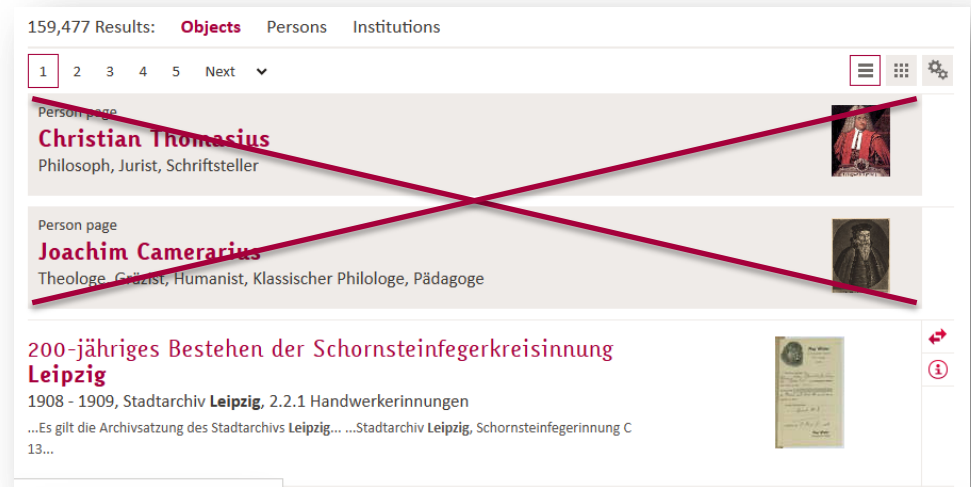
# Back in early 2013

We didn't have...

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

We did have ...

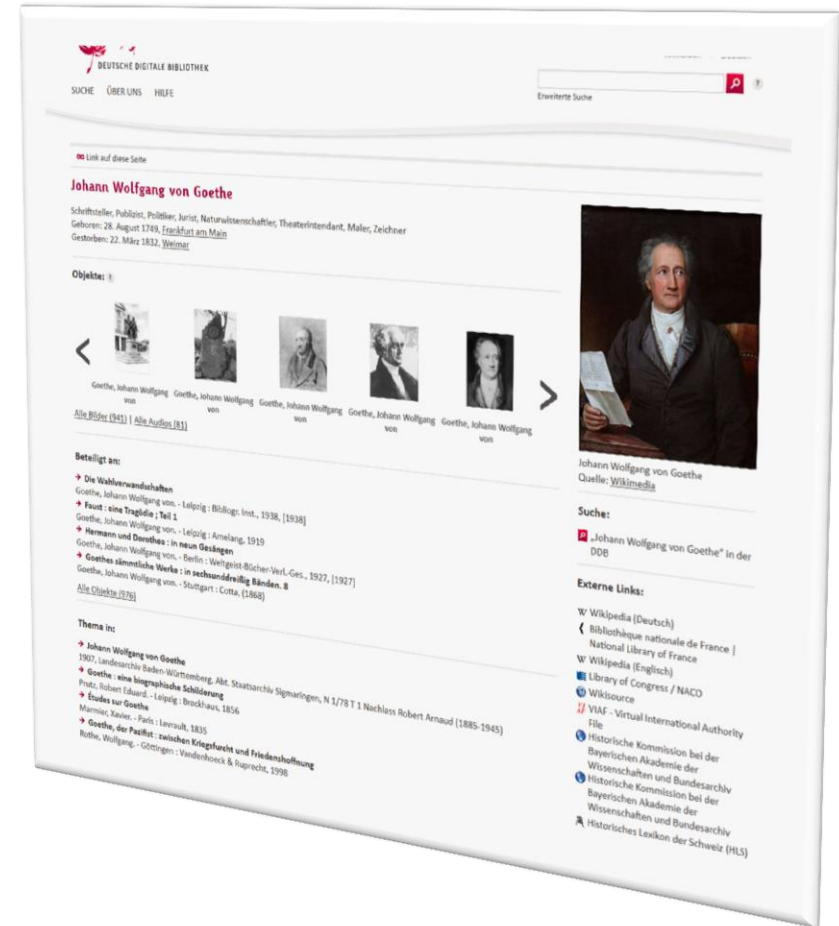
- URIs of **GND authority data** in the data of our providers
- URIs of other authority files



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



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 what 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 curation 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 → “Zero reasons not to use it!”
  - **ready-to-use** data to display for humans → “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

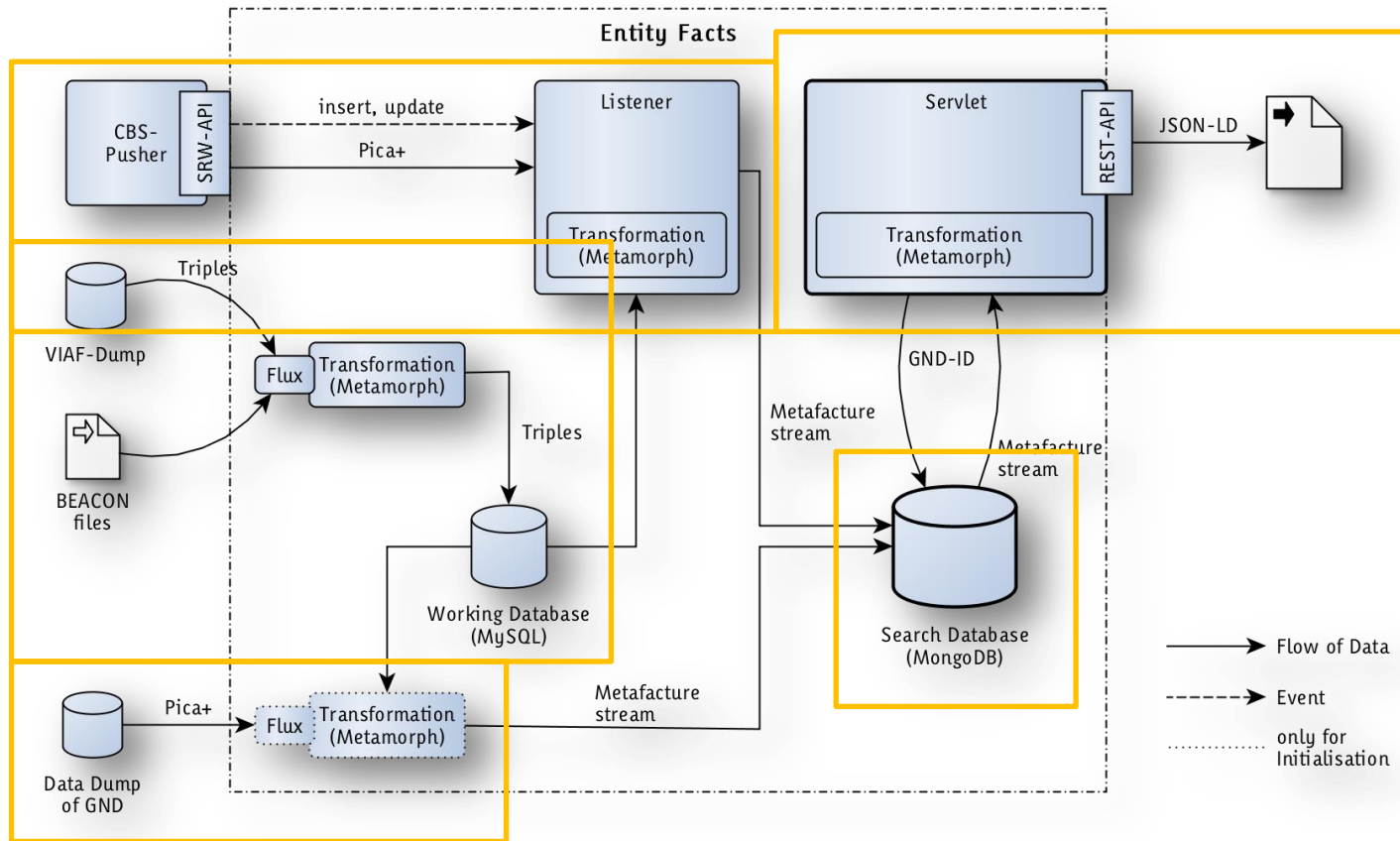
- **Single values:** preferredName, surname, prefix, forename, academicDegree, titleOfNobility, dateOfBirth, dateOfDeath, dateOfBirthAndDeath, periodOfActivity, biographicalOrHistoricalInformation
- **Arrays:** variantName
- **Single values with links to controlled vocabularies:** placeOfBirth, placeOfDeath, 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!