

A short introduction to the cooperation

GND meets Wikibase



content

1. The GND, the authority file in D-A-CH
2. GND meets Wikibase
3. First observations

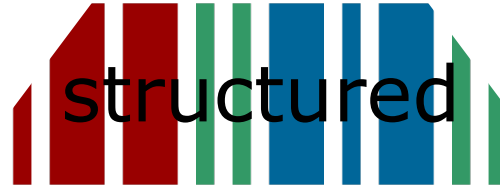
The GND, the authority file in D-A-CH

What is an authority control?

The triangle of data



authority file



structured

Wikidata



metadata

Wikipedia



all data

bits & bytes

Authority control



1. GND – The Gemeinsame Normdatei

- 16 Mio identifiers referring to names, persons, places, corporate bodies, works, headings and conferences
- A digital collaborative product embedded in a cooperative system of 1000 libraries in Germany, Austria and Switzerland is opening to GLAM, science and others
- The handy tool of librarians has to evolve into a cross domain tool: organization; data model; infrastructure & community building

GND meets Wikibase

What do we want to achieve?

Through the Labyrinth of Data

Tim Berners Lee:



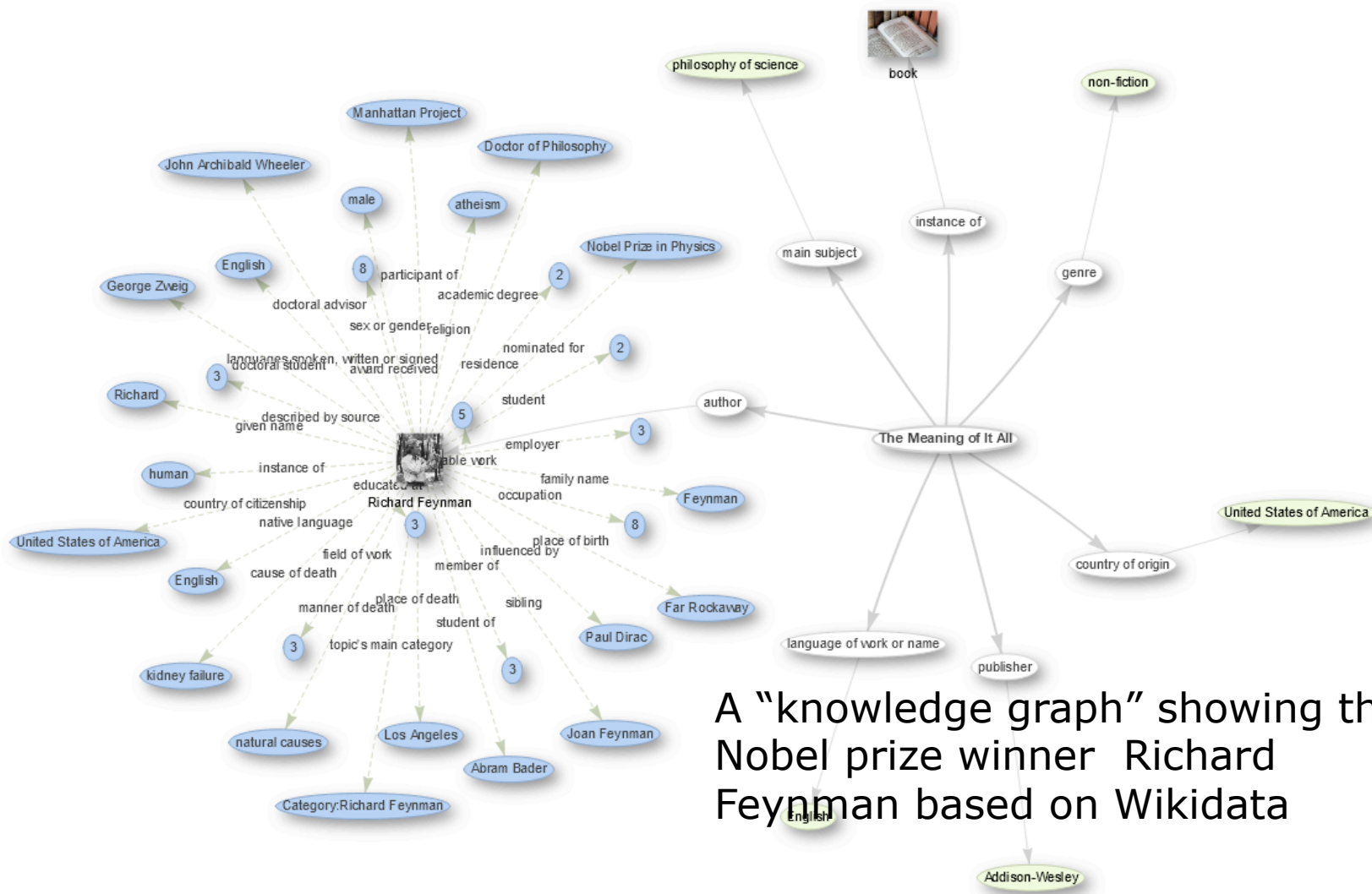
Linked Open Data is the Path

motivation



“The Semantic Web is a new form of Web content that is meaningful to computers will unleash a revolution of new possibilities” 2001





A "knowledge graph" showing the Nobel prize winner Richard Feynman based on Wikidata

CORE

Es gibt einen Kernbereich an **Eigenschaften, Relationen, Klassen** und **Regeln**, die für jede Anwendungsgemeinschaft verbindlich sind.

Im Zentrum steht das **Basismodell**.

Die **Eigenschaften, Relationen** und **Klassen** gruppieren sich mit zunehmender Differenziertheit (**Vererbungsprinzip**) um diesen Kern.

Beispiel für Plus-Regel (Erweiterung):

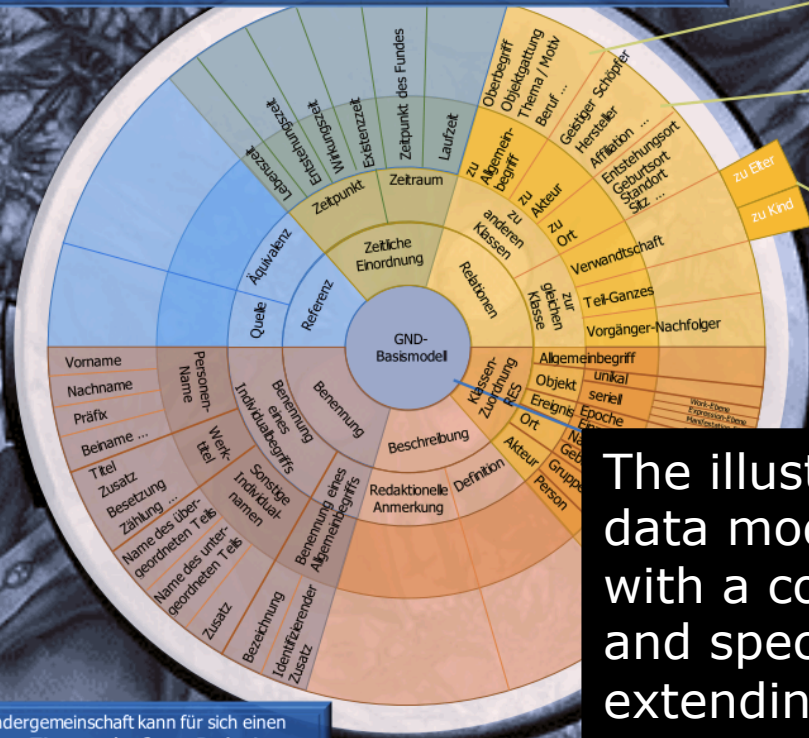
In einer Anwendungsgemeinschaft besteht ein besonderes Interesse an Personen einer bestimmten Berufsgruppe. Sie stellt für sich die Regel auf, diesen Beruf bei Personen immer zu erfassen, auch wenn dies zur eindeutigen Identifizierung der Person nicht notwendig wäre und / oder es sich nicht um den spezifischen Beruf der Person handelt.

Beispiel für Plus-Regel (Restriktion):

In einer Anwendungsgemeinschaft gibt es die Restriktion, nur eine per Regelwerk definierte Teilmenge der Akteurs-Relationen zu verwenden.

Beispiel für Plus-Relation:

In einer Anwendungsgemeinschaft besteht das Interesse, die genealogischen Beziehungen von Adelsgeschlechtern in der GND abzubilden. Der in der GND bereits definierte Relationstyp „Verwandtschaft“ ist für diesen Zweck zu unspezifisch. Die Anwendungsgemeinschaft definiert die diesem Typ untergeordneten Relationstypen „Elter“ und „Kind“ und wendet diese für genealogische Beziehungen an.



PLUS

Eine Anwendungsgemeinschaft kann für sich einen Plusbereich an **Eigenschaften, Relationen, Klassen** und **Regeln** definieren.

The illustration shows the idealic data model of the enlarged GND with a common core in the center and specifying the needs of some extending the circle

- Jede Eigenschaft bzw. Relation kann eine Quellenreferenz haben.
- ...

2. GND meets Wikibase

Two partners coinciding in two aims:

- making free structured cultural data more accessible and interoperable.
- Evaluating the software, originally conceived for a sole application, Wikidata, for a broader use

[Blog post](#)

bearing in mind: What is Wikibase?

- Wikibase is a free software by the Wikimedia Foundation to mold and collect structured data providing an URL to each item
- Wikibase is an extension of Mediawiki, the free collaborative software by the Wikimedia Foundation on which Wikipedia is based on. It includes other open source components.
- Wikibase is developed mainly by Wikimedia Deutschland to serve Wikidata

These questions are crucial to us:

- Will Wikibase meet the requirements of GLAM to implement and drive a collaborative authority file?
- Will Wikibase increase usability?
- Will the use of Wikibase make linked data easier?

First observations

What have we learned so far?

3. First observations

- Until now Wikibase was apparently used to create new databases (see “Factgrid” at University Erfurt)
- We would have loved to be able to import the GND ontology pressing a button. Every property has to be defined & matched.
- We would appreciate to be able to import relations already established in our source database.

Evaluation parameters

- Fluidity in evolution
- Expressivity of an item
- Velocity in performance
- Usability
- Stability in performance
- Scalability
- Compatibility
- Synchronization

Further criteria

- Rights- & access management
- Multi linguality
- Documentation
- Mappings
- WMDE's capacity to meet our requirements

First observations regarding the parameters

Expressivity of an item

- The ERM applied by GND is modeling in a more hierarchical way than Wikibase seems to be able to
- We need to learn how to increase the expressivity

First observations regarding the parameters

usability

- Record history is easy to follow up
- In the merging process it would be helpful to be able to to double check by presenting both records simultaneously

First observations regarding the parameters

velocity

- Entering new data, save and search shows a good performance
- To import via Quickstatements is far to slow

First observations regarding the parameters

fluidity

- To start from scratch is handy
- To import existing properties and their relations is not possible
- We will have to test further to option to adapt entity types in an ongoing process and to work with constraints on the property level

What's next?

- Checking the rights and roles assessment
- Going further into detail on constraints
- Learning more on the Wikibase capacity in its expressivity
- Learning on the usability on creating templates for user generated content
- Presenting further results on WikidataCon in October.

Discussion



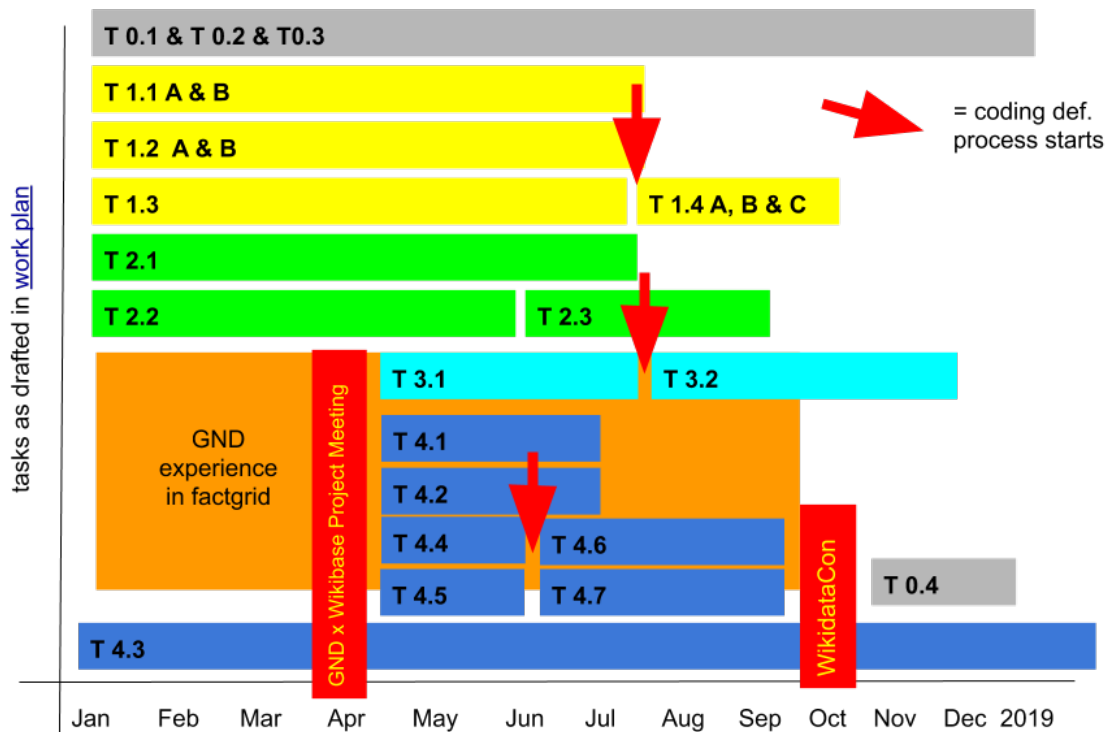
contact: b.k.fischer@dnb.de

License: Barbara Fischer (DNB AfS) [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)

What we do

- T1: Wikibase as an instance to the GND focussing on persons and corporate bodies (which is the “factgrid” focus too)
- T2: testing the expressivity of Wikibase in GND2.0
- T3: assessment of instances (access & user management)
- T4: evaluation of technical aspects in T1 and T2

Time beam



What are we doing at the moment

- We are comparing Wikidata roles to our needs
- We are testing the Ad-On [Quickstatements](#) for mass import
- We are testing a further [Ad-On](#), to transfer Marc XML into csv files, provided by Wikimedia Deutschland

Actual Focus: Format transfer

- Problem: We have ca. 15000 test GND items in MarcXML. As Wikibase only imports csv-data, we need to change the format.
- Solution: The properties are [matched](#). We have defined a procedure (see 9 examples on [Github](#)) to be adopted individually for each property.
- Result: We would like you to provide us the adopted source code to each property and paste it into the [spreadsheet](#).

Screenshot of the spreadsheet

Wikibase property	Bezeichnung / Erklärung	link zum Quellcode in github	who did it	status: not started, working on it, comple
p9	Entitätentyp		Leszek	completed
p16	Katalogisierungslevel		Leszek	completed
p15	GND-URI		Leszek	completed
p19	GND ID		Leszek	completed
p20	ORCID		Leszek	completed
p22	Land		Leszek	completed
p23	Geschlecht		Leszek	completed
p38	Geburtsort (Verknüpfung)		Leszek	completed
p39	Sterbeort (Verknüpfung)			not started
p6	Wirkungsort (Verknüpfung)			not started
p2	Geburtsort (Textstring)			not started
p41	Sterbeort (Textstring)			not started
p4	Wirkungsort (Textstring)			not started
p5	Beruf (Textstring)			not started
p8	Beruf (Verknüpfung)			not started

50 properties so far
need to be transferred