



WIKIDATA

giving more people more access to
more knowledge

Lydia Pintscher
Wikidata Portfolio Lead, Wikimedia Deutschland
lydia.pintscher@wikimedia.de - @nightrose
Knowledge Graph Conference, May 2023



What is Wikidata?

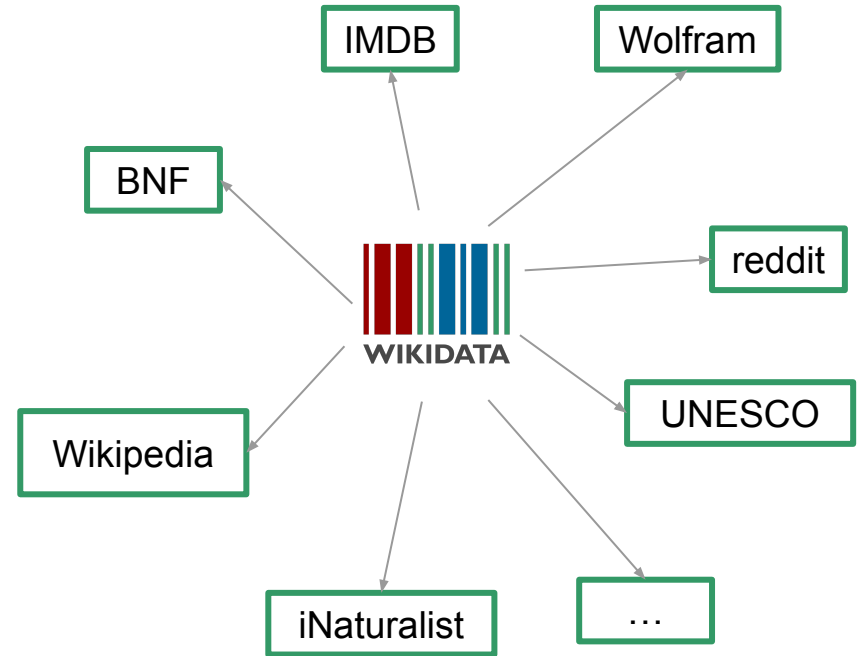


- Wikimedia project started in 2012
- Free and open knowledgebase
- Contains linked data and is linked to a lot of other databases, catalogs, etc.
- Data available under CC0
- Made for humans and machines
- Multilingual
- Collaborative

The screenshot shows the Wikidata main page. At the top left is the Wikidata logo. Below it is a navigation menu with links like 'Main page', 'Community portal', 'Project chat', 'Create a new item', 'Recent changes', 'Random item', 'Query Service', 'Nearby', 'Help', and 'Donate'. There are also sections for 'Lexicographical data' and 'Tools'. The main content area features a large network diagram with nodes and edges, overlaid with a central grey box that says 'Welcome to Wikidata' and 'the free knowledge base with 94,975,076 data items that anyone can edit.' Below this are three colored boxes: 'Welcome!' (green), 'Learn about data' (blue), and 'Get involved' (red). The 'Learn about data' box includes three images: Earth, a map, and Mount Everest, with labels for 'item: Earth (Q2)', 'Property: highest point (P612)', and 'custom value: Mount Everest (Q512)'. At the bottom, there is a link to 'Learn about Wikidata'.

What makes Wikidata special?

- You can be a part of it
- More nuanced modeling of the world and focusing on verifiability
- Multilingual
- Loosely enforced ontology
- Highly connected internally and to other databases, catalogs, etc. to open up a ton of additional data
- Closely connected to Wikipedia and the other Wikimedia Projects



Maya Angelou

(Q19526)

Item identifier (Q ID)

American poet, author, and civil rights activist (1928–2014)

 edit

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

▼ In more languages

[Configure](#)

Language	Label	Description	Also known as
English	Maya Angelou	American poet, author, and civil rights activist (1928–2014)	Marguerite Annie Johnson Marguerite Johnson Marguerite Ann Johnson Marguerite Anne Johnson
German	Maya Angelou	US-amerikanische Schriftstellerin, Professorin und Menschenrechtlerin	Marguerite Johnson Marguerite Annie Johnson
French	Maya Angelou	mémorialiste, essayiste, poète et universitaire afro-américaine	Marguerite Annie Johnson
Bavarian	Maya Angelou	No description defined	

[All entered languages](#)

Labels,
Descriptions,
Aliases

Maya Angelou (Q19526)

American poet and author (1928-2014)

 edit

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

[In more languages](#)

Statements

instance of

human

Value

 edit

[1 reference](#)

[+ add value](#)

Property

Statement

award received

Property

National Women's Hall of Fame

Value

edit

point in time

1998

Qualifier

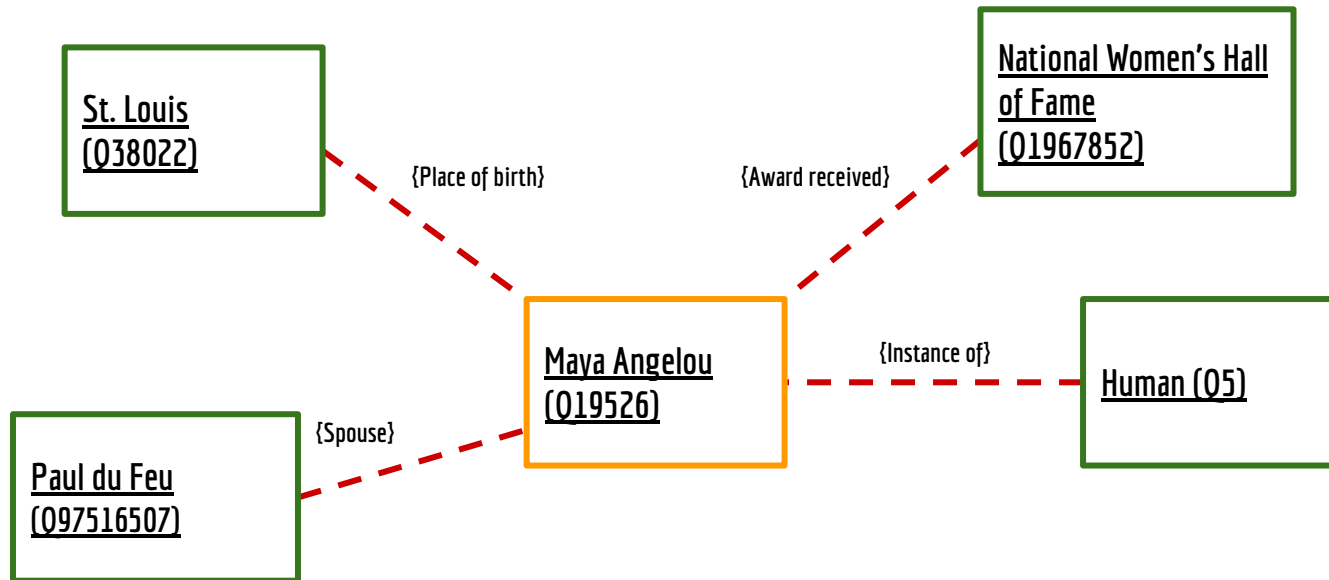
1 reference

reference URL

<https://www.womenofthehall.org/inductee/maya-angelou/>

Reference

+ add reference



102

Million

Items

Earth (Q2)...

ORES predicted quality: A (4.94)

third planet from the Sun in the Solar System

Planet Earth | the Earth | ☾ | 🌐 | World









 edit

Wikipedia (290 entries)  edit

ab	Адгъыл
ace	Bumoë
ady	Чытыгу
af	Aarde 
als	Erde
am	ጠሬት
ang	Eorðe
an	Tierra
arc	ܐܪܨܐ
ar	الأرض 
ary	أرض
arz	الأرض
ast	Tierra
as	পৃথিৱী
atj	Aski
avk	Tawava
av	Ракъ (планета)
awa	पृथ्वी
ay	Aka pacha
azb	بَر
az	Yer 
ban	Gumi
bar	Eadn
bat_smg	Žemė
ba	Ер
bcl	Kinaban
be_x_old	Зямля

[In more languages](#)

Statements

instance of	 terrestrial planet ...  edit
	0 references + add reference
inner planet of the Solar System ...	 inner planet of the Solar System ...  edit
	0 references + add reference
geographic region ...	 geographic region ...  edit
	0 references + add reference
	+ add value
part of	 Earth-Moon system ...  edit
	0 references + add reference
	+ add value

instance of (P31)

that class of which this subject is a particular example and member

 edit

is a | is an | has class | has type | is a particular | is a specific | is an individual | is a unique | is an example of | member of | unique individual of | distinct member of | unitary element of class | distinct element of | distinct individual member of | rdfs:type | type | main type | is a(n) | type of | is a type of | \in | example of

[In more languages](#)

Data type

Item

Statements

 instance of



Wikidata property

 edit

▼ 0 references

+ add reference



Wikidata property for the relationship

 edit

▼ 0 references

+ add reference

+ add value

 value hierarchy property



subclass of

 edit

▼ 0 references

+ add reference

+ add value

11k

Properties

1.45 Billion

Statements

› taxon name

› Rhincodon typus  edit

taxon author	Andrew Smith ...
year of taxon publication	1828 ...

▼ 1 reference

stated in	Integrated Taxonomic Information System
publication date	13 June 1996
retrieved	19 September 2013

+ add reference

+ add value

(L3271)

red

en

 edit

Language English

Lexical category adjective

Statements

word stem



red (English)

 edit

0 references

+ add reference

+ add value

derived from lexeme



red 

 edit

0 references

+ add reference

+ add value

Oqaasileriffik online dictionary ID



130246

 edit

0 references

+ add reference

+ add value

+ add statement

1 Million

Lexemes

12.5k

active editors





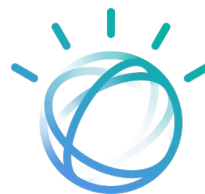
What do people and organisations do with
Wikidata's data?



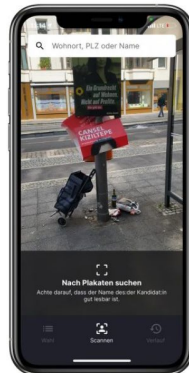
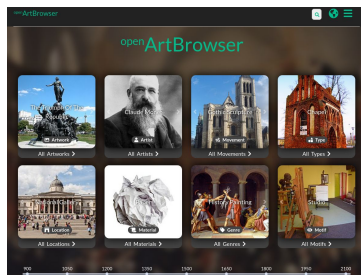
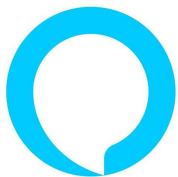


OPEN LIBRARY

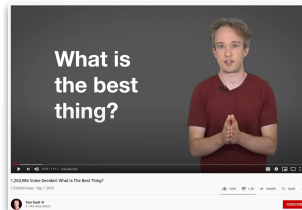
mySociety



Google



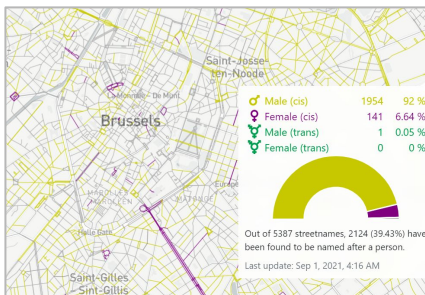
OCCRP



Quora

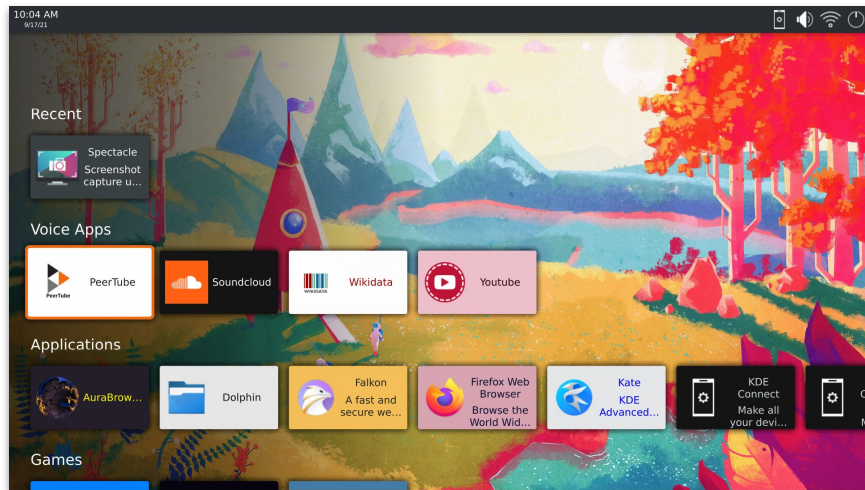


reddit



WolframAlpha™

Accessing basic information



- Use Wikidata to retrieve basic data on specific entities
- Example: MyCroft AI

Augmenting other data

Amplifying the Voices Behind Books With the Power of Data

By MEK | Published: SEPTEMBER 2, 2020

Exploring how Open Library uses author data to help readers move from imagination to impact

By [Nick Norman](#), Edited by [Mek](#) & [Drini](#)



Image Source: [Pexels / Pixabay](#) from [popsugar](#)

According to [René Descartes](#), a creative mathematician, “The reading of all good books is like a conversation with the finest [people] of past centuries.” If that’s true, then who are some of the people you’re talking to?

If you’re not sure how to answer that question, you’ll definitely appreciate the ‘Author Stats’ feature developed [by Open Library](#).



Search

[Follow @openlibrary](#)

Recent Posts

- [Book Talks: Watch Virtual Talks by Trailblazing Authors](#)
- [Reach Your 2023 Reading Goals with Open Library](#)
- [A Brand New My Books Experience](#)
- [Search Is Getting Smarter on Open Library](#)
- [2022 Review](#)

Archives

Select Month

- Use Wikidata to enrich data you already have
- Example: OpenLibrary

Training machine learning systems

Knowledge Graph based Analysis and Exploration of Historical Theatre Photographs

Tabea Tietz^{1,2}, Jörg Waitelonis³, Mehwish Alam^{1,2}, and Harald Sack^{1,2}

¹ FIZ Karlsruhe – Leibniz Institute for Information Infrastructure, Germany
firstname.lastname@fiz-karlsruhe.de

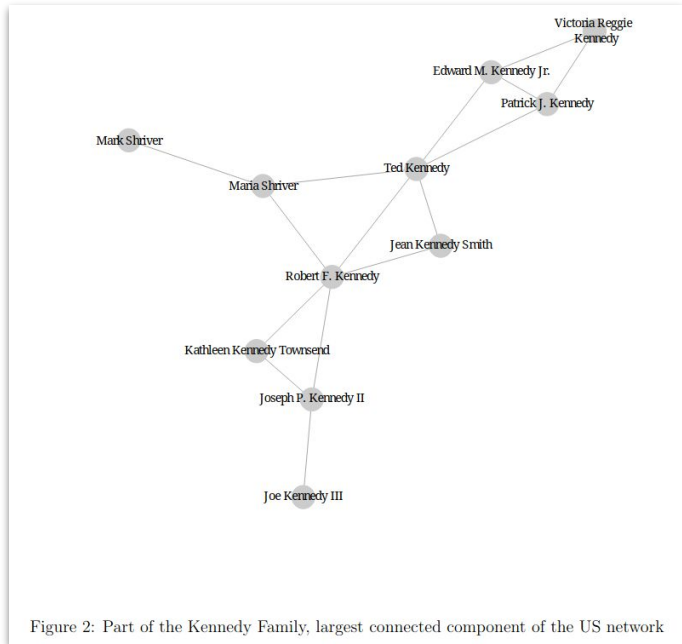
² Karlsruhe Institute of Technology, Institute AIFB, Germany

³ yovisto GmbH, Potsdam, Germany
joerg@yovisto.com

Abstract. Historical theatre collections are an important form of cultural heritage and need to be preserved and made accessible to users. Often however, the metadata available for a historical collection are too sparse to create meaningful exploration tools. On the use case of a historical theatre photograph collection, this position paper discusses means of automated recognition of historical images to enhance the variety and depth of the metadata associated to the collection. Moreover, it describes how the results obtained by image recognition can be integrated into an existing Knowledge Graph (KG) and how these generated structured image metadata can support data exploration and automated querying to support human users. The goal of the paper is to explore cultural heritage data curation techniques based on deep learning and KGs to make the data findable, accessible, interoperable and reusable in accordance with the F.A.I.R principles.

- Use Wikidata as a source of training data for machine-learning systems
- Example: Exploration of historical theatre photographs

Exploring and visualizing data



- Use Wikidata's data to give new insights and overviews in areas such as journalism, education and research
- Example: Measuring political elite networks by Omer Yalcin, OpenArtBrowser

openArtBrowser



Livro Dos Heróis E Heroínas Da Pátria

Artwork

All Artworks >



Thomas Rowlandson

Artist


All Artists >



Early Netherlandish Painting

Movement


All Movements >



Chapel-shrine

Type


All Types >



National Gallery Of Art

Location

All Locations >



Wood

Material

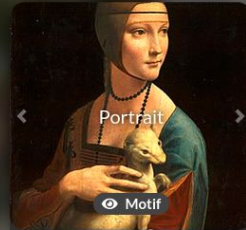
All Materials >



Animal Art

Genre

All Genres >



Portrait

Motif

All Motifs >



Gateway into the LOD web



SIDESTEPPING THE LIMITATIONS OF COLLECTION CATALOGUES WITH MACHINE LEARNING AND WIKIDATA

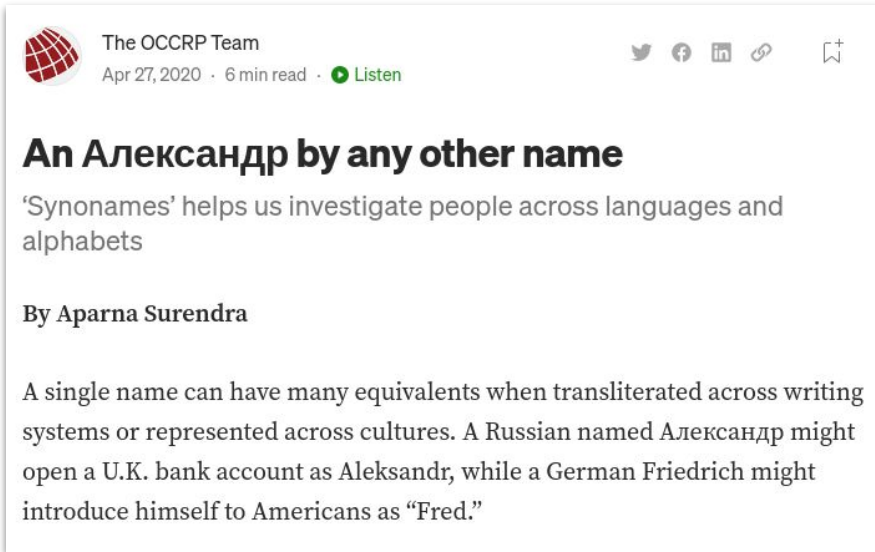
23 September 2020

Rhiannon Lewis and John Stack

The [Heritage Connector project](#) seeks to understand how existing digital tools and methods can be used to build relationships at scale between inconsistently, and at times thinly catalogued, digitised collection objects. Online collections have been with us for around twenty years now, and their digitisation has enabled access to databases with a wealth of collections knowledge. However, these databases have determined, and limited, how this collection knowledge was structured and accessed. Machine learning presents an opportunity to build links at scale through knowledge graphs between Wikidata and museum collections, so that we can begin to acknowledge and overcome these limitations.

- Use Wikidata's links to other websites, catalogs, archives and more to access additional information
 - Example: The Science Museum
-

Source of notable entities for disambiguation, cataloging, tagging etc



The screenshot shows a Medium article snippet. At the top left is the profile picture of 'The OCCRP Team', a globe icon. To its right is the text 'The OCCRP Team' and 'Apr 27, 2020 · 6 min read · Listen'. Further right are social media icons for Twitter, Facebook, LinkedIn, and a link icon, followed by a bookmark icon. The main title of the article is 'An Александр by any other name' in bold. Below the title is the subtitle '‘Synonames’ helps us investigate people across languages and alphabets'. The author's name 'By Aparna Surendra' is listed below the subtitle. The main body of text starts with 'A single name can have many equivalents when transliterated across writing systems or represented across cultures. A Russian named Александр might open a U.K. bank account as Aleksandr, while a German Friedrich might introduce himself to Americans as “Fred.”'

The OCCRP Team
Apr 27, 2020 · 6 min read · Listen

An Александр by any other name

‘Synonames’ helps us investigate people across languages and alphabets

By Aparna Surendra

A single name can have many equivalents when transliterated across writing systems or represented across cultures. A Russian named Александр might open a U.K. bank account as Aleksandr, while a German Friedrich might introduce himself to Americans as “Fred.”

- Use Wikidata’s stable identifiers to clearly identify concepts in a language-independent manner
- Example: OCCRP

Internationalisation

- Use Wikidata as a source of names for various concepts across languages
 - Example: Mapbox, YLE
-



How to get to the data



There are various ways to get at that data. Depending on your needs & what you're trying to do, some ways are better than others.

- Wikidata Query Service (WDQS)
- Linked Data Fragments (LDF)
- Linked Data Interface
- Search (Elastic)
- Action API
- REST API
- Dumps
- Recent Changes stream

Network best practices

When interacting with Wikimedia servers over the internet:


- follow the [User-Agent policy](#) (send a good User-Agent header)
- follow the [Robot policy](#) (send Accept-Encoding: gzip, don't make too many requests at once, ...)
- if you get a *429 Too Many Requests* response, stop sending further requests for a while (see the Retry-After response header)

Wikidata Query Service


Wikidata Query Service Examples Query Builder Help More tools English

```
1 #Cats, with pictures
2 #defaultView:ImageGrid
3 SELECT ?item ?itemLabel ?pic
4 WHERE
5 {
6   ?item wdt:P31 wd:Q146 .
7   ?item wdt:P18 ?pic
8   SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en" }
9 }
```


Image grid - 42 results in 300 ms Code Download Link



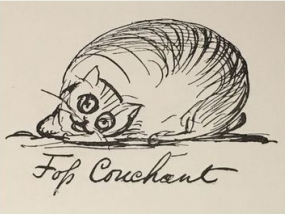
[commons:Fat cat, asleep \(319313958\).jpg](#)
Toffee



[commons:Crimean Tom.jpg](#)
Crimean Tom



[commons:Stationmaster NITAM...](#)
Nitama



[commons:Foss Couchant.jpg](#)
Foss

Wikidata Query Service

- SPARQL endpoint backed by Blazegraph
- UI: query.wikidata.org
- API: query.wikidata.org/sparql (GET and POST)

Useful to know:

- You can write federated queries with a limited number of other SPARQL endpoints
- You can set up your own instance to avoid timeouts and other limitations
- You can embed the live query result visualizations in other websites
- You can get code snippets for various programming languages in the UI

Use when:

- You don't know the specific entities you're interested in, but you know their characteristics

Don't use when:

- You're performing a text or fuzzy search
 - `FILTER(REGEX(...))` is an antipattern
- You have millions of users, each executing queries in your application
 - Consider running your own instance!
- You expect the result to be a large percentage of Wikidata's total entities

Wikidata Query Service

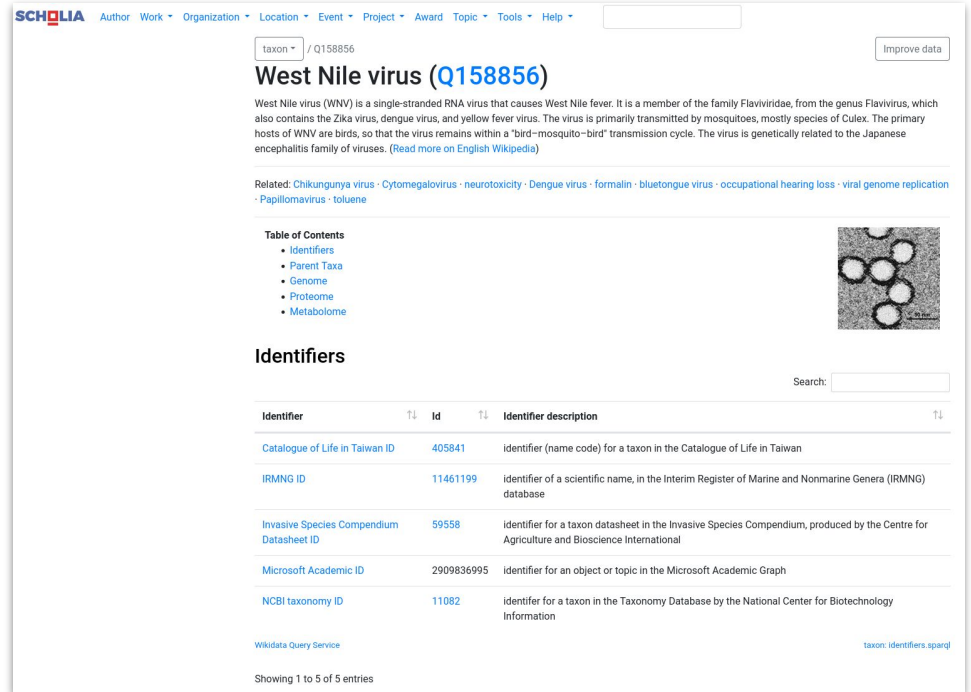
Policies and recommendations:

- Robot and user agent policies apply
- If your query times out, get help from the community to optimize it (there is a limit of 60s for query execution time)
- If you get a 429 Too Many Requests response, back off for a bit :)
- Add ?timeout=5 (seconds) to make the query time out earlier – useful in cases where fast response is required, and a late response wouldn't be usable anyway

Wikidata Query Service

Used for example by:

- Scholia
- scholia.toolforge.org
- github.com/WDScholia/scholia



SCHOLIA Author Work Organization Location Event Project Award Topic Tools Help

taxon / Q158856 Improve data

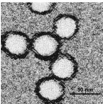
West Nile virus (Q158856)

West Nile virus (WNV) is a single-stranded RNA virus that causes West Nile fever. It is a member of the family Flaviviridae, from the genus Flavivirus, which also contains the Zika virus, dengue virus, and yellow fever virus. The virus is primarily transmitted by mosquitoes, mostly species of Culex. The primary hosts of WNV are birds, so that the virus remains within a "bird-mosquito-bird" transmission cycle. The virus is genetically related to the Japanese encephalitis family of viruses. [\(Read more on English Wikipedia\)](#)

Related: Chikungunya virus · Cytomegalovirus · neurotoxicity · Dengue virus · formalin · bluetongue virus · occupational hearing loss · viral genome replication · Papillomavirus · toluene

Table of Contents

- Identifiers
- Parent Taxa
- Genome
- Proteome
- Metabolome



Identifiers

Search:

Identifier	Id	Identifier description
Catalogue of Life in Taiwan ID	405841	identifier (name code) for a taxon in the Catalogue of Life in Taiwan
IRMNG ID	11461199	identifier of a scientific name, in the Interim Register of Marine and Nonmarine Genera (IRMNG) database
Invasive Species Compendium Datasheet ID	59558	Identifier for a taxon datasheet in the Invasive Species Compendium, produced by the Centre for Agriculture and Bioscience International
Microsoft Academic ID	2909836995	identifier for an object or topic in the Microsoft Academic Graph
NCBI taxonomy ID	11082	identifier for a taxon in the Taxonomy Database by the National Center for Biotechnology Information

Wikidata Query Service taxon: identifiers.sparql

Showing 1 to 5 of 5 entries

Linked Data Fragments

Wikidata

Wikidata

Query Wikidata by triple pattern

subject: _____

predicate: _____

object: _____

[Find matching triples](#)

Matches in Wikidata for

Showing triples 1 to 101 of ± 13,691,622,200 with 100 triples per page. [next](#)



Linked Data Fragments

- query.wikidata.org/bigdata/ldf

Useful to know:

- Computation is done on the client side, taking less resources on the server
- More experimental service with less support

Use when:

- You're looking for a list of entities based on triple patterns
- Your result set is likely to be larger
- You're okay with doing computation of result sets on your side instead of the server

Don't use when:

- You need a stable endpoint
- You need a complete result set

Linked Data Interface

▼ entities:	
▼ Q42:	
pageid:	138
ns:	0
title:	"Q42"
lastrevid:	1591415695
modified:	"2022-03-11T12:36:46Z"
type:	"item"
id:	"Q42"
▶ labels:	{...}
▶ descriptions:	{...}
▶ aliases:	{...}
▶ claims:	{...}
▶ sitelinks:	{...}

Linked Data Interface

- wikidata.org/entity/Q42 (redirects to wikidata.org/Special:EntityData/Q42)
- Available formats: .json, .rdf, .ttl, .nt or .jsonld

Useful to know:

- LDI performs content negotiation and responds in the appropriate format
- You can force a specific format by appending the file extension to the URI
- You can get a specific revision by appending `?revision=112` to the URI
- Append `?flavor=dump` for a less verbose response (not applicable for JSON)

Use when:

- You want data on a smallish set of entities, especially RDF data
- You already know the IDs of the entities you are interested in
- You want each whole entity

Don't use when:

- You don't know exactly which entities you want
 - you need to query or search first
- You want large amounts of data

Linked Data Interface

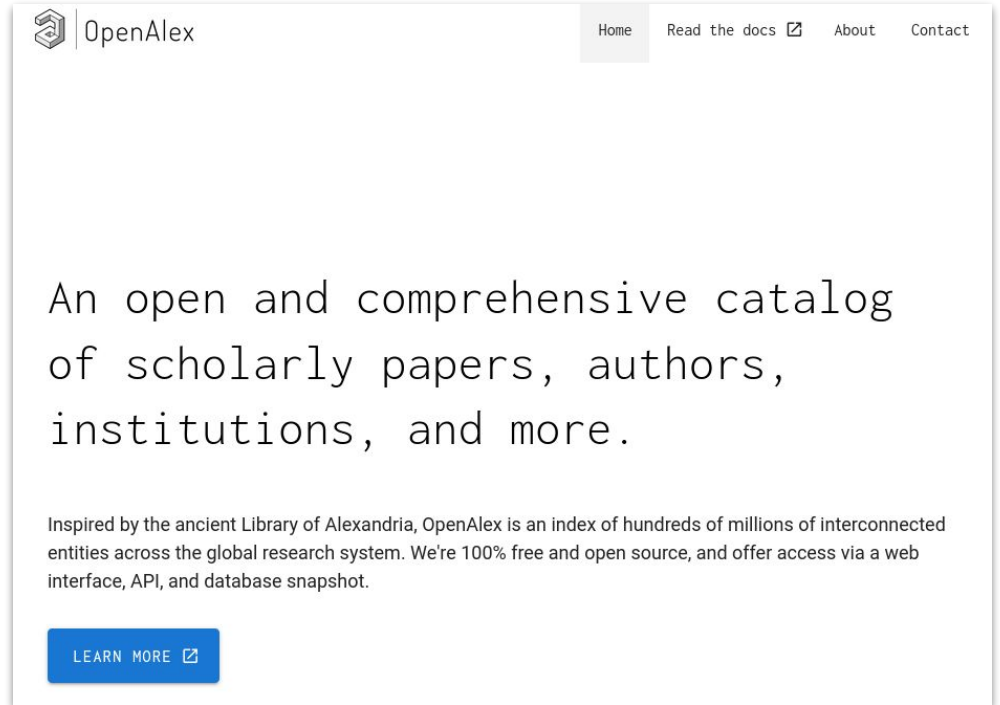
Policies and recommendations:

- Robot and user agent policies apply
- The following URLs for a specific revision and format are likely to be cached already:
 - wikidata.org/wiki/Special:EntityData/Q42.json?revision=123
 - wikidata.org/wiki/Special:EntityData/Q42.ttl?flavor=dump&revision=123
- URLs without *?revision* always return the latest data

Linked Data Interface

Used for example by:

- OpenAlex
- openalex.org
- github.com/ourresearch/openalex-guts



The screenshot shows the OpenAlex website homepage. At the top left is the OpenAlex logo, a stylized 'A' inside a hexagon, followed by the text 'OpenAlex'. To the right is a navigation menu with links for 'Home', 'Read the docs' (with an external link icon), 'About', and 'Contact'. The main content area features a large heading: 'An open and comprehensive catalog of scholarly papers, authors, institutions, and more.' Below this is a paragraph of text: 'Inspired by the ancient Library of Alexandria, OpenAlex is an index of hundreds of millions of interconnected entities across the global research system. We're 100% free and open source, and offer access via a web interface, API, and database snapshot.' At the bottom left is a blue button with the text 'LEARN MORE' and an external link icon.

Search

Special page

Search results

To search for Wikidata items by their title on a given site, use [Special:ItemByTitle](#).

Advanced search:

Search in:

[Luna \(Q27897338\)](#)
family name
15 statements, 1 sitelink - 15:49, 6 November 2021

[Luna \(Q4963425\)](#)
female given name
22 statements, 10 sitelinks - 12:45, 11 November 2021

[Luna County \(Q489652\)](#)
county in New Mexico, United States
46 statements, 44 sitelinks - 03:51, 23 August 2021

[Luna Sea \(Q24760\)](#)
Japanese rock band
81 statements, 21 sitelinks - 21:16, 24 February 2022

Search

- We're running Elasticsearch
- UI: **wikidata.org/wiki/Special:Search**
- API:
wikidata.org/w/api.php?action=query&list=search

Useful to know:

- You can make your search more powerful with these additional keywords specific to Wikidata:
haswbstatement, inlabel, wbstatementquantity, hasdescription, haslabel

Use when:

- You're searching for a specific text string
- You know the name of entities you're looking for, not the exact entities themselves
- You can filter your search based on some simple relations within the data

Don't use when:

- Your search involves complex relations within the data

Action API

```
▼ {
  ▼ "entities": {
    ▼ "Q42": {
      "pageid": 138,
      "ns": 0,
      "title": "Q42",
      "lastrevid": 1591415695,
      "modified": "2022-03-11T12:36:46Z",
      "type": "item",
      "id": "Q42",
      ▼ "labels": {
        ▼ "en": {
          "language": "en",
          "value": "Douglas Adams"
        }
      },
      ▼ "descriptions": {
        ▼ "en": {
          "language": "en",
          "value": "English writer and humorist (1952-2001)"
        }
      },
      ▼ "aliases": {
        ▼ "en": [
          ▼ {
            "language": "en",
            "value": "Douglas Noel Adams"
          },
          ▼ {
            "language": "en",
            "value": "Douglas Noël Adams"
          },
          ▼ {
            "language": "en",
            "value": "Douglas N. Adams"
          }
        ]
      },
      ▼ "claims": {
        ▼ "P31": [
```


Action API

Use when:

- You need to edit Wikidata
- You need JSON data of a batch of entities (up to 50 entities per request)

Don't use when:

- You want large sections of all entities (use a dump instead)
- You just want to retrieve the current state of entities in JSON
 - Consider using the Linked Data Interface: responses will more likely be cached resulting in faster requests

- MediaWiki's own API
- Has been extended to include Wikibase-specific actions
- **wikidata.org/w/api.php**
- Explore it at Special:ApiSandbox

Useful to know:

- With the **props** parameter you can filter on parts of entities: labels, descriptions, claims, etc.
- The Wikidata UI uses the API for all editing
- You can retrieve entities using a combination of *client site id* & *page name*: e.g., enwiki & Berlin to get data for Q64

Action API

Policies and recommendations:

- Robot and user agent policies apply
- Use the *maxlag* parameter
- Keep in mind the other recommendations mentioned in [API:Etiquette](#)

Action API

Used for example by:

- Monumental
- monumental.toolforge.org
- github.com/hatnote/monumental

MONUMENTAL Brandenburg Gate LOGIN

Brandenburg Gate


Brandenburger Tor

WIKIDATA EDIT

Dorotheenstadt

110 images and 19 subcategories 76 languages

architectural heritage monument visitberlin.de/de/ort/brandenburger-tor



Wikipedia

EN DE

The **Brandenburg Gate** (German: *Brandenburger Tor* [ˈbrʌndŋˌbʊʁɐ ˈtoːɐ] (listen)) is an 18th-century neoclassical monument in Berlin, built on the orders of Prussian king Frederick William II after the temporary restoration of order during the Batavian Revolution. One of the best-known landmarks of Germany, it was built on the site of a former city gate that marked the start of the road from Berlin to the town of Brandenburg an der Havel, which used to be the capital of the Margraviate of Brandenburg.

REST API

Wikibase REST API 0.1 OAS3

OpenAPI definition of Wikibase REST API
Wikimedia Deutschland - Wikibase Product Platform Team - Website
GNU General Public License v2.0 or later

Servers
https://wikibase.example/w/rest.php/wikibase/v0

items Wikibase Items Wikibase Data Model - Items ^

GET /entities/items/{item_id} Retrieve a single Wikibase Item by ID

labels Wikibase Labels Wikibase Data Model - Terms ^

GET /entities/items/{item_id}/labels Retrieve an Item's labels

descriptions Wikibase Descriptions Wikibase Data Model - Terms ^

GET /entities/items/{item_id}/descriptions Retrieve an Item's descriptions

aliases Wikibase Aliases Wikibase Data Model - Terms ^

GET /entities/items/{item_id}/aliases [WIPI] Retrieve an Item's aliases

statements Wikibase Statements Wikibase Data Model - Statements ^

GET /entities/items/{item_id}/statements Retrieve Statements from an Item

POST /entities/items/{item_id}/statements Add a new Statement to an Item

GET /entities/items/{item_id}/statements/{statement_id} Retrieve a single Statement from an Item

PUT /entities/items/{item_id}/statements/{statement_id} Replace a single Statement of an Item

PATCH /entities/items/{item_id}/statements/{statement_id} Change elements of a single Statement of an Item

DELETE /entities/items/{item_id}/statements/{statement_id} Delete a single Statement from an Item

REST API

Use when:

- You want to access the current data of a Wikidata Item (or part of it)
- You need to edit Wikidata (under active development currently)

Don't use when:

- You want large sections of all entities (use a dump instead)
- You need JSON data of a batch of entities (currently not possible)

- [RESTful](#) API allowing basic accessing and editing of Wikibase/Wikidata data

Useful to know:

- New API that is currently in development, replacing Action API long-term

Dumps

Index of /wikidatawiki/entities/

../			
20220126/	29-Jan-2022 12:09		-
20220128/	28-Jan-2022 23:30		-
20220131/	03-Feb-2022 18:31		-
20220202/	05-Feb-2022 13:49		-
20220204/	04-Feb-2022 23:29		-
20220207/	10-Feb-2022 15:28		-
20220209/	12-Feb-2022 10:21		-
20220211/	11-Feb-2022 23:31		-
20220214/	17-Feb-2022 15:25		-
20220216/	19-Feb-2022 09:26		-
20220218/	18-Feb-2022 23:28		-
20220221/	24-Feb-2022 18:03		-
20220223/	26-Feb-2022 10:12		-
20220225/	25-Feb-2022 23:26		-
20220228/	03-Mar-2022 17:12		-
20220302/	02-Mar-2022 03:43		-
20220304/	04-Mar-2022 23:28		-
20220307/	10-Mar-2022 16:46		-
20220309/	12-Mar-2022 10:57		-
20220311/	11-Mar-2022 23:31		-
dcatap.rdf	12-Mar-2022 11:29	84751	
latest-all.json.bz2	10-Mar-2022 02:22	72787147780	
latest-all.json.gz	09-Mar-2022 17:52	110300189465	
latest-all.nt.bz2	10-Mar-2022 16:46	145354436870	
latest-all.nt.gz	09-Mar-2022 22:24	186891820793	
latest-all.ttl.bz2	10-Mar-2022 03:59	93072933618	
latest-all.ttl.gz	09-Mar-2022 17:52	112846180363	
latest-lexemes.json.bz2	09-Mar-2022 03:42	206381467	
latest-lexemes.json.gz	09-Mar-2022 03:41	286668473	
latest-lexemes.nt.bz2	11-Mar-2022 23:31	582918167	
latest-lexemes.nt.gz	11-Mar-2022 23:25	783971111	
latest-lexemes.ttl.bz2	11-Mar-2022 23:27	319665811	
latest-lexemes.ttl.gz	11-Mar-2022 23:23	404945905	
latest-truthy.nt.bz2	12-Mar-2022 10:57	32685992234	
latest-truthy.nt.gz	12-Mar-2022 07:51	53922332817	

Dumps

- **dumps.wikimedia.org**
- Various formats available: JSON (recommended), RDF (*all* and *truthy*), XML
- Various mirrors available

Useful to know:

- *Truthy* dumps contain only best-ranked statements and no references or qualifiers
- Wikimedia retains dumps from the last three months
 - Older dumps are often available from the Internet Archive or via torrents

Use when:

- You need data on a significant proportion of entities
- You want to set up your own query service

Don't use when:

- You are severely restricted in bandwidth, storage space or processing power
- You need very current data

Dumps

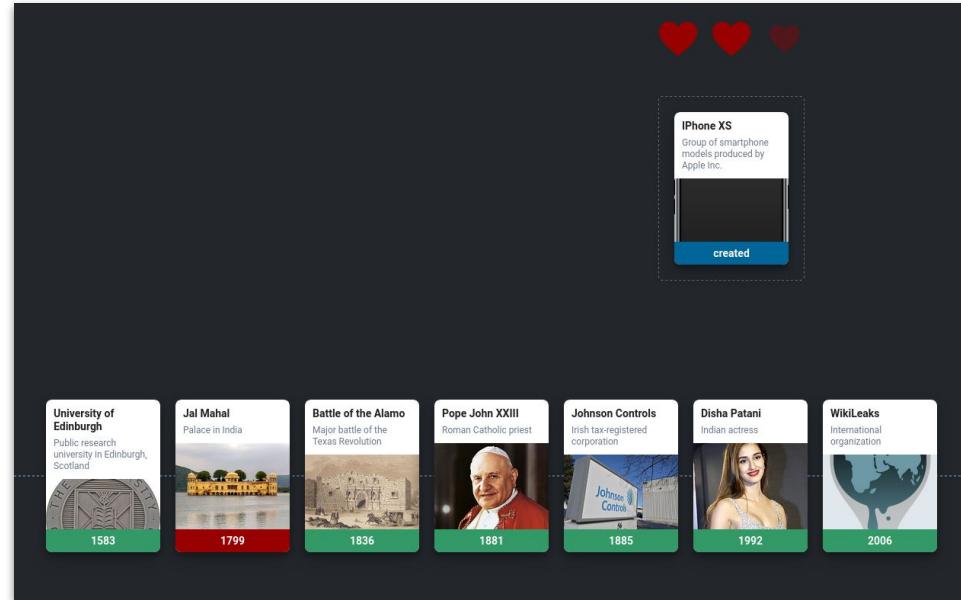
Policies and recommendations:

- We advise against using MediaWiki XML dumps for working with Wikidata's data as these contain the internal entity representation that is not stable
- You can use **wdumper** to get partial custom RDF dumps

Dumps

Used for example by:

- Wikitrivia
- wikitrivia.tomjwatson.com
- github.com/tom-james-watson/wikitrivia



Recent changes stream

- **stream.wikimedia.org** (over HTTP using chunked transfer encoding)
- Per-wiki feeds available in the Action API (*list=recentchanges*)
- Legacy streams available on IRC

Useful to know:

- Returns data for all wikis; filter the stream on your end if you only want Wikidata
- Includes many events, you want “mediawiki.revision-create” to know when entities has changed
- UI available providing an overview / example

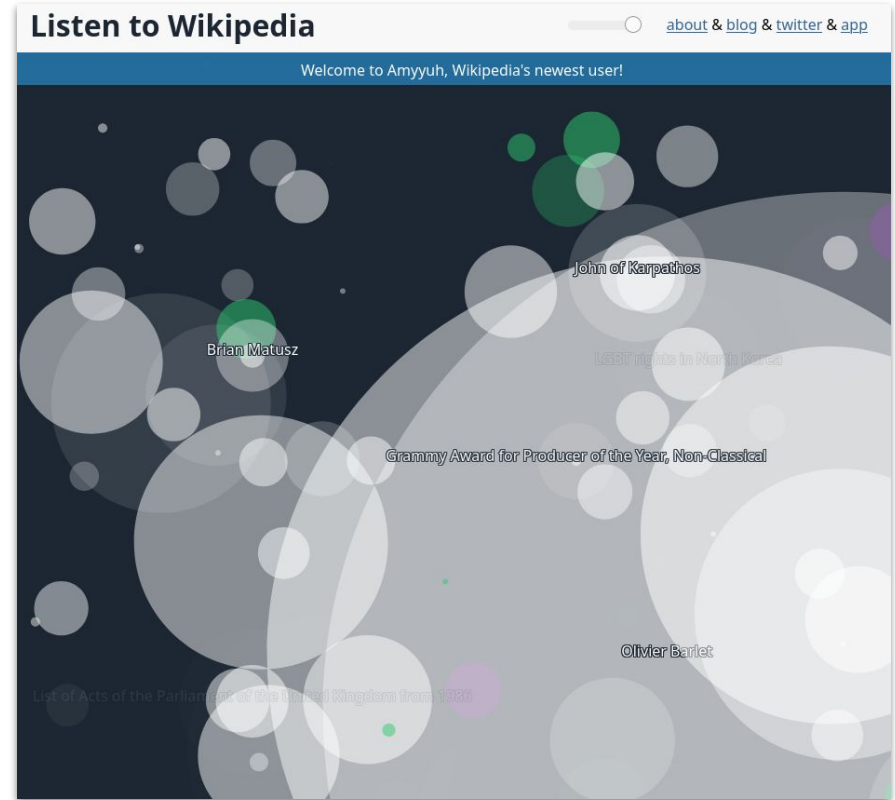
Use when:

- You need to react to changes in real time
- You want to keep up with everything happening on Wikidata (for example, to keep your own query service up to date)

Recent changes stream

Used for example by:

- Listen to Wikipedia
- listen.hatnote.com
- github.com/hatnote/listen-to-wikipedia





Useful tools to know



Constraints Checks

Constraints

property constraint

subject type constraint [edit](#)

class

- human
- character that may or may not be fictional
- fictional character
- mythical character
- fictional human formerly considered to be historical
- prosopographical phantom
- hypothetical person
- individual animal

relation

- instance of

position held

member of the Wisconsin State Assembly ... [edit](#)

0 references

Suggestions [Help](#) [Discuss](#)

required qualifier constraint

This position held statement is missing a qualifier start time.

- Way to define how specific Properties should be used
- Notification is shown when a statement violates a constraint right next to the statement

Query Builder

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...

<input type="button" value="With"/>	<input type="button" value="Without"/>	Property [⊙] <input type="text" value="Enter a property"/>	matching	Value [⊙] <input type="text" value="Enter a value"/>	References [⊙] <input type="button" value="with and without references"/>	<input type="button" value="✕"/>
-------------------------------------	--	--	----------	--	---	----------------------------------

Settings

Limit the number of results to

Show IDs instead of labels (may prevent timeout)

Get a shareable link

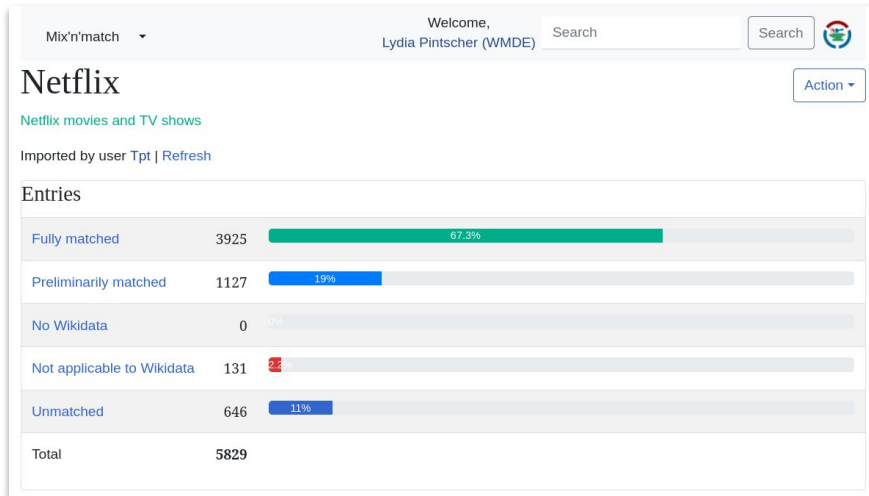
[Show query in the Query Service](#)

Results

Results will be displayed here

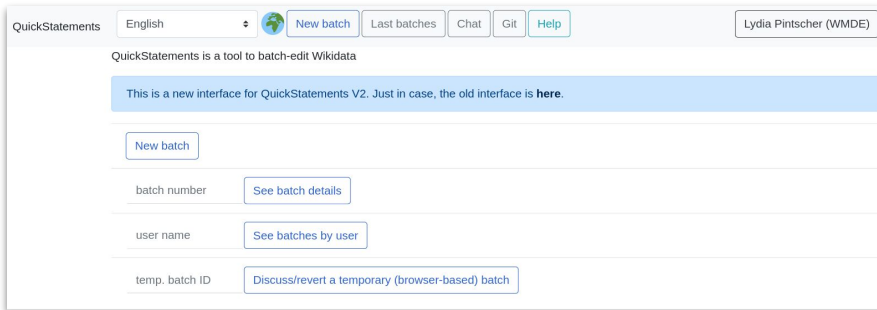
- query.wikidata.org/querybuilder
- Visual interface to create SPARQL queries for Wikidata

Mix'n'Match



- mix-n-match.toolforge.org
- Tool for matching external catalogs to Wikidata

QuickStatements



- quickstatements.toolforge.org
- Tool to import data into Wikidata and make other edits

Mismatch Finder

The screenshot shows the Wikidata Mismatch Finder web application. At the top left is the Wikidata logo and the title 'WIKIDATA MISMATCH FINDER'. A 'Log in' link is in the top right. Below the title is an 'About this tool' section with a 'More information' link. The main text explains that the tool finds data in Wikidata that differs from other sources like catalogs or websites. Below this is a section 'Which Items should be checked?' with a 'Random mismatches' button. A text input field contains the instruction 'Please add one Item Identifier per line' and an example list of QIDs: Q80378, Q33602, Q1459, and Q4524. A 'Check Items' button is at the bottom right of the input area. The footer contains three columns of links: 'About the Wikidata Mismatch Finder', 'About us', and 'More data quality tools'.

WIKIDATA MISMATCH FINDER Log in

About this tool [More information](#)

The Mismatch Finder shows you data in Wikidata that differs from the data in another database, catalog or website (for example, someone's date of birth in Wikidata doesn't match the corresponding entry in the German National Library's catalog). Mismatches like this need fixing, and the Mismatch Finder helps you to do just that.

Which Items should be checked? Random mismatches

Please add one Item Identifier per line

For example:
Q80378
Q33602
Q1459
Q4524

[Check Items](#)

About the Wikidata Mismatch Finder **About us** **More data quality tools**

Licensed under [BSD 3-Clause License](#) [Privacy policy](#) [Query Builder](#)

[Get source code](#) [Wikimedia Deutschland](#) [Item Quality Evaluator](#)

[Report an issue](#) **Made with** [by the Wikidata Team](#) [Curious Facts](#)

[Constraint Violation Checker](#)

- mismatch-finder.toolforge.org
- Tool for suggesting and reviewing corrections to Wikidata's data based on comparisons to other databases
- Can also be used to suggest missing data

Wikxhibit

Wikxhibit

Show off your Wikidata!

Author interactive applications of Wikidata and other sources of data on the web



Documentation

Paintings
Wikidata
Download <HTML <Play!

Artist Page
Wikidata Spotify YouTube Songkick
Download <HTML <Play!

A game to guess the countries
Wikidata
Download <HTML <Play!

- wikxhibit.org
- Simple way to build websites with Wikidata's data
- Especially useful for specialized views on Wikidata's data

Snowman

- github.com/glaciers-in-archive/s/snowman
 - Static site generator for SPARQL endpoints
-

Toolkits

Wikidata Toolkit

Java CI passing codecov 0% maven central 0.13.1 [Project Stats](#)

Wikidata Toolkit is a Java library for accessing Wikidata and other Wikibase installations. It can be used to create bots, to perform data extraction tasks (e.g., convert all data in Wikidata to a new format), and to do large-scale analyses that are too complex for using a simple SPARQL query service.

Documentation

- [Wikidata Toolkit homepage](#): project homepage with basic user documentation, including guidelines on how to setup your Java IDE for using Maven and git.
- [Wikidata Toolkit examples](#): stand-alone Java project that shows how to use Wikidata Toolkit as a library for your own code.
- [Wikidata Toolkit Javadocs](#): API documentation

Knowledge Graph Toolkit (KGTK)

KGTK is a Python library for easy manipulation with knowledge graphs. It provides a flexible framework that allows chaining of common graph operations, such as: extraction of subgraphs, filtering, computation of graph metrics, validation, cleaning, generating embeddings, and so on. Its principal format is TSV, though we do support a number of other inputs.

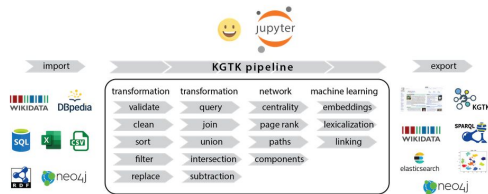
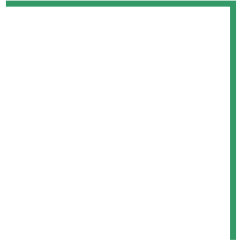
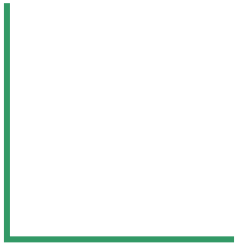


Figure 1: Overview of the usage workflow and features included in KGTK.

- github.com/Wikidata/Wikidata-Toolkit
- kgtk.readthedocs.io
- Make it easier to work with and analyze Wikidata's data dumps

Tips and best practices



Wikidata is a commons

and we all have a role to play to ensure it
stays around for a long time

It's in all our interests to be good citizens.
For people and organisations using
Wikidata's data that specifically means:

- You get better data to build your products and services
- You are doing right by your users by getting them the best data they can
- You protect your reputation
- You help ensure Wikidata stays around for a long time

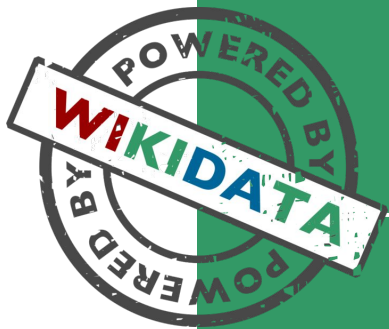
But ultimately it's also just the right thing to do!

Give something back to Wikidata

- Attention and publicity
 - Data improvements (e.g. from your internal quality assurance processes or error reports from your users)
 - Maintenance work (e.g. keeping an eye on changes to the data you are using)
 - Expertise
 - Feedback about what is (not) working well when building on top of our data
 - Money to support development and programmatic work
 - ...
-

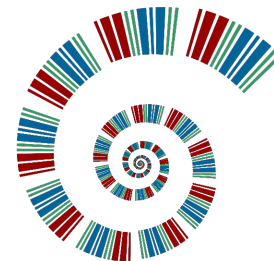
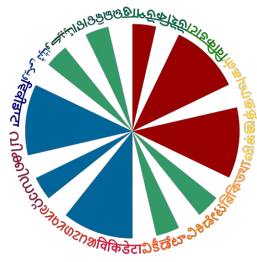
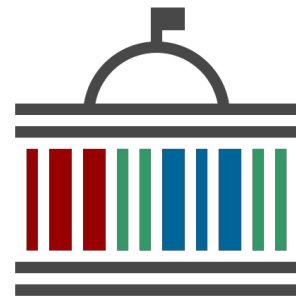
Indicate where the data in your application is coming from

If your users know where the data they see is coming from, they have a chance to improve it for everyone and they will better understand that some mistakes are not on you.



Introduce yourself and your work on your user page

- Disclose if you are paid to edit Wikidata (required by the Terms of Use)
 - Let others know who you are and what you do
 - Be honest and upfront about your motives
-



Keep an eye on
changes to content that
is relevant to you

- [Watchlist](#)
 - [Sparqlrc](#)
 - [Listeria](#)
 - [Integraality](#)
 - Your own internal change tracking
-

Let us know about errors you find

- Small scale: bring it up on-wiki (on Project chat or the applicable Wiki Project)
 - Large scale: publish regular reports, contribute mismatches to Mismatch Finder, ...
-

Fix errors you find

Preferably upstream

- Wikidata is a wiki. You are encouraged to edit!
 - If you are unsure if something should be changed, discuss your edit on the Property talk page, in the appropriate Wiki Project or on Project chat
-

Where to get help?

- Documentation:
[Wikidata:Data access](#)
 - Writing SPARQL queries:
[Wikidata: Request a query](#)
 - General help:
 - [Wikidata mailing list](#)
 - [Wikidata project chat](#)
 - [Wikidata Telegram channel](#)
-

Staying up to date

- [Weekly Summary](#)
 - Social media
 - Mastodon: [@wikidata@wikis.world](#)
 - Twitter: [@Wikidata](#)
-



Where is Wikidata going?



What are we focusing on now?

- Empower editors to increase data quality
 - Facilitate equity in decision making
 - Increase re-use for impact
 - Strengthen underrepresented languages
 - Enable Wikimedia Projects to share their workload
-

Empower editors to increase data quality

- Ensure that the content on Wikidata is of high quality for anyone who re-uses our data.
 - Ensure that the socio-technical system is set up to help editors increase the quality of existing data and contribute high-quality new data.
-

Facilitate equity in decision making

- Ensure that fundamental decisions are made taking into account a diverse set of perspectives
-

Increase re-use for impact

- More people should benefit from the data Wikidata provides
 - Our data is available for anyone to re-use. We want to especially support projects that are aligned with our mission and values and/or that give back to Wikidata.
-

Strengthen underrepresented languages

- More people should have access to technology that supports their language
 - More people should have access to content in their language
-

Enable Wikimedia Projects to share their workload

- Wikimedia projects should be able to rely on Wikidata much more to provide content to their readers and maintain their content
-

Thank you

See you on Wikidata!

Email:

lydia.pintscher@wikimedia.de

Mastodon:

@nightrose@mastodon.online

Twitter:

@nightrose

Wikidata:

Q18016466
