

Wikidata: Mismatch Finder

March 2023

Alan Ang
Partner Manager
alan.ang@wikimedia.de



WIKIMEDIA
DEUTSCHLAND

Agenda

- About Wikimedia Deutschland
- Our Linked Open Data vision
- About Wikidata
- Data Reuse
- Wikibase
- Wikibase and GLAM



Wikimedia Deutschland

We are a **non-profit** organisation based in **Berlin**. The **German chapter** of the global **Wikimedia Movement**.

We work to provide more **people** with more **access** to more **knowledge**.



Software development

We have a dedicated software development department. Our role is to ensure that Wikimedia Deutschland **gives everyone free access to more knowledge** by delivering innovative and useful **open source software**.

We develop and maintain **Wikibase**. We also develop and maintain **Wikidata**, which is run on the servers of the Wikimedia Foundation.



Software development

- The WMDE software teams build products for a large international community of users
- To ensure we build the right products, these users must play a central role in our product development processes
- We adopt a **community-centered development** approach
- In practice this means that our community of users inform product development and we work to ensure transparency around our work

Our Linked Open Data vision

“We work towards a future in which people share the power to collect and organize the data that shapes humanity's understanding of the world.

Diverse communities around the world participate in Wikidata and in a network of specialized Wikibases, co-creating an open and free global knowledge graph in a thriving Linked Open Data web.

This is the basis upon which people, companies and institutions of all sizes can generate new insights, build new apps and services, and change the world for the better.”

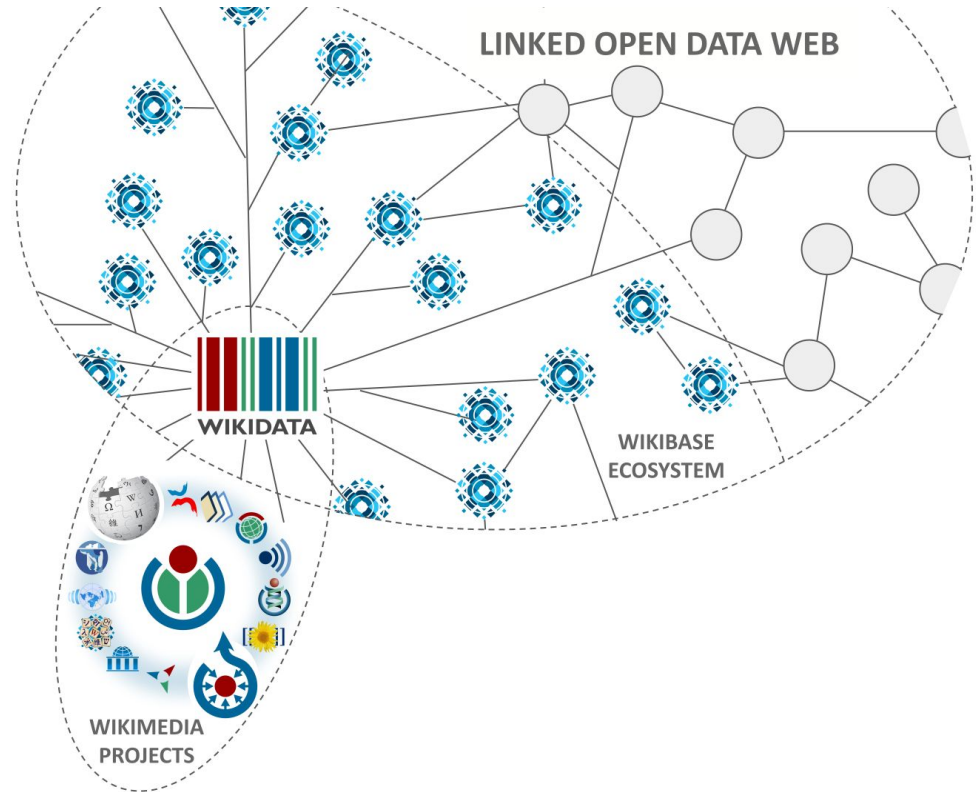


Source: [Wikimedia Linked Open Data Strategy 2021](#)

Wikidata and the Wikibase ecosystem are the foundation of our Linked Open Data work

Wikidata provides data and an ontology that others can use, and build upon.

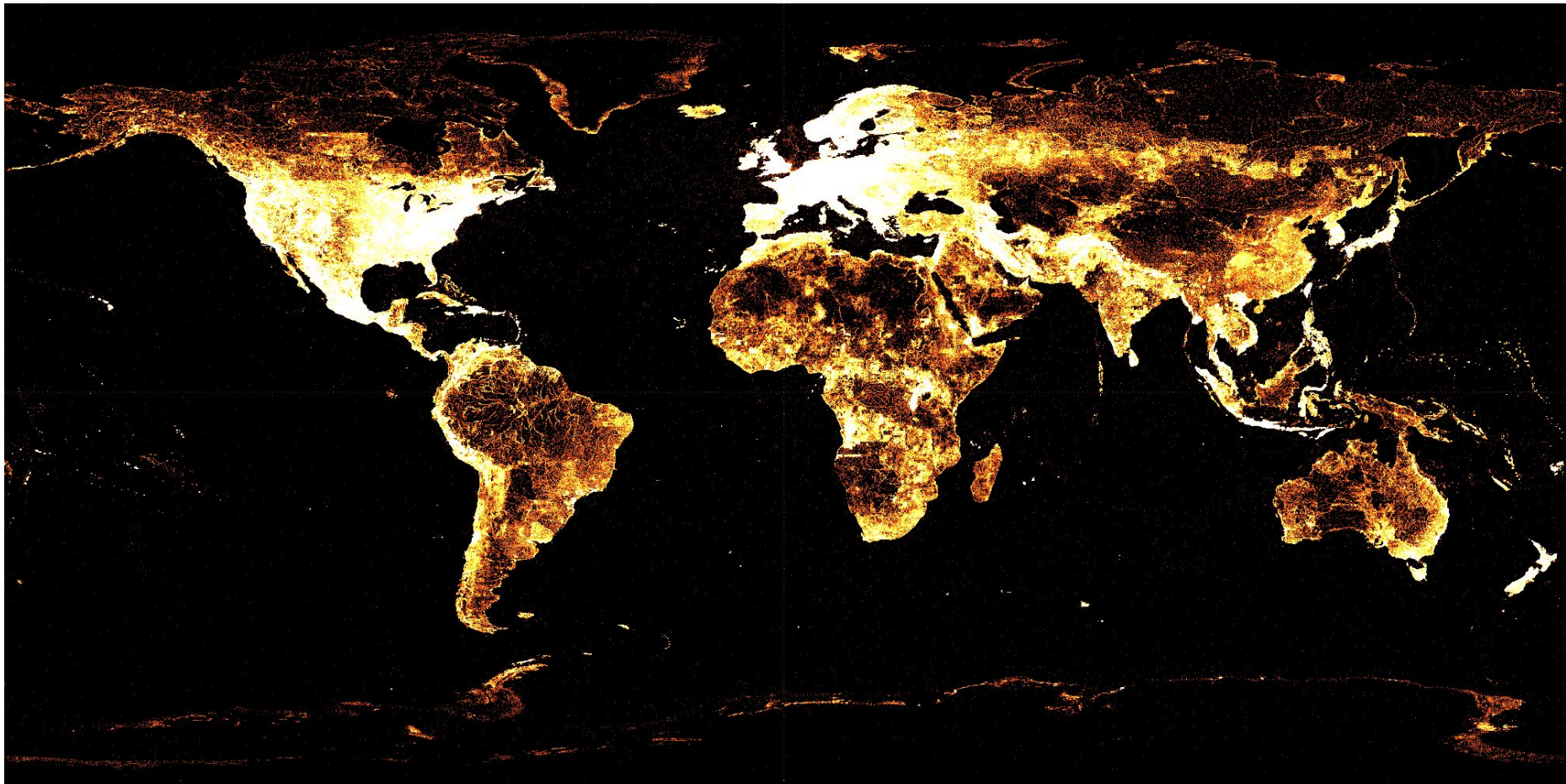
The Wikibase products expand the knowledge available through Linked Open Data



Wikidata



WIKIMEDIA
DEUTSCHLAND



100M data items

Wikidata: the basics

- **Wikimedia project** started in 2012
- **Free** and **open** knowledge base
- Contains **structured data**
- **Linked** to other databases
- Data available under **CC0 licence**
- Based on **statements** and **references**
- Made for **humans** and **machines**
- **Multilingual**
- **Collaborative**



How is data modeled?

Users model the world in terms of “**triples**”. Triple structure: **Item, Property, Value**

Maya Angelou (Q19526) — **Item identifier (Q item)**

American poet and author (1928-2014)

 edit

Marguerite Annie Johnson | Marguerite Johnson | Marguerite Ann Johnson | Marguerite Anne Johnson

[In more languages](#)

Statements

Property — instance of

 human — **Value**

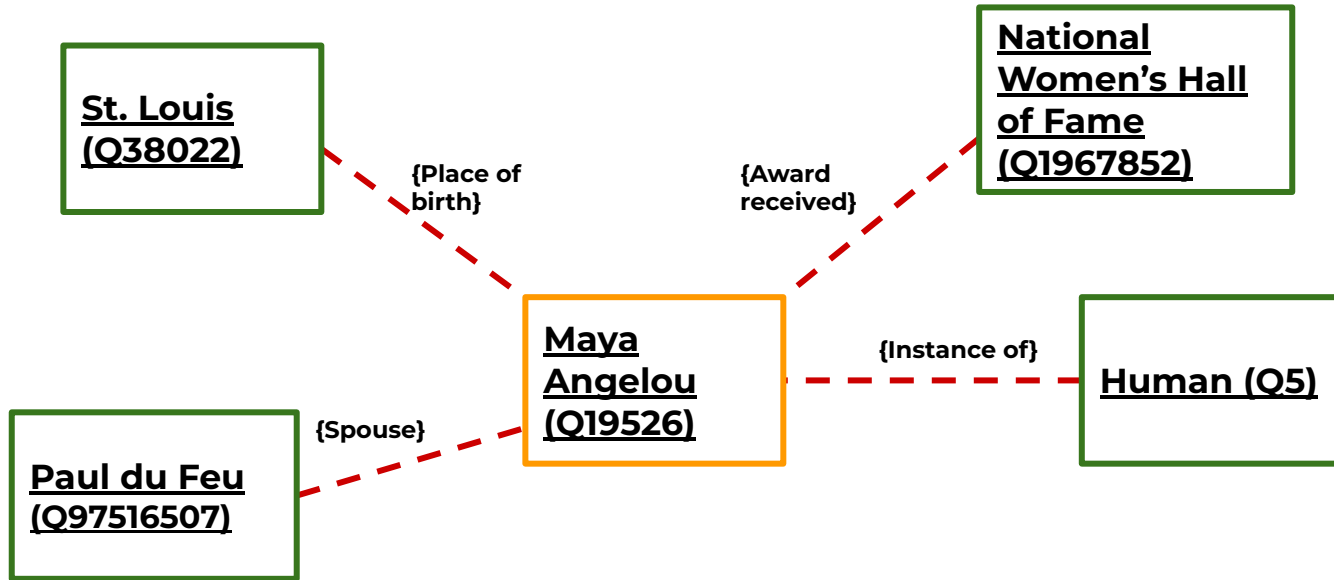
 edit

[1 reference](#)

[+ add value](#)

How is data modeled?

Linking data: items are linked to each other



Querying Wikidata



Wikidata Query Service

Examples

Help

More tools

Wikidata Query Builder

The Wikidata Query Builder provides a visual interface for building a simple Wikidata query. It is ideal for users with little or no experience in [SPARQL](#), the powerful query language. The Query Builder doesn't offer SPARQL's full functionality, but you can always open your query in the Query Service, where you can view, edit or expand it via the link above the results. [Feedback is welcome here](#).

Query

Find all items...

With Without Property matching Value References

Add condition

Settings

Limit the number of results to
 Show IDs instead of labels (may prevent timeout)

Run query Get a shareable link

[Show query in the Query Service](#)

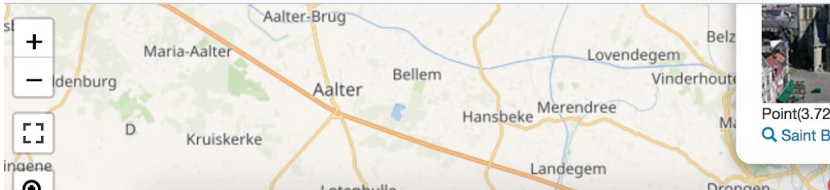
Results

Results will be displayed here

```
1 #defaultView:Map
2 SELECT ?item ?itemLabel ?coord ?image
3 WHERE
4 {
5   ?item wdt:P708 wd:Q869096.
6   ?item wdt:P625 ?coord.
7   OPTIONAL {
8     ?item wdt:P18 ?image
9   }
10  SERVICE wikibase:label { bd:serviceParam wikibase:language "en".
11 }
```



Map



About the Query Builder

Licensed under [BSD 3-Clause License](#)
Last build at [Thu, 19 Aug 2021 08:44:30 GMT](#)
[View Source](#)
[Report a bug](#)

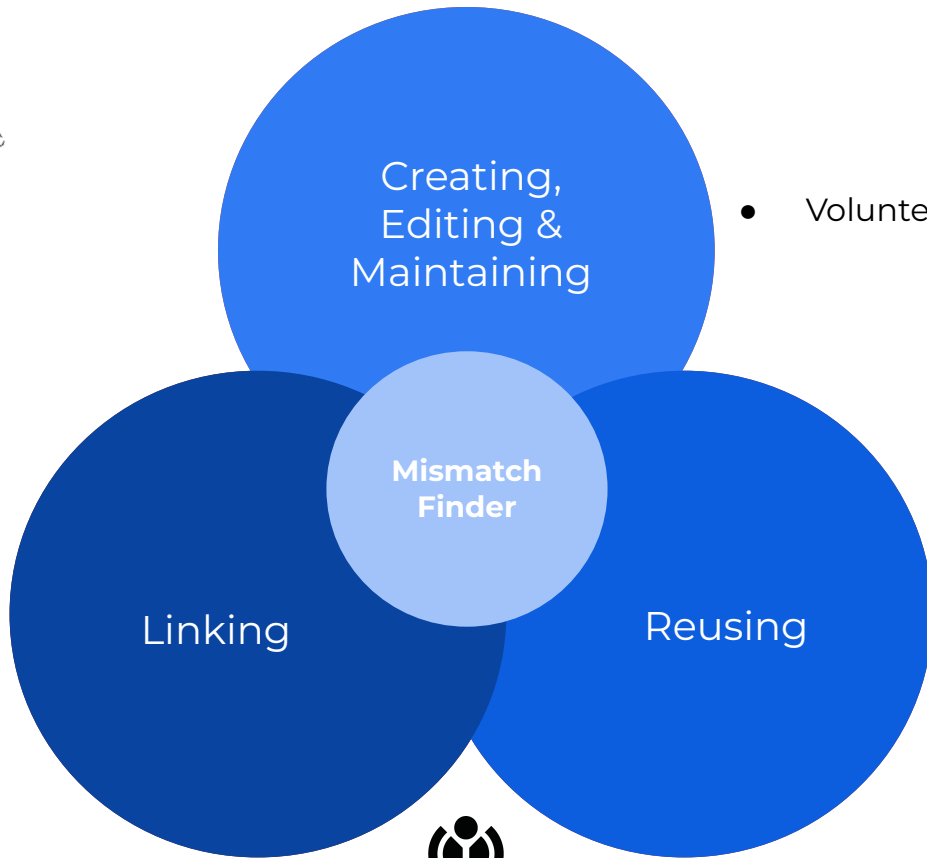
About us

[Privacy Policy](#)
[Wikimedia Deutschland](#)
[Made by the Wikidata team](#)

Reusing Wikidata's Data



WIKIMEDIA
DEUTSCHLAND



- Volunteers

- Galleries, Libraries, Archives, & Museums
- Open government projects
- Media organisations

Organisations

- Commercial: large tech, AI companies
- Non-profit: OSM, Government agencies, Wikimedia projects

Projects

- Govdirectory, Histropedia

Mismatch Finder

[Mismatch Finder](#) is a tool, developed by Wikimedia Deutschland, that helps editors work on **mismatches** between Wikidata's data and other databases/websites.

Process Flow

1. Identify discrepancies between your data and Wikidata's statements
2. Generate a mismatch CSV file with the mismatching statements
3. Add your mismatches to the [Mismatch Finder store](#) via API
4. Editors review your mismatches and fix statements in Wikidata/ report issues to your data to you

[Video](#)

 WIKIDATA MISMATCH FINDER



Wikibase



WIKIMEDIA
DEUTSCHLAND

Wikibase is the software underlying Wikidata

It is a **free** and **open source** software suite launched in 2012 by **Wikimedia Deutschland**. We continue to develop and maintain it today.

Wikibase powers **Wikidata** and increasingly, a wide range of other **linked open data projects**.



Suite and Cloud overview



	Wikibase Suite (manual install/ via Docker images)	Wikibase.cloud (SaaS)
Hosting	Self managed	Hosted by WMDE (Europe, GDPR-compliant)
Technical experience (installation)	Intermediate to Expert	Novice
Number of Wikibases	Unlimited	6 (per Wikibase.cloud user account)
Customisation	High	Generic

Wikibase & GLAM

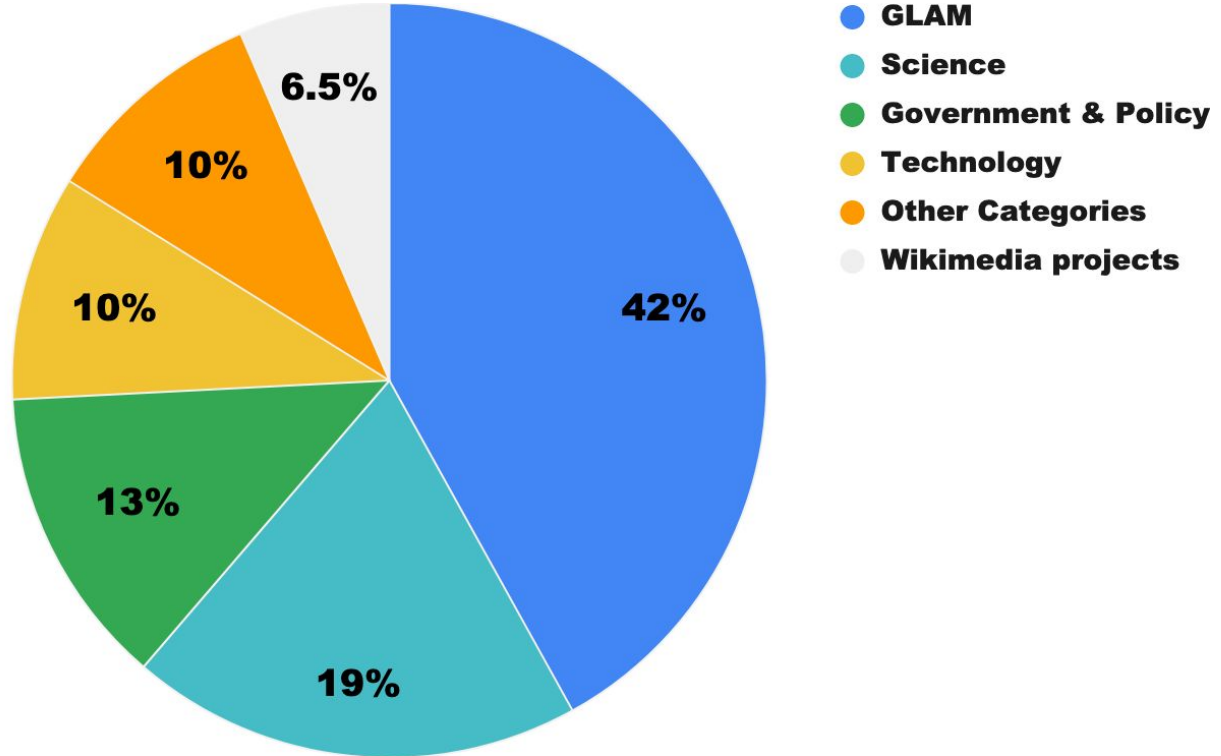


WIKIMEDIA
DEUTSCHLAND

Wikibases at a glance

The active Wikibases instances that we currently track span a number of disciplines.

This pie chart represents the % of Wikibase instances in each category, not the % of data they contribute to the decentralised knowledge graph.



*GLAM stands for galleries, libraries, archives, and museums

Wikibase in the library context

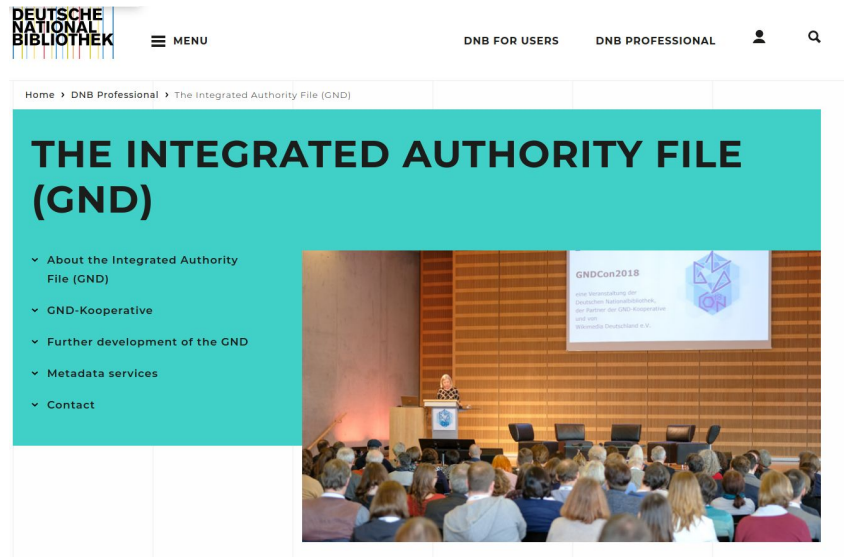
Wikibase is being evaluated by libraries as a tool to help them **store and manage** their structured data, as well as **connect to the world of linked open data**.

- German National Library (DNB)
- French National Library
- Finnish National Library
- National Library of the Czech Republic
- National Library of Luxembourg
- The Smithsonian Libraries



German National Library

- The German National Library is piloting Wikibase for its **integrated authority file**
- The ***Gemeinsame Normdatei*** represents and describes entities then used in cataloguing. It is authoritative **across German-speaking** countries
- The idea is to enable **collaborative maintenance** of the authority file
- See Wikibase instance [here](#).
- DNB talks Wikibase [video](#).



The screenshot shows the website of the Deutsche Nationalbibliothek (German National Library). The header includes the library's name, a menu icon, and navigation links for 'DNB FOR USERS' and 'DNB PROFESSIONAL'. The main content area features a teal banner with the title 'THE INTEGRATED AUTHORITY FILE (GND)'. Below the banner is a navigation menu with links: 'About the Integrated Authority File (GND)', 'GND-Kooperative', 'Further development of the GND', 'Metadata services', and 'Contact'. To the right of the menu is a photograph of a conference room with a speaker at a podium and an audience. A presentation slide in the background reads 'GNDCon2018' and '100 Jahre Wertschöpfung der Deutschen Nationalbibliothek'.



French National Library

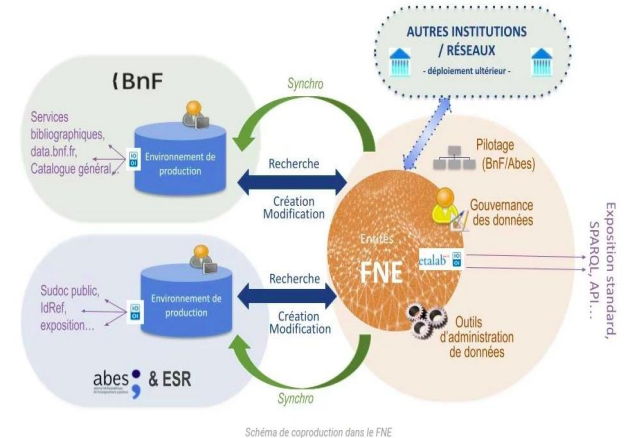
- The **FNE** is a project launched in 2017 and jointly steered by Abes (Bibliographic Agency for Higher Education) and BnF. It aims at setting up an open data co-production platform between cultural and research public institutions.
- The **French national entities' file** will bring together in a central database all entities created by holding institutions. It will enable these institutions to create directly in this shared database.

Le projet Fichier national d'entités (FNE) a pour objet la mise en place d'une plate-forme de production mutualisée d'entités, données d'autorité structurées selon le modèle IFLA LRM. Le projet a été lancé en 2017 sous l'impulsion de l'Abes et de la BnF qui en assurent le co-pilotage. Il est rattaché au programme national Transition bibliographique.

Un périmètre élargi pour des services mutualisés

Le FNE est appelé à devenir un outil majeur non seulement dans le paysage documentaire et bibliographique français, mais aussi à l'échelle du web de données. Ses services s'adressent aux bibliothèques de lecture publique et centres de documentation, aux archives et à toutes institutions culturelles françaises, voire francophones.

Le FNE est un projet stratégique pour accompagner la mise en œuvre de la Transition bibliographique à l'échelle nationale.



TIB

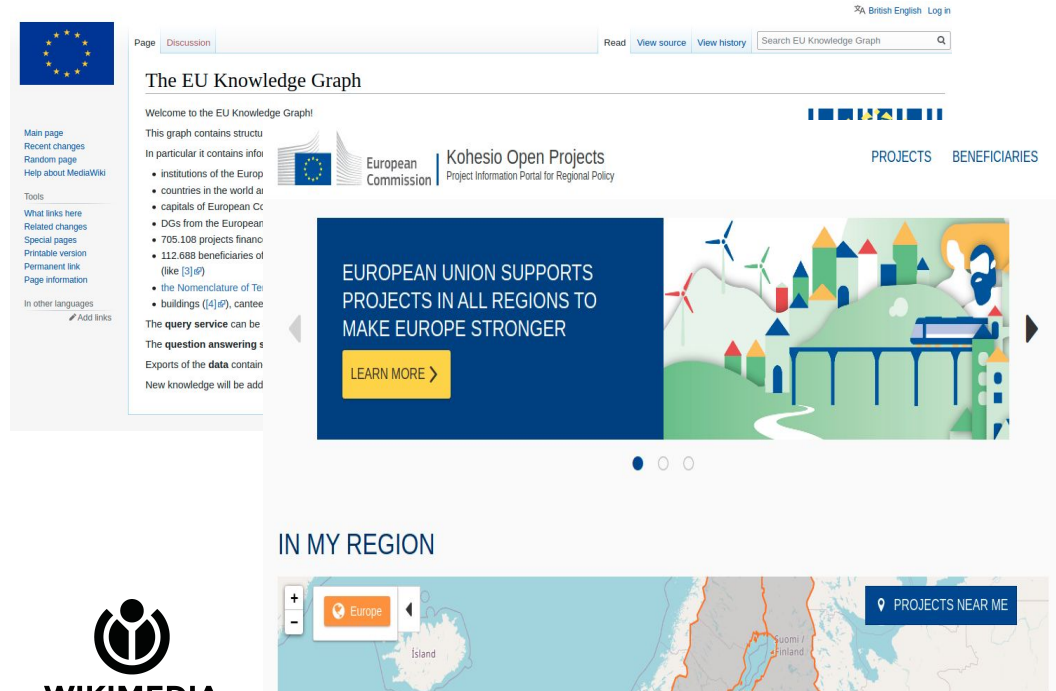
- Technical Information Library (Hannover) – [Leibniz Information Centre for Science and Technology and University Library](#) provides science, research, industry and business with literature and information - in both printed and electronic form.
- Also known as the **German National Library of Science and Technology** as well as Architecture, Chemistry, Computer Science, Mathematics and Physics.
- The library is committed to **openness, open access** to information, publications and scientific data: it aims to make information and knowledge more easily accessible, more transparent and better reusable.
- See [here](#) for TIB's video on Commons.

Ministry of Culture, Luxembourg

- **Shared authority file** on persons
- **8 GLAM institutions** joining forces in a competency network on digital cultural heritage topics
 - Sharing of authority information as such, and as shared access points to holdings.
 - Using CIDOC-CRM ontology as a federation tool for data modelling and interfaces, and RDA-FR as normative rulebook for content.
- [Wikibase EDTF extension](#): adds support for the Extended Date/Time Format (EDTF) Specification to Wikibase via a new data type. See video [here](#).
- [Automated Values extension](#): that allows defining rules to automatically set labels or aliases based on Statement values. See video [here](#).
- See video [here](#) and deck [here](#).

EU Commission

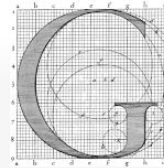
- The [EU Knowledge Graph](#) is a Wikibase instance managed and maintained by a team at the EU Commission
- It contains **general purpose data** about the European Union and its institutions, as well as data about **EU funding and its beneficiaries**
- It acts as a **data repository** for the [Kohesio](#) website .
- Video [here](#).



The screenshot shows the 'The EU Knowledge Graph' page. At the top, there's a navigation bar with 'Page', 'Discussion', 'Read', 'View source', 'View history', and a search box for 'Search EU Knowledge Graph'. The main heading is 'The EU Knowledge Graph'. Below it, a welcome message states 'Welcome to the EU Knowledge Graph!'. A sidebar on the left contains navigation links like 'Main page', 'Recent changes', 'Random page', 'Help about MediaWiki', 'Tools', 'What links here', 'Related changes', 'Special pages', 'Printable version', 'Permanent link', 'Page information', and 'In other languages'. The main content area features the European Commission logo and 'Kohesio Open Projects' with the subtitle 'Project Information Portal for Regional Policy'. A large blue banner reads 'EUROPEAN UNION SUPPORTS PROJECTS IN ALL REGIONS TO MAKE EUROPE STRONGER' with a 'LEARN MORE' button. Below this is a map section titled 'IN MY REGION' showing a map of Europe with a 'PROJECTS NEAR ME' button.

Factgrid

- Run by the **Gotha Research Centre** and hosted by the **University of Erfurt**
- Started out as a project to document the history of the Illuminati
- Now a **collaborative historical database** used by historians
- See Wikibase instance [here](#).



Main page
Query service
Sample queries
Advanced text search
Recent changes
Random page
Reasonator

Project Spaces

Blog
Projects
Troubleshooting
Data modeling
SPARQL Lab

Work

New Item
Mergelteme
Batch fragments
QuickStatements
Directory of Properties
New Property
Help

Tools

What links here
Related changes
Special pages
Printable version
Permanent link
Page information

Main page Discussion

Main Page



Welcome!

on the FactGrid database, a project of the Gotha Research Centre Germany operated by the data lab of the University of Erfurt are using a MediaWiki with Wikidata's "wikibase" extension. Our main product are data which we collect on "items" such as the

- [Adam Weishaupt](#)
- [Paris](#)

You will need to be logged in to see the "add statement" link with which you can add information in the form of triple based data explored with a data mining language, SPARQL, at our [SPARQL Endpoint](#). All FactGrid data are [CC0-licensed](#). You can do the aim to explore FactGrid data in other software environments or visualise searches with various tools on our site.

Visit our

- [FactGrid FAQ](#) for more information on why you might love to use FactGrid as your research platform
- [Help Section](#) for further assistance
- [Terms of Service](#) if you want to have an account on FactGrid
- [Project Section](#) to take a look at work done on the collective platform

The FactGrid database is an open collaborative and multilingual project provided by — as of August 2020 — [140 participants](#) and the database might be a cool candidate. For more information and account details contact:

Dr. Olaf Simons
Gotha Research Centre at the University of Erfurt
Am Schloßberg 2
99867 Gotha
office: +49-361-737-1722
mobile: +49-179-5196880
olaf.simons@pierre-marteau.com

Enslaved.org

- Enslaved is a linked open database about the **history of transatlantic slave trade**
- Using Wikibase as a **linking and reconciliation hub**, Enslaved unites seven separate scholarly databases
- It contains more than **950 thousand* records** about (People, Events, Places and sources)
- See Wikibase instance [here](#).

Enslaved | Peoples of the Historical Slave Trade

Explore or reconstruct the lives of individuals who were enslaved, owned slaves, or participated in the historical trade. [Learn More](#)

Start a search across 613458 records from the historical slave trade [Go to Advanced Search](#)

Search People, Events, and Places

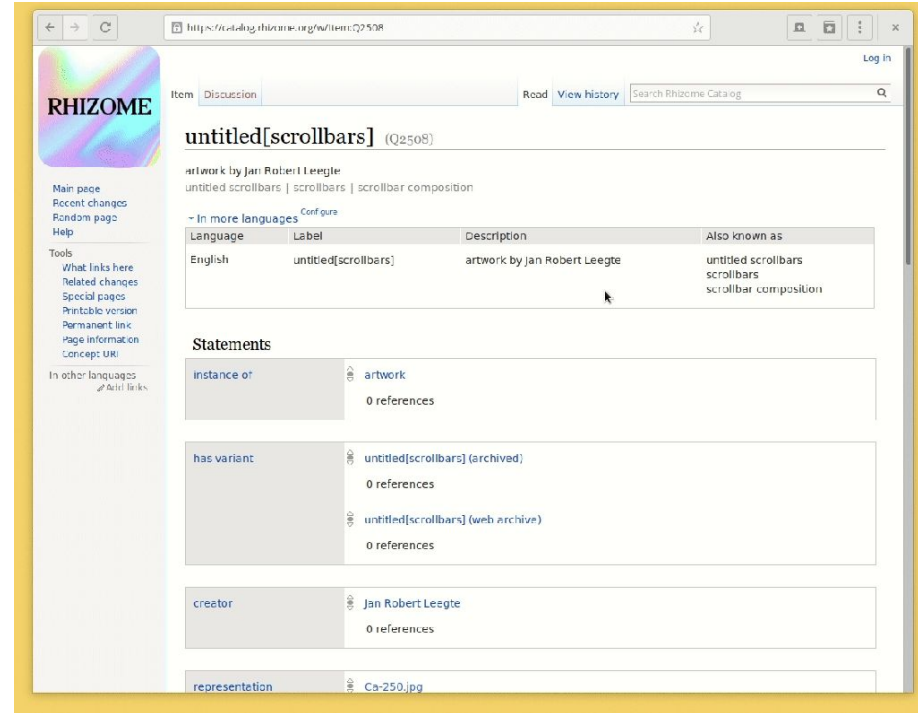
Explore

Browse People, Events, Places, and Sources

👤 People	384380	>
📅 Events	212537	>
📍 Places	14033	>
📄 Sources	2508	>

Rhizome

- Rhizome, an art organization in New York City, was one of the **early adopters** of Wikibase
- Started using Wikibase in 2015 for its **archive of born-digital art** and **digital preservation activities**
- The **flexible** model of Wikibase captures the constantly changing nature of digital art



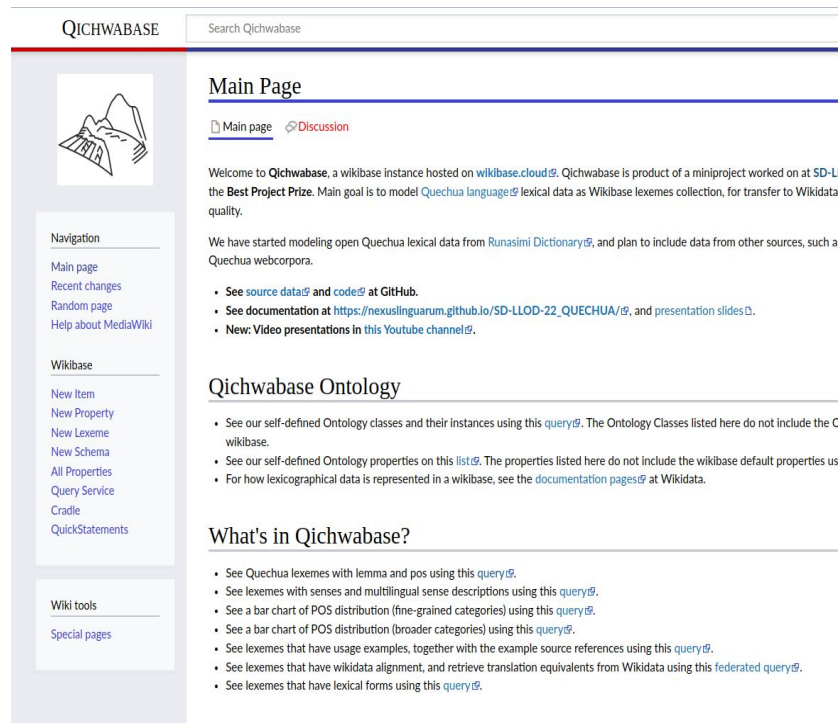
The screenshot shows a Wikibase item page for 'untitled[scrollbars]' (Q2508). The page is titled 'untitled[scrollbars] (Q2508)' and includes a description: 'artwork by Jan Robert Leegte' and 'untitled scrollbars | scrollbar composition'. A table lists the item in multiple languages, with English being the primary language. The 'Statements' section shows various relationships, including 'instance of', 'has variant', 'creator', and 'representation'.

Language	Label	Description	Also known as
English	untitled[scrollbars]	artwork by Jan Robert Leegte	untitled scrollbars scrollbars scrollbar composition

Property	Value	References
instance of	artwork	0 references
has variant	untitled[scrollbars] (archived)	0 references
	untitled[scrollbars] (web archive)	0 references
creator	Jan Robert Leegte	0 references
representation	Ca-250.jpg	

Qichwabase

- **Qichwabase**, a wikibase instance hosted on **wikibase.cloud**, is product of a miniproject worked on at the 4th Summer Datathon on Linguistic Linked Open Data (**SD-LLOD-22**) in June 2022, where it was awarded the **Best Project Prize**.
- Main goal is to model **Quechua language** lexical data as Wikibase lexemes collection, for transfer to Wikidata, as soon the dataset reaches the envisaged quality.
- See video [here](#).



The screenshot shows the main page of the Qichwabase Wikibase instance. At the top, there is a search bar and the title 'QICHWABASE'. Below the search bar is a logo featuring a mountain range and the word 'QICHWA'. The page is titled 'Main Page' and includes links for 'Main page' and 'Discussion'. The main content area contains a welcome message, a description of the project, and a list of links to source data, documentation, and video presentations. There are also sections for 'Qichwabase Ontology' and 'What's in Qichwabase?' with lists of queries and descriptions.

QICHWABASE Search Qichwabase

Main Page

[Main page](#) [Discussion](#)

Welcome to **Qichwabase**, a wikibase instance hosted on [wikibase.cloud](#). Qichwabase is product of a miniproject worked on at SD-L the **Best Project Prize**. Main goal is to model **Quechua language** lexical data as Wikibase lexemes collection, for transfer to Wikidata quality.

We have started modeling open Quechua lexical data from [Runasimi Dictionary](#), and plan to include data from other sources, such as Quechua webcorpora.

- See [source data](#) and [code](#) at [GitHub](#).
- See [documentation](#) at https://nexuslinguaram.github.io/SD-LLOD-22_QUECHUA/, and [presentation slides](#).
- **New: Video presentations in this Youtube channel**.

Qichwabase Ontology

- See our self-defined Ontology classes and their instances using this [query](#). The Ontology Classes listed here do not include the Q wikibase.
- See our self-defined Ontology properties on this [list](#). The properties listed here do not include the wikibase default properties us
- For how lexicographical data is represented in a wikibase, see the [documentation pages](#) at Wikidata.

What's in Qichwabase?

- See Quechua lexemes with lemma and pos using this [query](#).
- See lexemes with senses and multilingual sense descriptions using this [query](#).
- See a bar chart of POS distribution (fine-grained categories) using this [query](#).
- See a bar chart of POS distribution (broader categories) using this [query](#).
- See lexemes that have usage examples, together with the example source references using this [query](#).
- See lexemes that have wikidata alignment, and retrieve translation equivalents from Wikidata using this [federated query](#).
- See lexemes that have lexical forms using this [query](#).

Discussion



WIKIMEDIA
DEUTSCHLAND

Dedicated community spaces

The Wikibase user community **continues to grow**. Whether to remain informed about our continued work with **Wikibase.cloud** or to **engage with other community members**, there are a number of avenues to connect with the community:

- [Wikibase Community User Group](#)
- Mailing lists
 - [General user group](#)
 - [Wikibase.cloud](#)
- [Telegram channel](#)
- [Monthly Live Sessions](#)
- [Wikibase stakeholder group](#)



Joining the Wikibase community

Further support resources are available on mediawiki.org, including detailed **documentation** on how to set up Wikibase using the **docker images** or **manually**. You can also find a useful **FAQ**.

If you're looking for **technical support**, we maintain a space on [meta](#) where third party support providers can list their services.





Let's stay in touch!

Reach out to us!

swe_partnerships@wikimedia.de



WIKIMEDIA

DEUTSCHLAND