
Integrating Wikibase into research workflows

[Daniel Mietchen](#) - [Wikibase Stakeholder Group](#) meeting - 7 July 2022
[DOI: 10.5281/zenodo.6807104](https://doi.org/10.5281/zenodo.6807104)

Structure of the talk

- Research ecosystem
 - Wikibase ecosystem
 - Use cases at the intersections
 - Discussion
-

Research ecosystem



[Research cycle](#)
[CC BY 4.0](#).

- Research cycle
 - think / plan / do / analyze / learn / share
- Researchers and research communities
 - local / national / global / disciplines / research fields
- Research infrastructure
 - coverage / quality / access / performance / security / licensing
- Research subjects & research methods
- Research questions, hypotheses and nanopublications
- Research gaps
- Research funding
- ...

See also [Wikimedia ecosystem & open science landscape](#) and [Opening up the research process](#)

Wikibase ecosystem

- Software ecosystem

- [Wikibase software stack](#)
 - What about mapping its [dependency trees](#)?
- Wikibase instances
 - test/ stable / [Wikibase registry](#)
- Wikibase farms like [Wikibase Cloud](#)
- [Wikibase extensions](#), tools and frontends

- Interoperability

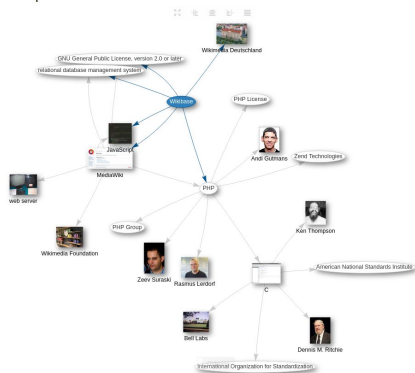
- software / data models / standards / policies/ communities

- Wikibase communities

- individual Wikibase communities / Wikibase Cloud users
- [Wikibase Community User Group](#) / [Wikibase Stakeholder Group](#)

- ...

Software dependencies



Wikibase Query Service

software: software-dependencies.org

Wikibase dependency tree, [via Scholia](#).

Use cases at the intersections (1)

OAI-PMH 2.0 Request Results

Identify	ListMetadataFormats	ListIdentifiers (oai_dc)	ListRecords (oai_dc)	Resume
Datestamp of Response	2022-07-07T12:17:19Z			
Request URL	https://enkore.toolforge.org/oai_pmh/			
Request Parameters	verb = ListRecords metadataPrefix = oai_dc			

ListRecords

The request was completed with the following results.
This is a list of records available for the metadata format oai_dc.

Record [wikidata-Q100416402](#)

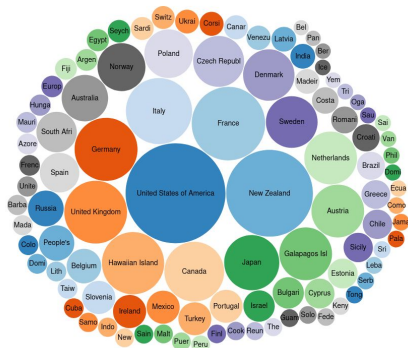
ListMetadataFormats	GetRecord
Identifier	wikidata-Q100416402
Datestamp	2022-04-14T09:46:38Z

[OAI-PMH endpoint for WikiProject Invasion biology](#)

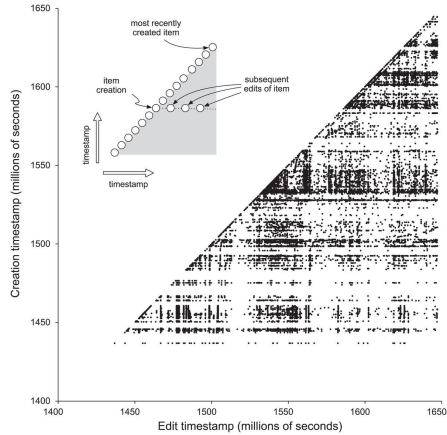
- Reproducibility/ replicability
 - documentable / reusable sample datasets
 - [100 publications from the field of invasion biology](#)
 - [this paper](#) used a [sample of 10k papers](#)
 - run queries over the sample dataset
- License incompatibilities
 - Wikibases for each of the standard licenses?
- Export a variety of formats for data sharing
 - OAI-PMH endpoint for Wikibase
 - [prototyped for invasion biology and Wikidata](#)
 - but perhaps better turn it into an extension?
 - [exposed via BASE](#)
 - reuse by [Open Knowledge Maps](#) is being [worked on](#)

Use cases at the intersections (2)

- Data import
 - Work with [Wikibase data in various formats](#)
 - Wikibase dumps / generic RDF data / tabular data
- Throw-away Wikibases for
 - Data modeling & cleanup & integration
 - Education
- Federation
 - items / properties / lexemes / queries etc.
- learning about / exploring
 - Wikibase / MediaWiki / SPARQL
 - triple stores / FAIR data / open data
 - content
 - communities



Intermediate results of import from a [CSV source](#) about invasive species, as seen [via Scholia](#). [CC0](#).



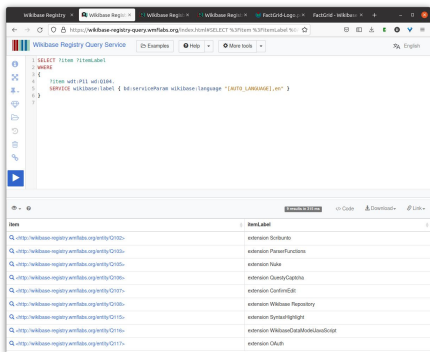
[Community curation of a set of 10k taxonomic papers](#)
(Rod Page, [CC BY 4.0](#))

Use cases at the intersections (3)

- Using Scholia to profile
 - [Wikibase as a software](#)
 - [Wikibase as a topic](#)
 - [Usage of Wikibase](#)
- Running Wikibase frontends or tools like Scholia on
 - a standalone Wikibase
 - a Wikibase farm
- Community curation
 - quality control workflows
- Visualizations beyond WDQS defaults?
- Documentation
- Internationalization
- Wikibase equivalent(s) of [InstantCommons](#)?

Use cases at the intersections (4)

- Spam protection / non-public content
- Wikibase instances to serve the Wikibase community
 - index of
 - instances (Wikibase registry) and their specs
 - Wikibase dumps / tools / extensions/ events?
 - reference instances for tech specs, e.g. triple stores?
 - SPARQL queries?
 - some basic default queries for a fresh Wikibase?
 - see also [Documented queries](#)
 - OpenRefine manifests / ... ?
 - benchmark data / setups / queries
- Usage stats
- [Standardization of property proposals](#)



[List of Wikibase extensions](#) known to the Wikibase Registry as of today. [CCO](#).

Thank you!

Questions / comments?

Use the meeting's pad or

[@EvoMRI](#) or daniel.mietchen@wikipedia.de
