Abandon your database, use Wikidata instead!

Jakob Voß

Verbundzentrale des GBV (VZG)

ELAG "EXIT", Copenhagen, 2016-06-07/08

VZG | Verbundzentrale des GBV
www.gbv.de

Overview

Full agenda

- Tuesday 14:00-15:30 *introduction to Wikidata*
- Wednesday 11:00-13:00 *discussion of use cases*
- Thursday 13:30-14:30 *workshop report*

Agenda day 1

1. Introduction to Wikidata
   - History & goals
   - Bits & pieces
   - Exercise

2. Usage
   - Search Wikidata
   - Edit Wikidata

3. Query Wikidata

Working together

- ▶ Please interrupt me!
- ▶ Ask and help each other!
- ▶ https://etherpad.wikimedia.org/p/elag2016

Introduction to Wikidata

What is Wikidata?

Wikipedia open encyclopedia
Wikidata open database

ore specialized Wikis: Wikimedia Commons, Wikisource. . .

What is Wikidata?

- A collaboratively edited, free knowledge database that can be read and edited by to both humans and machines
- Run by Wikimedia Foundation
- Powered by Wikibase, based on MediaWiki software
  - Not as fancy as Google & Apple products
  - It's still a Wiki (revision history etc.)

Wikidata's goals

1. Centralize links between Wikipedia language editions (interwiki/sitelinks)
2. Centralize Infoboxes
3. Provide an interface for rich queries

Structure the sum of all human knowledge

History of Wikidata

- Blame Denny Vrandečić (@vrandezo, Q18618629)
- Idea around at least since 2004
- 1.3M € grant from Paul Allen, Moore & Google (4/2012)
- Development of Wikibase at Wikimedia Germany
- Launch with limited features (10/2012)
- Freebase merged into Wikidata 2015
- Ongoing development!

Core Wikidata vocabulary

- Items (Q...)
  - e.g. Q18618629 "Denny Vrandečić"
  - e.g. Q2013 "Wikidata
- Properties (P...)
  - e.g. P112 "founder"
  - *see property discussion pages*
- Pages
  - discussion, help...

Examples

- Items can be connected
  - Q856638 "library catalog"
  - P279 "subclass of"
  - Q2352616 "catalog"
- Everything is (or should be) connected
  - subclasses of catalogs
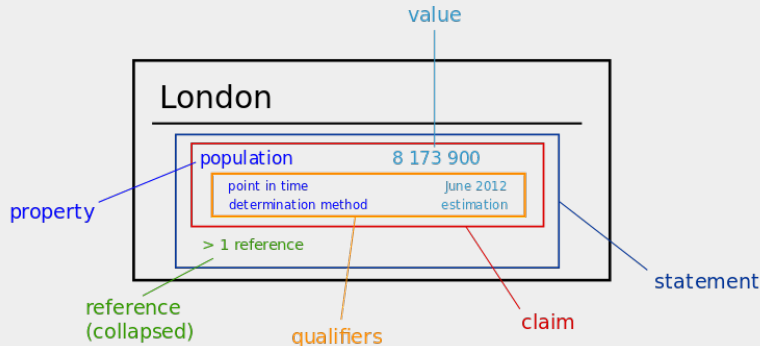
Elements of a Wikidata statement

image: CC-BY-SA User:Kaganer

Advanced Wikidata vocabulary

- ▶ Statement
    - ▶ List of claims with ranks and references
- ▶ Claim
    - ▶ item, property, value
      (subject, predicate, object)
    - ▶ can have qualifiers
- ▶ Qualifier
    - ▶ claim, property, value ("meta-claim")
    - ▶ allows for n-ary relationship

First exercise (in groups of 2-3)

- ▶ Find catalog subclasses (or sibling classes)
  without label (or unclear label) in your language(s)
- ▶ Add labels and/or descriptions!

More advanced Wikidata vocabulary

- Rank
  - preferred/normal/deprecated
  - useful to mark outdated claims

- Reference
  - source of a claim
  - for instance P248 "stated in"

- Value
  - with Datatype
  - NoValue (None)
  - SomeValue (Unknown)

`https://www.wikidata.org/wiki/Wikidata:Glossary`

Wikidata usage

First steps

- ► You already edited Wikidata!
- ► It's easier to try out than to get explained
- ► I will skip the obvious parts but
  concentrate on tips & tricks

Wikidata user account

- Global account for all Wikimedia projects
  (Wikipedia, Commons, Wikidata...)
- Applications can be granted access via OAuth
  ("connected apps")
- Gadgets & beta features

User rights

- Read and (re)use: everyone
- Edit and create items and statements: all users
- Define properties: community
- Defined datatypes: developer

Browsing Wikidata

- ► Default interface
- ► Mobile interface
- ► Reasonator
- ► Wikidata Graph Builder
- ► SQID
- ► *many more specialized views. . .*

Better search before editing!

- ▶ Find item
  - ▶ from Wikipedia
  - ▶ via labels
  - ▶ via label and language
  - ▶ via sitelinks
- ▶ Find properties
  - ▶ Advanced search
  - ▶ Browse list
    `https://www.wikidata.org/wiki/Wikidata:`
    `List_of_properties`
    `https://tools.wmflabs.org/hay/propbrowse/`
    `https://www.wikidata.org/wiki/Wikidata:`
    `Database_reports/List_of_properties/Top100`
    . . .

Exercise: Let's bring ELAG to Wikidata!

- ► What properties do events in Wikidata typically have?
  - ► Try SQID to find out!
  - ► Collect in Etherpad!

- ► Add ELAG to Wikidata!
  - ► https://elag.org/history/
  - ► Two ELAG's for each participants
  - ► See create a new item in the menu left

Exercise

- If done: join and add ELAG as overall concept
- Add instance-of, follows, followed by...

Query Wikidata

Find anything via APIs

- Wikidata API
  `https://www.wikidata.org/w/api.php`
- Wikidata Query service
  `https://query.wikidata.org`

Sample API client: wdq

```
https://metacpan.org/pod/App::wdq

wdq ELAG
wdq psearch instance
wdq '?elag wdt:P31 wd:Q...'
wdq -n ...
```
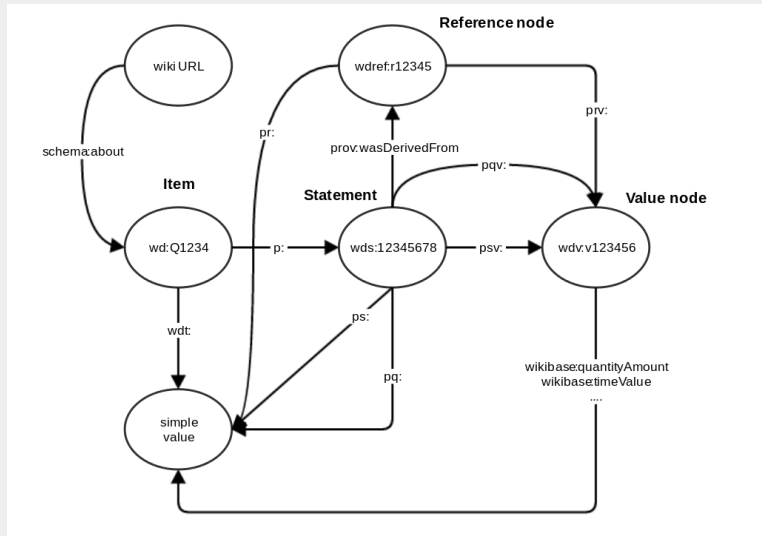
Wikidata Query Service

https://query.wikidata.org

- ▶ SPARQL access to Wikidata
- ▶ Live updates (with caching)
- ▶ Examples included
- ▶ Ongoing development
- ▶ Better integration of queries planned

I always avoided SPARQL. Then there was query.wikidata.org!

Wikidata data model in RDF

Wikidata data model in RDF

- https://www.mediawiki.org/wiki/Wikibase/DataModel
- https://www.mediawiki.org/wiki/Wikibase/Indexing/RDF_Dump_Format
- `wdq help ontology`

Exercice: SPARQL queries (in groups of 2-5)

- ► Get a list of ELAG conferences
- ► Get a list of library catalogs
- ► Get the number of countries ELAG took place in
- ► Any other interesting queries?

https://www.mediawiki.org/wiki/Wikibase/Indexing/
SPARQL_Query_Examples

Advanced SPARQL queries

Libraries with GitHub account and the administrative place they
are located in (http://tinyurl.com/gop4jnk)

```
SELECT ?item ?itemLabel ?github ?place ?placeLabel ?locatio
  ?item wdt:P31/wdt:P279* wd:Q7075 . # a library
  ?item wdt:P131 ?place .            # located in
  ?place wdt:P625 ?location .        # coordinates
  ?item wdt:P2037 ?github .          # with GitHub account
  SERVICE wikibase:label {           # add labels
    bd:serviceParam wikibase:language "en"
  }
}
```

Advanced SPARQL queries

Where do most events take place?
(http://tinyurl.com/gvbjjv4)

```
SELECT ?place ?placeLabel ?location (count(*) AS ?count) {
  ?item wdt:P31/wdt:P279* wd:Q1656682 . # an event
  ?item wdt:P131 ?place .              # located in
  ?place wdt:P625 ?location .          # coordinates
  SERVICE wikibase:label {             # add labels
    bd:serviceParam wikibase:language "en"
  }
} GROUP BY ?place ?placeLabel ?location
```

Very advanced SPARQL queries

- living people in Wikidata per country population
  (ORDER BY, GROUP BY, FILTER, subquery, arithmetic. . .)
  (source)

See you tomorrow!

ELAG Workshop: Abandon your database, use Wikidata instead! (day 2)

Agenda day 2

1. Projects and tools using Wikidata
2. Brainstorming use cases

Addendum

Mind the community

- Community portal
- Project chat
- Property proposals
- . . .
- User page discussions
- Be kind and open!

User contributions

- List user contributions
- No easy access to provenance information yet
  https://phabricator.wikimedia.org/T136799

Projects and tools

Projects build on Wikidata

- Histropedia
- The Sum of all Knowledge
- . . .

Software frameworks exist to start your own!

Tools

Long lists at
https://www.wikidata.org/wiki/Wikidata:Tools

- ▶ Reasonator nice display of Wikidata items
- ▶ SQID browse and analyze Wikidata properties and classes
- ▶ WikiData RemoteEditor
  - ▶ The Wikidata Game
  - ▶ Mix'n'match
  - ▶ ...

Authority control on Wikidata

- Properties link items to entities in external databases
  - P214 "VIAF Identifier"
  - P496 "ORCID"
  - . . .
- 970 properties with datatype external-id (of 2450 in total)

Example: BARTOC ID

- BARTOC http://bartoc.org/
  register of knowledge organization systems
- P2689 "BARTOC ID"
  - in SQID
  - database report
- Comparision of content BARTOC and Wikidata
  to improve data quality!

Brainstorming use cases

. . .

what kind of data to put in or move into Wikidata, and what consequences to expect or fear?

▶ What is actually in Wikidata?
▶ What is required to use Wikidata content?
▶ How can be measure data quality?
▶ Other arguments against putting everything in Wikidata?

Appendix

Credits & License

- ▶ Parts of the slides are copied or dervided from User:Emw's *An Ambitious Wikidata Tutorial* and *Wikidata for libraries and archives*

- ▶ My slides are available (Pandoc Markdown source & PDF) at `https://github.com/jakobib/elag2016-wikidata` under CC-BY-SA