

Barbara Fischer | Sarah Hartmann

GND meets Wikibase

GND meets Wikibase

- We want to make our free structured authority data easier accessible and interoperable
- We are testing Wikibase on its functionality as toolkit for regulations



[Blog post](#)

Gemeinsame Normdatei (GND)

- authority file used by CHI (mainly libraries) in D-A-CH
 - 16 mio identifiers referring to persons, (names of persons), corporate bodies, conferences, geographic names, subject headings, works
- run cooperatively by GND agencies
 - active user: ~1.000 institutions
- Open data (CC0), APIs and documentation
- opening up to GLAM, science and others
 - the handy tool of librarians has to evolve into a cross domain tool: organization; data model; infrastructure & community building

On Wikibase

- Open source on behalf of the Wikimedia Foundation
- Developed by staff of Wikimedia Deutschland e.V.
- Based on Mediawiki
- An extension basically serving Wikidata needs
- Yet on the very start to become a standardized product

GND meets Wikibase

The project is an evaluation in two steps

- Part one: proof of concept
- Part two: testing the capacity



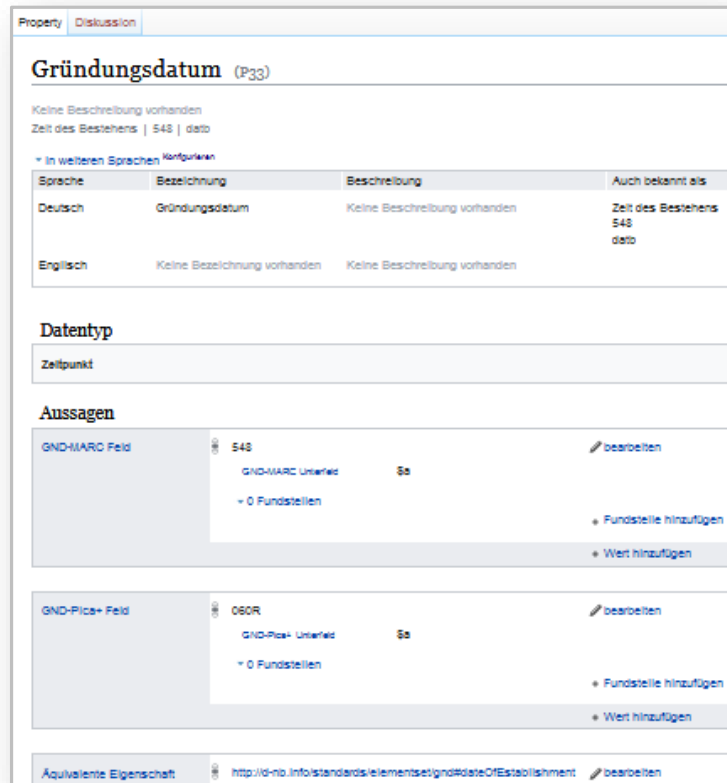
[Blog post](#)

Proof of concept – Questions in 2019

- Is Wikibase convenient for the collaborative production and maintenance of authority data?
 - Both actual GND and „GND2.0“
- Will Wikibase increase the usability in comparison to the software in use?
- Does Wikibase ease linkage to other authority data and further structured data instances?

What have we done?

- Installation of GND Wikibase instance(s)
- Defined and created properties
 - descriptive metadata for different kinds of entities
 - administrative metadata
 - mapping to GND fields: internal format (Pica) and export format (MARC 21)
 - mapping to GND Ontology and RDA properties



Property [Diskussion](#)

Gründungsdatum (P33)

Keine Beschreibung vorhanden
Zeit des Bestehens | 548 | dato

[In weiteren Sprachen konfigurieren](#)

Sprache	Bezeichnung	Beschreibung	Auch bekannt als
Deutsch	Gründungsdatum	Keine Beschreibung vorhanden	Zeit des Bestehens 548 dato
Englisch	Keine Bezeichnung vorhanden	Keine Beschreibung vorhanden	

Datentyp

Zeitpunkt

Aussagen

GND-MARC Feld 548 bearbeiten

GND-MARC Unterfeld 54

+ 0 Fundstellen Fundstelle hinzufügen

Wert hinzufügen

GND-Pica+ Feld 060R bearbeiten

GND-Pica+ Unterfeld 54

+ 0 Fundstellen Fundstelle hinzufügen

Wert hinzufügen

Äquivalente Eigenschaft http://id-nb.info/standards/elements/getgndOfEstablishment bearbeiten

What have we done?

- Created items (examples) in order to test the data model(s)
- Including items for controlled vocabs

Item
Diskussion

Angela Merkel (Q18)

Keine Beschreibung vorhanden

[In weiteren Sprachen](#)
[Konfigurieren](#)

Sprache	Bezeichnung	Beschreibung
Deutsch	Angela Merkel	Keine Beschreibung vorhanden
Englisch	Keine Bezeichnung vorhanden	Keine Beschreibung vorhanden

Aussagen

GND ID	119545373 0 Fundstellen
GND-URI	http://d-nb.info/gnd/119545373 0 Fundstellen

What have we done?

- Documentation
 - datamodel, rules, other vocabs
- Import of existing GND data
(testset of ~19300 records)
- Conceptual mapping
 - MARC21_GNDWikibase
 - Pica_GNDWikibase
- User and rights management



The screenshot shows the main page of the GND Wikibase project. At the top left is a blue geometric logo. Below it is a navigation menu with links: [Hauptseite](#), [GND-Entitätstypen](#), [GND-Datenfelder](#), [GND-Relationstypen](#), [GND-Katalogisierungslevel](#), [GND-Teilbestände](#), [GND-Systematik](#), [Bearbeiten](#), [neues Datenobjekt](#), [neue Eigenschaft](#), [Importieren](#), and [Quickstatements](#). On the right, there are tabs for [Hauptseite](#) and [Diskussion](#). The main heading is

Hauptseite

. The text below explains that GND is a community enterprise of libraries, museums, and other cultural institutions, serving the German-speaking region. It details the goal of creating a unified vocabulary for collections and resources, and mentions that GND entities are uniquely identified and linked to local catalogs to improve discoverability. It also states that GND is cooperatively managed by the Deutsche Nationalbibliothek (DNB) and that its data is available to all interested parties in MARC 21 Authority, MARC 21-xml, and RDFxml formats.

What have we done?

- Import of data
 - first approach: Quickstatements
 - second approach: python script
 - read: pymarc (<https://pypi.org/project/pymarc>)
 - write: WDI (<https://github.com/SuLab/WikidataIntegrator>)

```
data = list()

if P01 != 'UNSET': # monolingual text
    P01_obj = wdi_core.WDMonolingualText(value=P01, prop_nr='P1', language='de')
    data.append(P01_obj)

wd_item = wdi_core.WDItemEngine(data=data, new_item=True, mediawiki_api_url='http://evalbase0:18181/w/api.php')
wd_item.set_label(name, lang='de')
QID = wd_item.write(login_instance, edit_summary='initial import', retry_after=0)
```

What have we done?

– Import

- Quickstatements: 1331 records - 358 seconds
- WikidataIntegrator (WDI): 1331 records - 203 seconds -> **44%**

```
self.data = []  
if "success" in json_data and "entity" in json_data and "lastrevid" in json_data["entity"]:  
    self.lastrevid = json_data["entity"]["lastrevid"]  
# 20191202/hjb - why the hell sleep here!?!?  
#time.sleep(.5)  
return self.wd_item_id
```

– Plan for import interlinked data (incremental)

- create initial items (QIDs) for GND-IDs
- SPARQL-Output for mapping of QIDs to GND-IDs
- add statements to these items

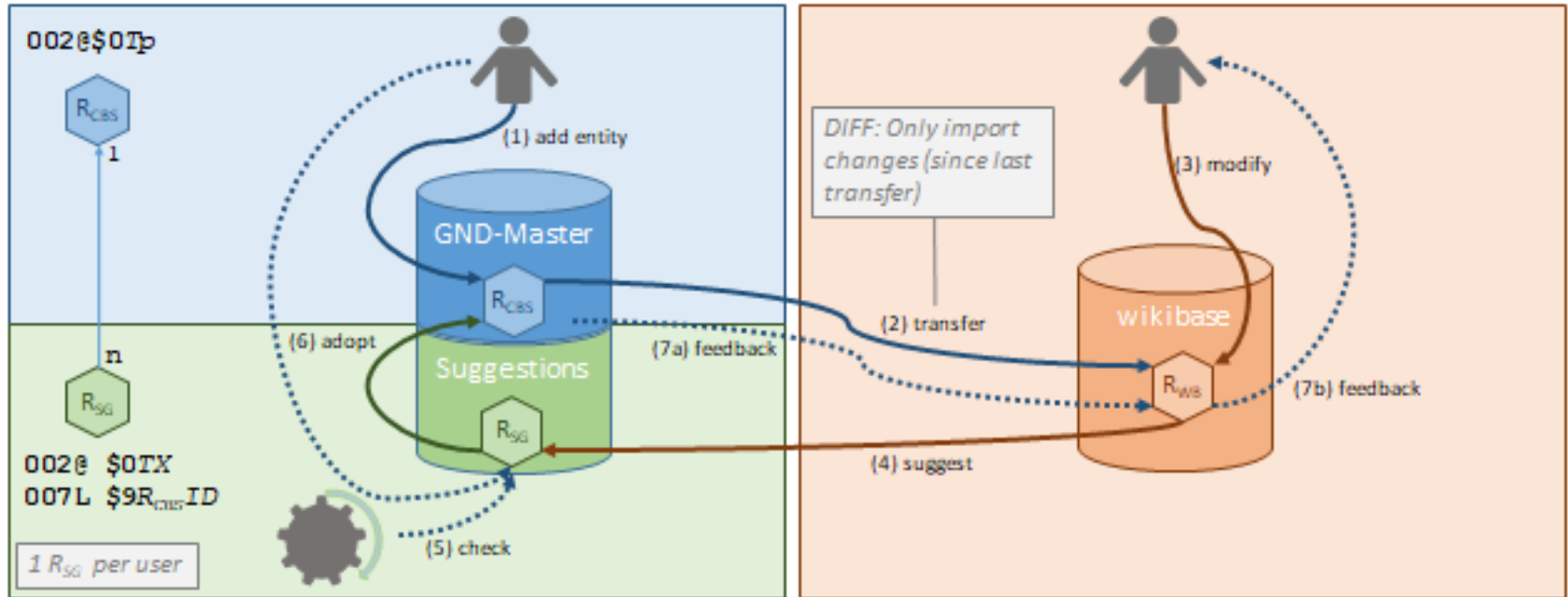
Proof of concept - learnings

- Wikibase used primarily for Wikidata or for creation of new „databases“ so far
 - import of existing (interlinked) data requires some time and effort
- Data modelling and creation of data is really easy
- Ideal solution of data model for import and export (e. g. names) needs further testing
- Creation of editing templates is more complex than expected

Part 2: Testing the capacity in 2020

- 1) “GND-all-in” mirroring 8 Mio items and their relations in the fresh installed Wikibase instance
 - import, new items, editing, redirects etc.
 - synchronization with the master base in CBS (GND)
 - compiling the requirements for User Interfaces
 - user roles and access management

CBS-Wikibase synchronization - idea



Part 2: Testing the capacity in 2020

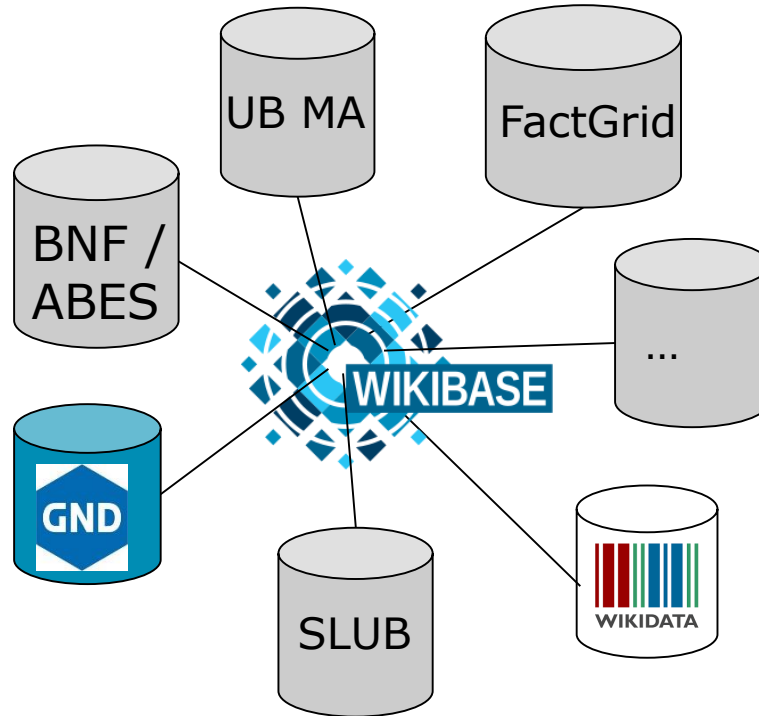
- 2) GND-documentation: make it easier to find, to understand, to follow and to change
 - creating a toolkit of rules, how-to guides und further documentation using Wikibase as a database
 - structuring and Linking all statements for each entity type together
 - access via entity type, property or relation as in Wikidata

Part 2: Testing the capacity in 2020

3) Further more topics relevant to us

- How will Wikimedia Deutschland define its governance model for the software product Wikibase beyond a community driven extension of Mediawiki?
- Who else is interested in the use of Wikibase and looking for an exchange on knowledge and skills?

Being part of something greater





Thank you!

Questions and remarks please

gnd-info@dnb.de

<http://www.dnb.de/gnd>

<http://wiki.dnb.de/display/GND>

Backup

Rules for cataloguing should be by principle independent from formats and the authority data **GND**.

The use of the GND is thus regulated separately in the blue section.

Analogously the specifics of cataloguing will be shown in the yellow section.

ET RAVI

The GND will also be usable through alternative cataloguing tools (i.e. CIDOC-CRM). We will provide specific access points

Regulation on specific topics could refer to diversified labels, relations, categories and further vocabulary

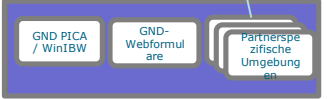
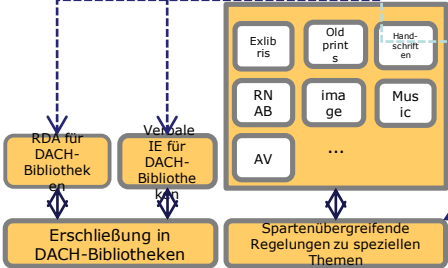
The aim is to centralize documentation on regulations and rules albeit cross domain use, traditions and requirements.

Project „3R for DACH-libraries“

Out of the requirements of librarian cataloguing rules and regulations derived for the GND.

„Common Platform to WGs“

Project „GND4C“



Grundprinzipien und -regeln (RDA & allgemeine DACH-Empfehlungen)

Core principles and rules (overview)

Erfassungsumgebungen (Formate, Systemumgebung)

Erschließung allgemein

Editing the authority file GND

Formates and Systems

