

Tracy Arndt

Practical Approaches to Large-Scale LOD

Lessons from Maintaining DNB's Linked Open Data Service

Community

Data Provider

Vocabulary

Data Producer

The library as a Data Producer

- Data about publications (bibliographic data)
- Gemeinsame Normdatei (GND) – Integrated Authority File

- Decentralized editorial team

- Uniform rules



Conference or event

<https://d-nb.info/gnd/6074048-6/about/rdf>



Work

<https://d-nb.info/gnd/7864443-4/about/rdf>



Corporate body

<https://d-nb.info/gnd/5055902-3/about/rdf>



Geographical entity

<https://d-nb.info/gnd/4011882-4/about/rdf>



Subject Heading

<https://d-nb.info/gnd/4056366-2/about/rdf>



Person

<https://d-nb.info/gnd/118514091/about/rdf>

Data Volume & Triples

- CC0 (★★★★★ Linked Data)
- GND 249.551.502 Triple*
- Bibliography 1.008.108.873 Triple*
+ x Triple
= ~ 1.4 billion triples (increasing every day)



Vocabularies & Ontologies

GND Ontology

<https://github.com/deutsche-nationalbibliothek/gnd-ontology>

DNB Metadata Terms

Agent Relationship Ontology

GND Gender

GND Geographic Area Codes

GND Subject Categories

GND Type of Coordinates

GND Description Level

<https://github.com/deutsche-nationalbibliothek/lds-vocabulary>



Data Provider

- APIs (OAI/PMH, SRU, http) 
- Dumps <https://data.dnb.de/opendata/> 
- SPARQL Endpoint (powered by qlever) <https://sparql.dnb.de/> 
- Infrastructure
 - Kubernetes for qlever

Data Provider

- Domain-specific tool for transformation from internal format to RDF



011@	Date of publication or creation
\$a	(First) year of publication or creation, sorted (YYYY)
\$b	Year of final publication or creation in sorted format (YYYY)
\$n	Date in original format, if different from the date in sorted format
\$r	Date of the original in sorted form (YYYY) (for reproductions only)



`http://purl.org/dc/terms/issued „$a“^^xsd:string`
OR

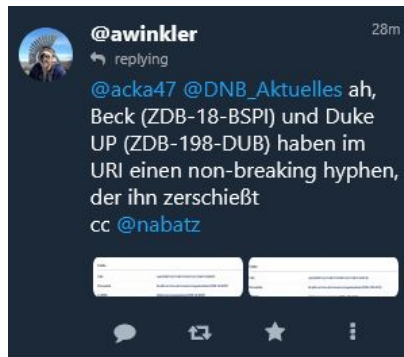
`http://purl.org/dc/terms/issued „$a-“^^xsd:string`
OR

`http://purl.org/dc/terms/issued „$a-$b“^^xsd:string`

~ 400 Fields and x subfields
= ~ 800 Rules

Community

- Researcher
- Library Services
- Data Communities



- Community Feedback
- Workshops
e.g. GND SPARQL Workshop
for the Wikidata Community
(17.7.)



Community

Researcher

Library Services

Data Communities



Data Provider

Dumps

APIs

SPARQL Endpoint



Vocabulary

Open development



Data Producer

Entities

Publications



Challenges

- Data Quality 
- Updates 
- In general: AI Bots 

Lessons learned after 16 years of LOD

- Avoid over engineering
 - LOD -> LOUD (Linked Open Usable Data)
- Automation! Automation! Automation!
- Community Feedback
- Documentation

Outlook

- Improve service and documentation
- Use our RDF data for AI projects

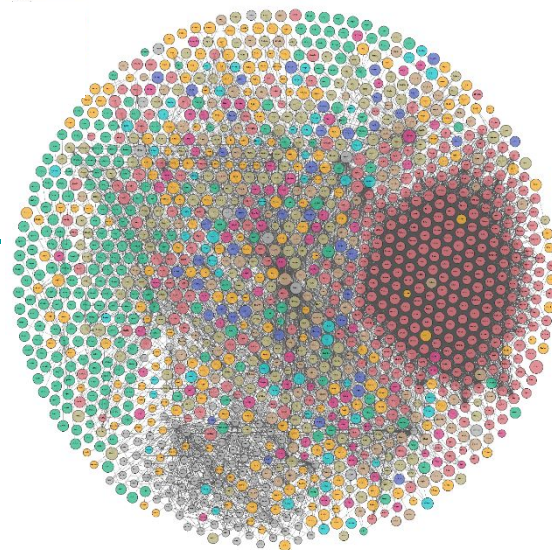
Thanks!

Linked Data Service:

- Metadata Service metadatendienste@dnb.de
- Tracy Arndt t.arndt@dnb.de ✉
@nabatz@openbiblio.social

Websites:

<https://www.dnb.de/metadatendienste>
<https://www.dnb.de/lds>
<https://data.dnb.de/opendata/>
<https://sparql.dnb.de>



The Linked Open Data Cloud from lod-cloud.net (CC-BY)