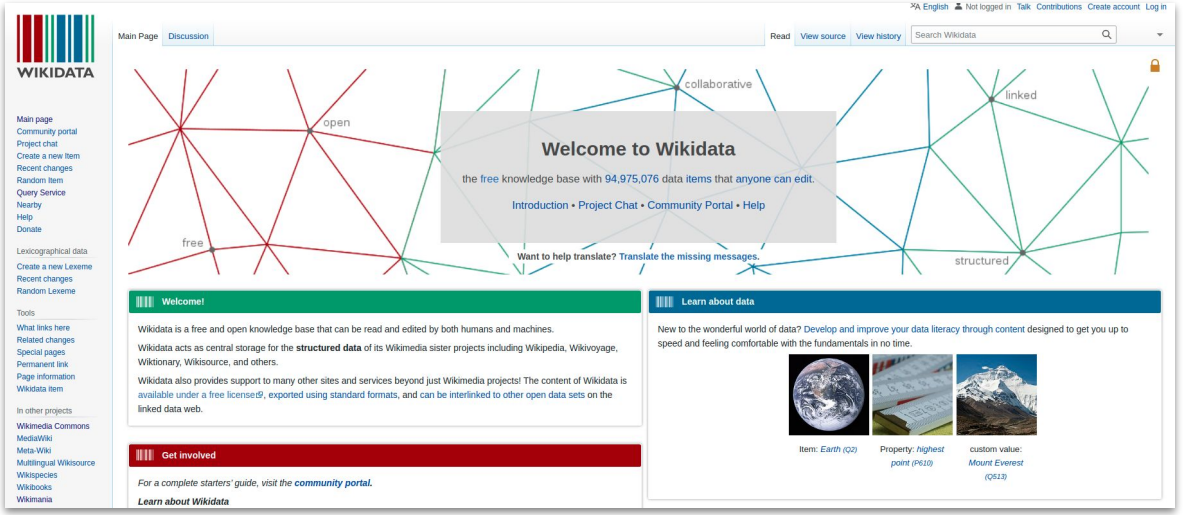
giving more people more access to more knowledge

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

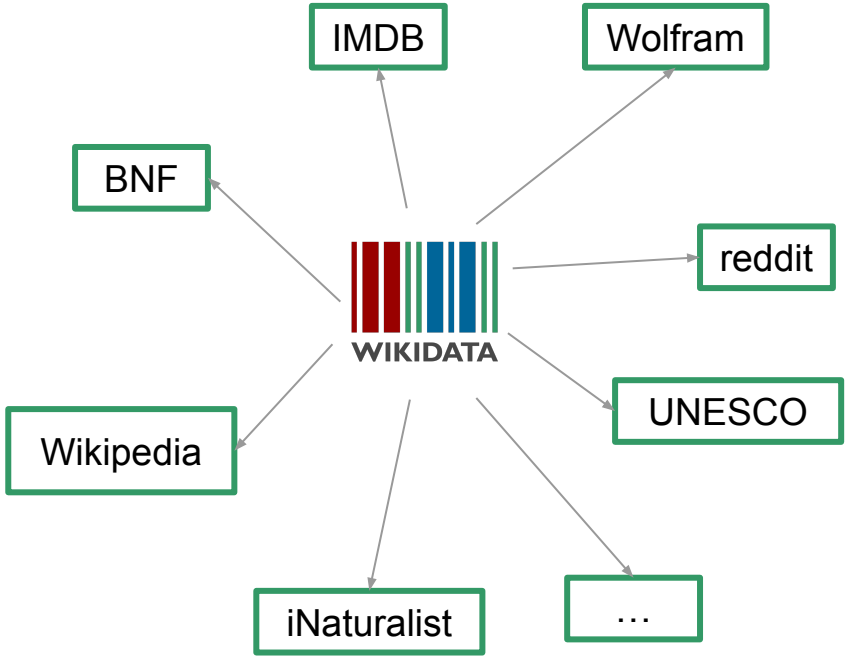
What is Wikidata?

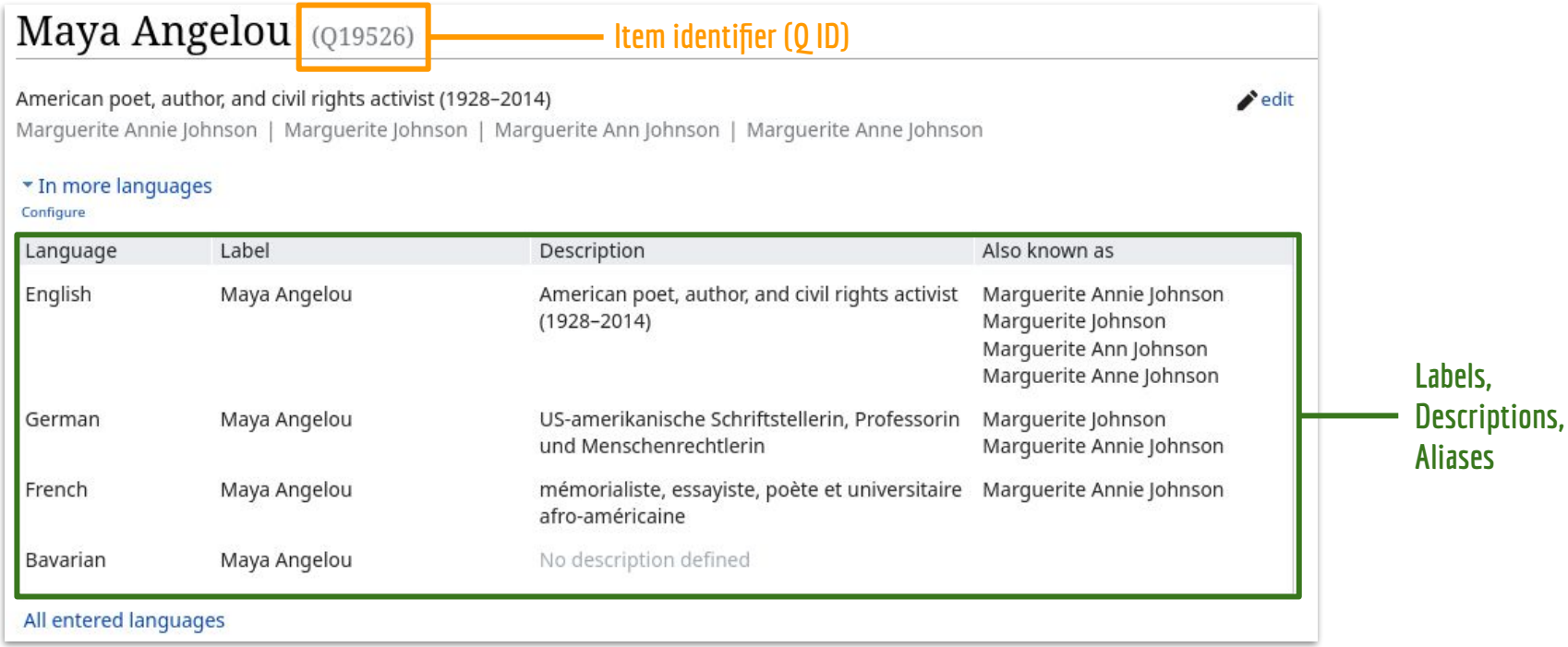
- Wikimedia project started in 2012
- Free and open knowledge graph
- Contains linked data and is linked to a lot of other databases, catalogs, etc.
- Data available under CC0
- Made for humans and machines
- Multilingual
- Collaborative

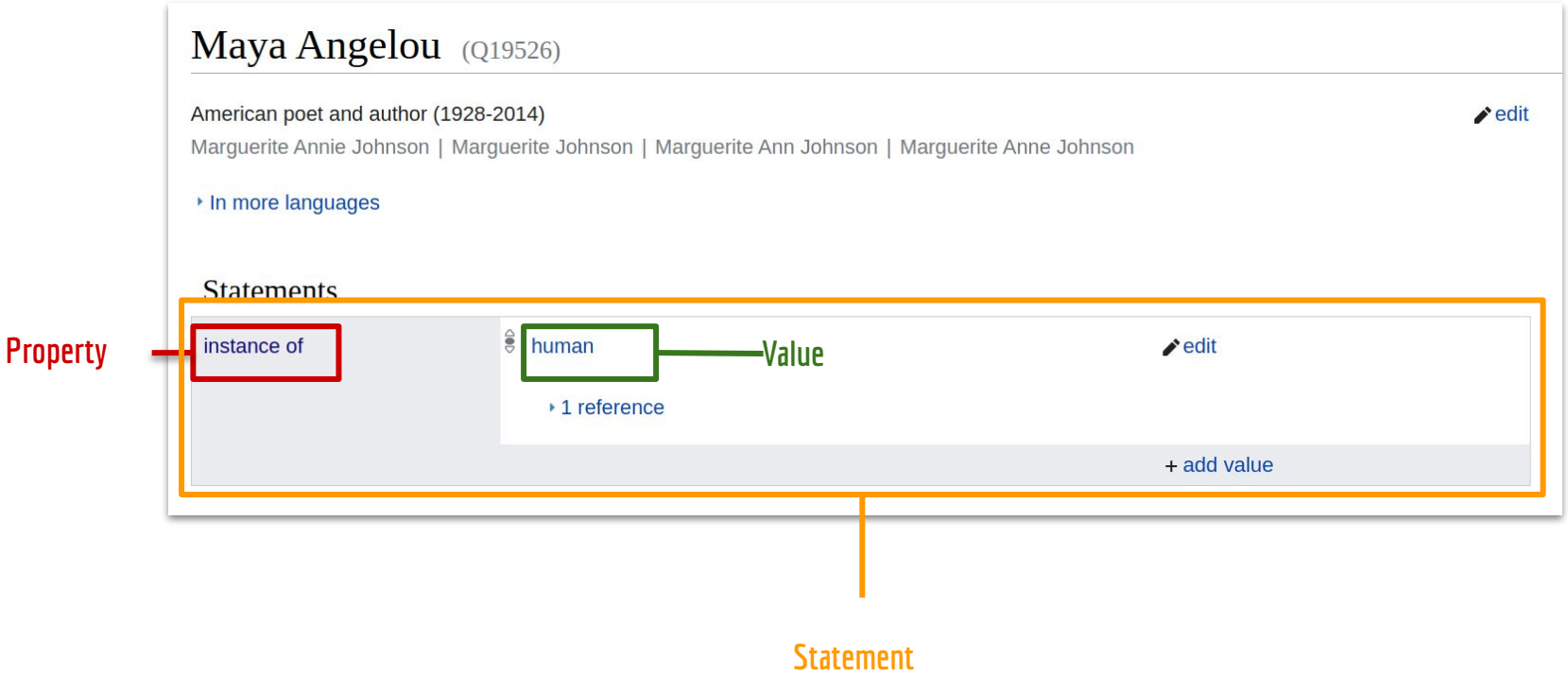


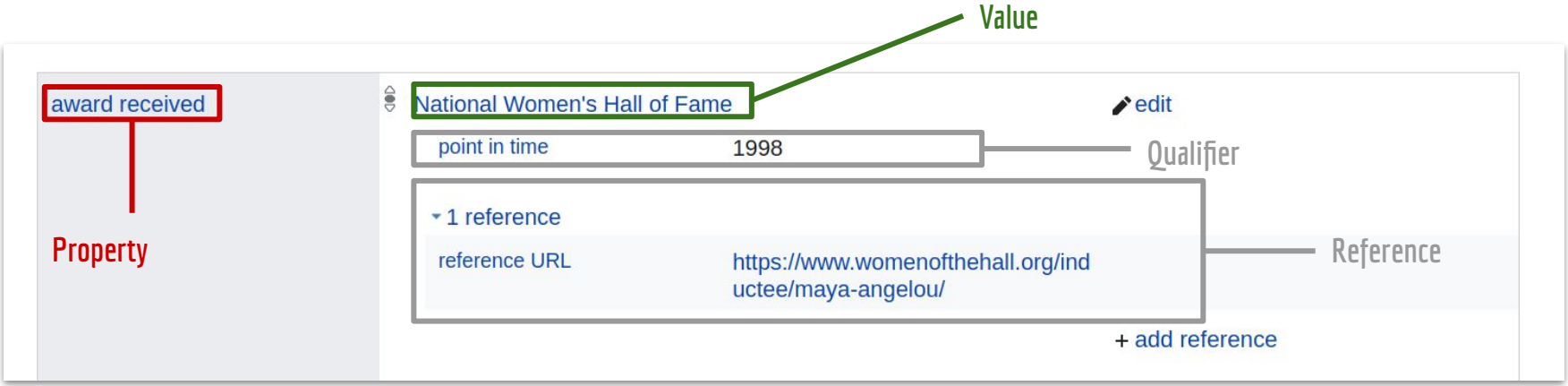
What makes Wikidata special?

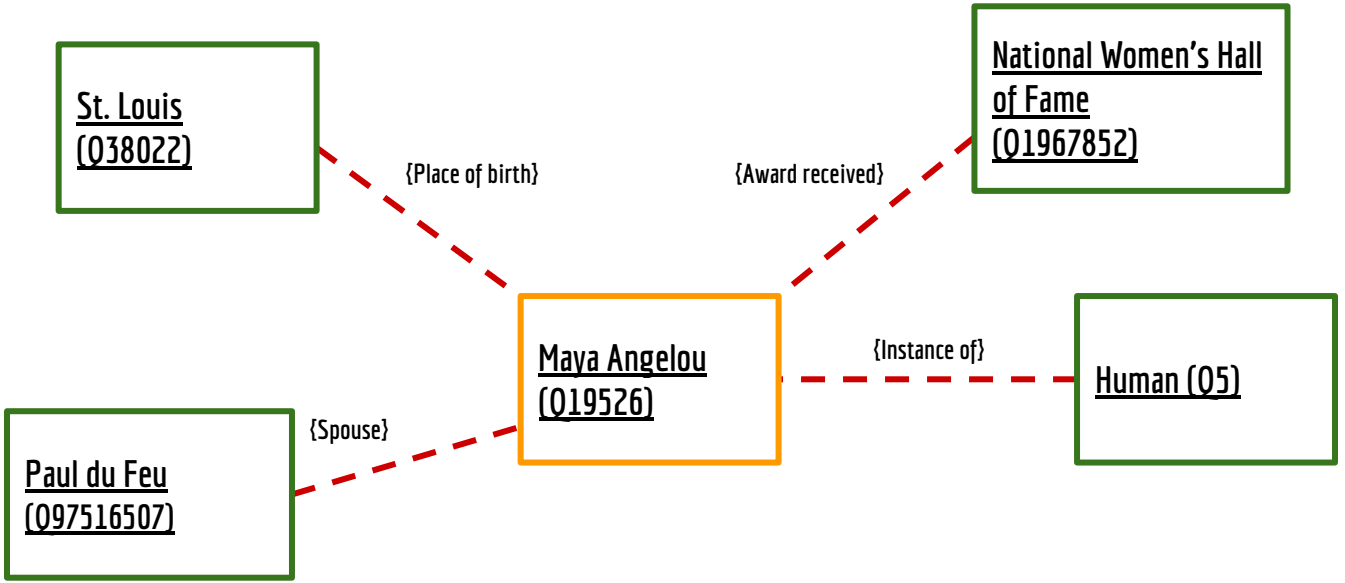
- You can be a part of it
- More nuanced modeling of the world and focusing on verifiability
- Multilingual
- Loosely enforced ontology
- Highly connected internally and to other databases, catalogs, etc. to open up a ton of additional data
- Closely connected to Wikipedia and the other Wikimedia Projects




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

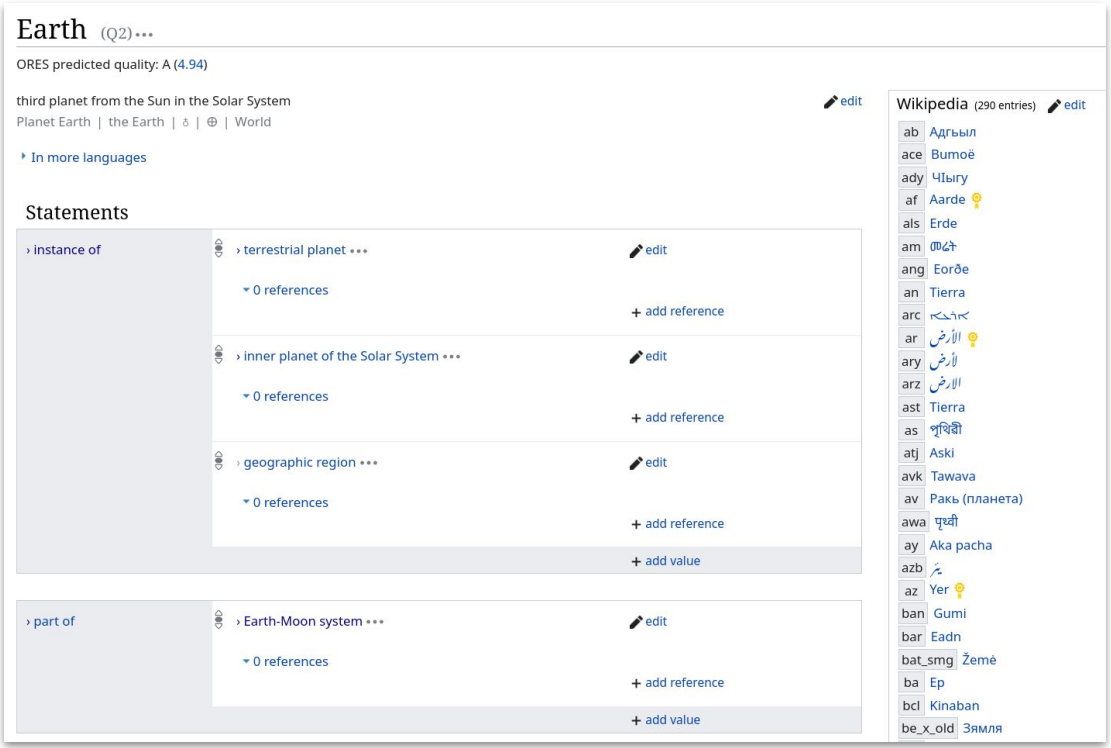

Maya Angelou (Q19526)
American poet and author (1928-2014)
edit
Marguerite Annie Johnson | Marguerite Johnson | Marguerite Ann Johnson | Marguerite Anne Johnson
In more languages
Statements
Property
instance of
human
Value
edit
1 reference
+ add value
Statement


Value
award received
National Women's Hall of Fame
edit
point in time
1998
Qualifier
1 reference
reference URL
https://www.womenofthehall.org/inductee/maya-angelou/
Reference
Property
+ add reference


St. Louis (Q38022)
{Place of birth}
National Women's Hall of Fame (Q1967852)
{Award received}
Maya Angelou (Q19526)
{Instance of}
Human (Q5)
Paul du Feu (Q97516507)
{Spouse}

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

In more languages

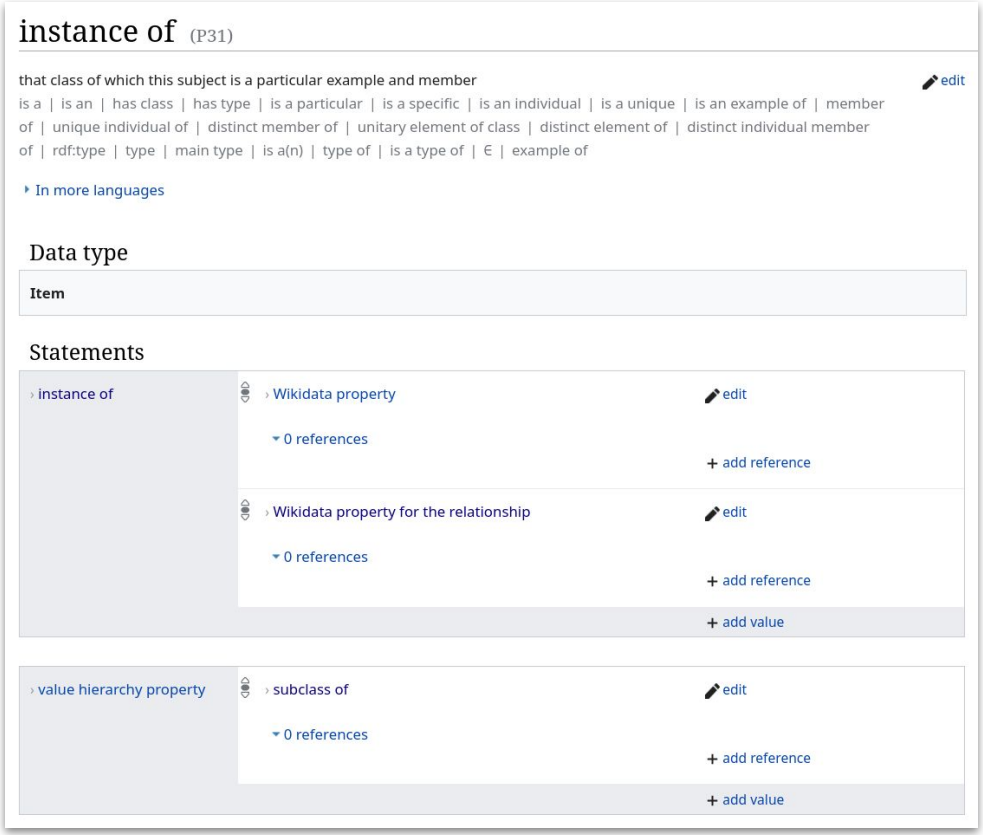
Statements

› instance of
› terrestrial planet
0 references
add reference
edit
› inner planet of the Solar System
0 references
add reference
edit
› geographic region
0 references
add reference
edit
add value

› part of
› Earth-Moon system
0 references
add reference
edit
add value

Wikipedia (290 entries) edit

ab Адгьыл
ace Bumoë
ady ЧIыгу
af Aarde
als Erde
am ሙሪት
ang Eorðe
an Tierra
arc ארעא
ar الأرض
ary الأرض
arz الارض
ast Tierra
as পৃথিৱী
atj Aski
avk Tawava
av Ракь (планета)
awa पृथ्वी
ay Aka pacha
azb یئر
az Yer
ban Gumi
bar Eadn
bat_smg Žemė
ba Ер
bcl Kinaban
be_x_old Зямля

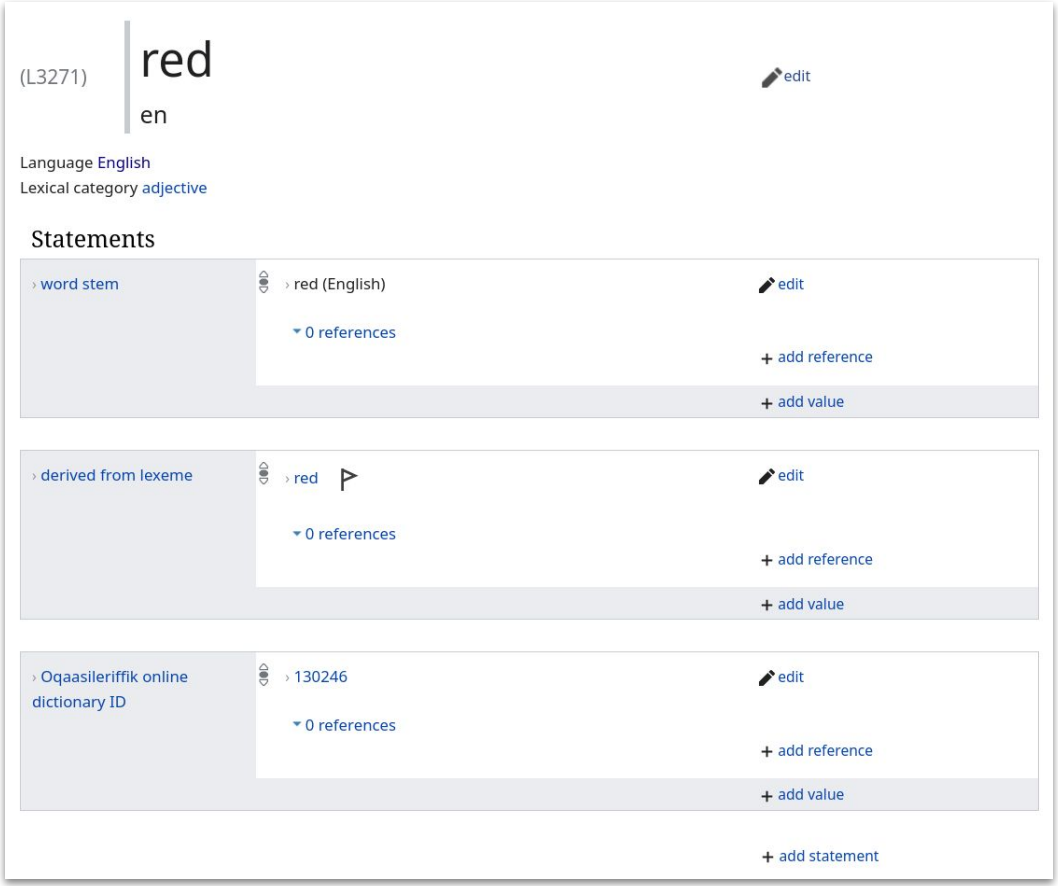


11k

Properties

1.45 Billion

Statements



1 Million

Lexemes

12.5k

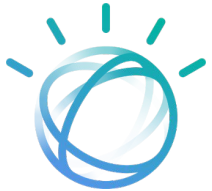
active editors

Photo by Pierre-Selim Huard (CC BY 4.0)

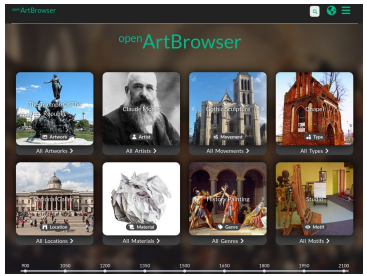
What do people and organisations do with Wikidata's data?

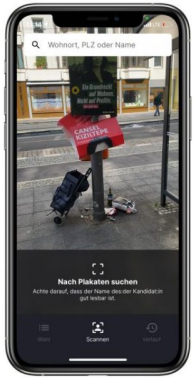

OPEN
LIBRARY


mySociety

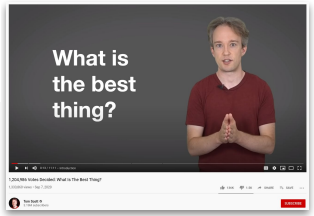



Google


openArtBrowser

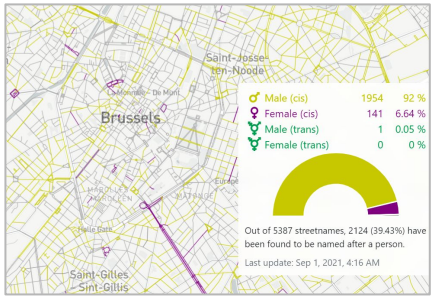

Wohnort, PLZ oder Name
Nach Plakaten suchen
Scannen


OCCRP


What is
the best
thing?

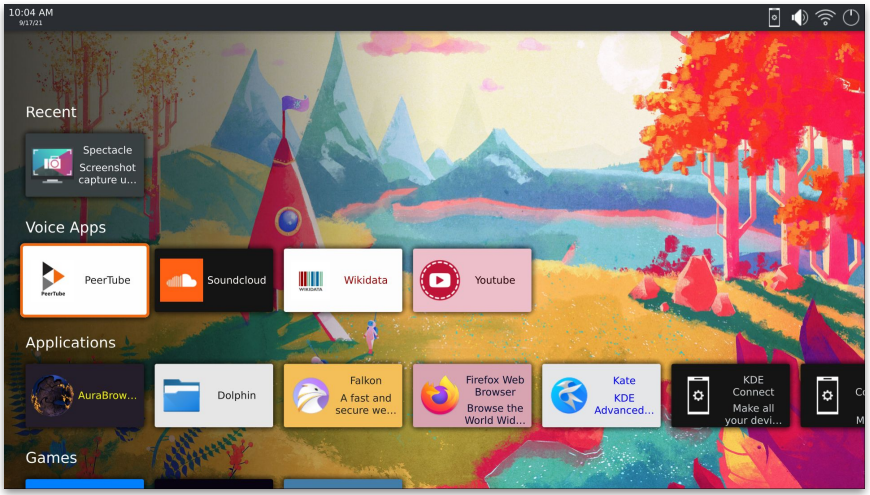

Quora


reddit


Saint-Josse-ten-Noode
Brussels
Saint-Gilles
Sint-Gillis
Male (cis) 1954 92 %
Female (cis) 141 6.64 %
Male (trans) 1 0.05 %
Female (trans) 0 0 %
Out of 5387 streetnames, 2124 (39.43%) have been found to be named after a person.
Last update: Sep 1, 2021, 4:16 AM

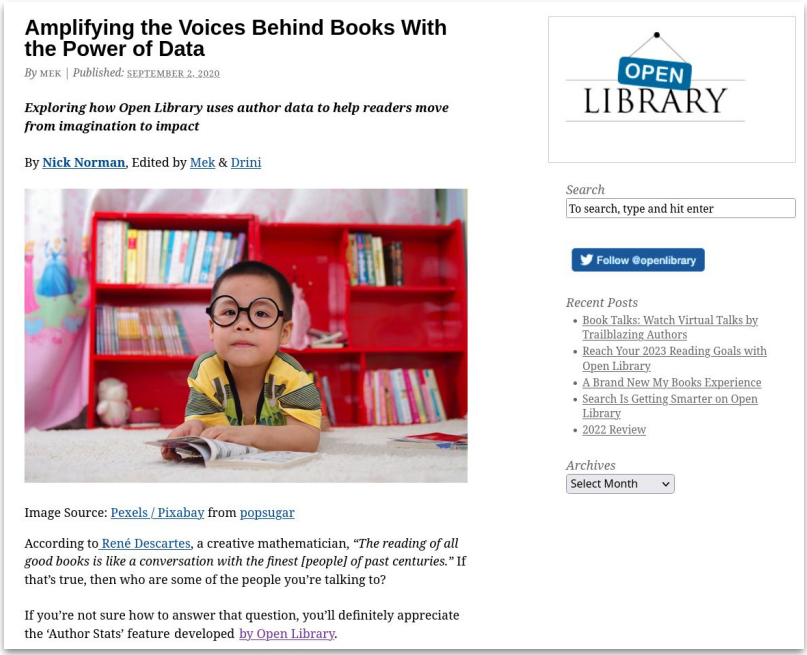

WolframAlpha™

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.

OPEN LIBRARY

Search

To search, type and hit enter

Follow @openlibrary

Recent Posts

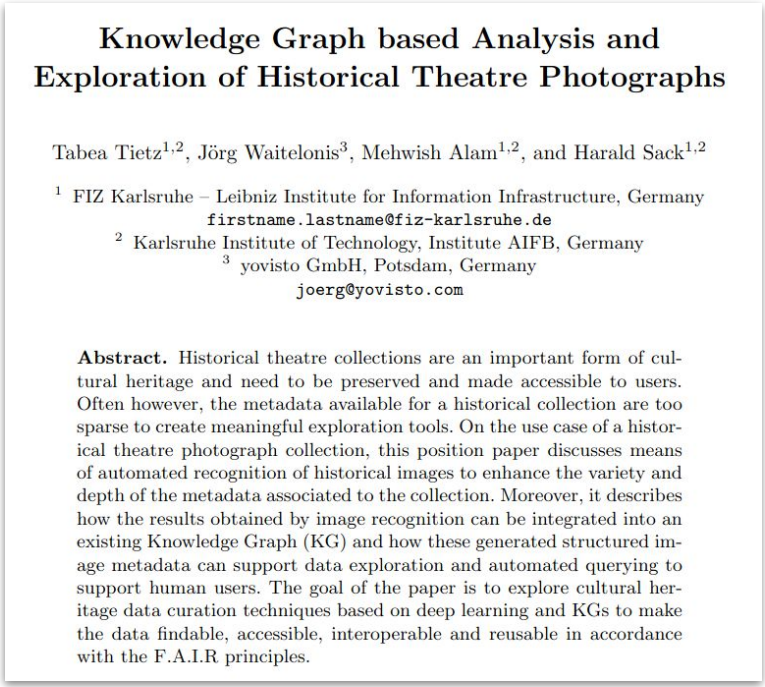
- Book Talks: Watch Virtual Talks by Trailblazing Authors
- Reach Your 2023 Reading Goals with Open Library
- A Brand New My Books Experience
- Search Is Getting Smarter on Open Library
- 2022 Review

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 Waitelonis[3], 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

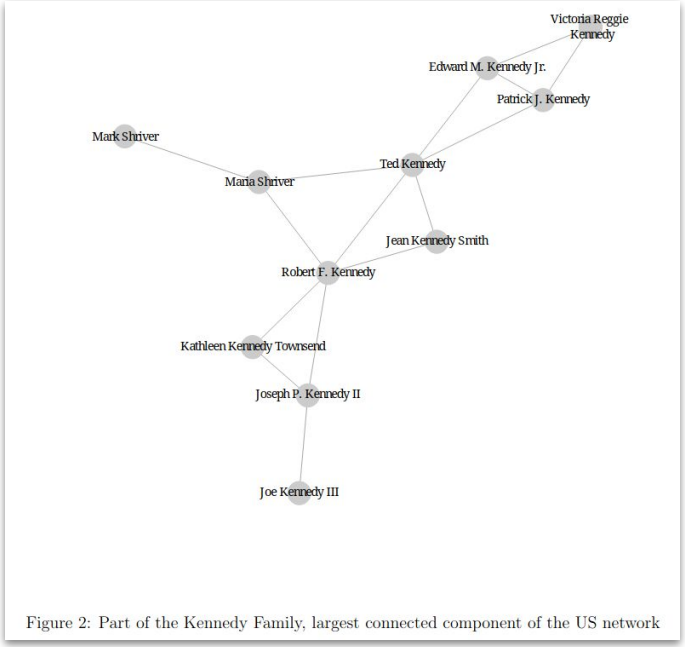
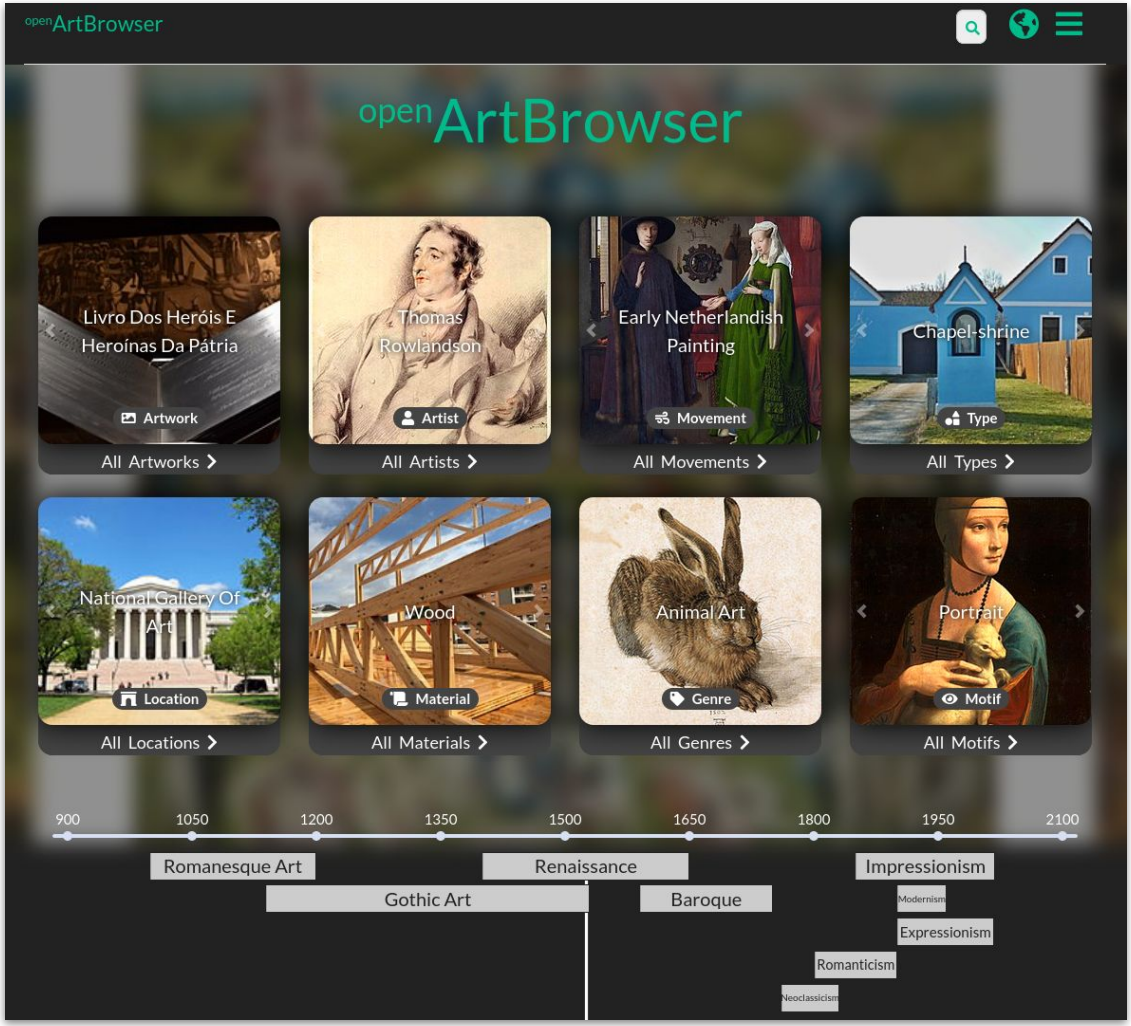


Figure 2: Part of the Kennedy Family, largest connected component of the US network

- 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
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
900
1050
1200
1350
1500
1650
1800
1950
2100
Romanesque Art
Gothic Art
Renaissance
Baroque
Impressionism
Modernism
Expressionism
Romanticism
Neoclassicism

Gateway into the LOD web

HERITAGE CONNECTOR BLOG

SIDESTEPPING THE LIMITATIONS OF COLLECTION CATALOGUES WITH MACHINE LEARNING AND WIKIDATA

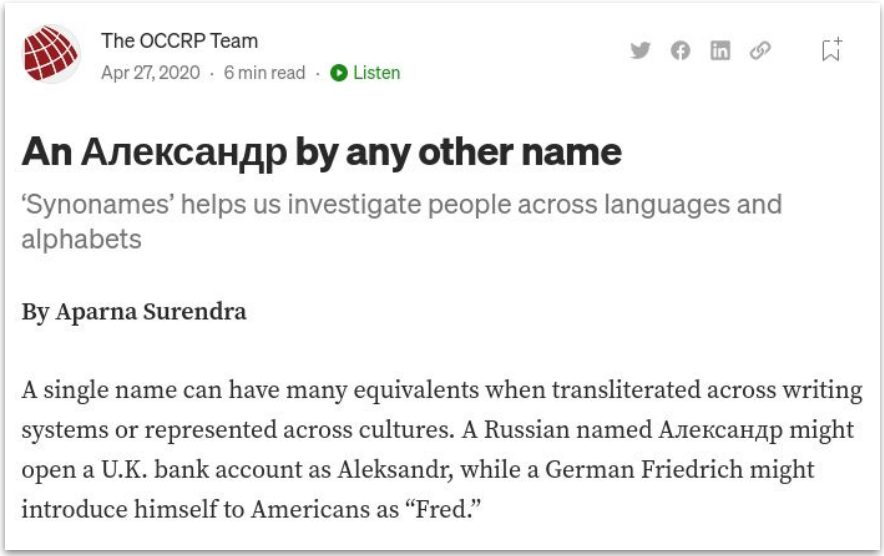
23 September 2020

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 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.

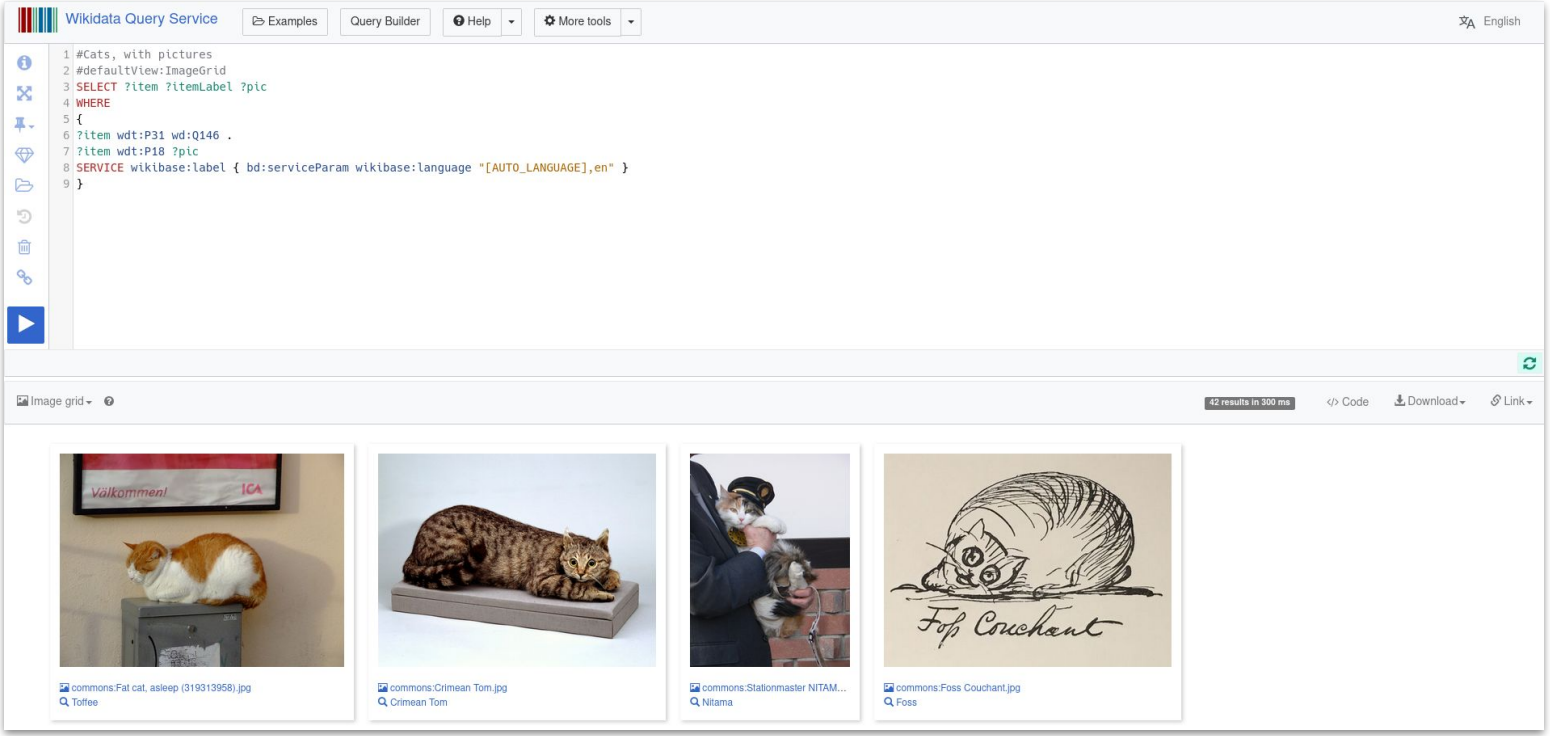
- Wikidata Query Service (WDQS)
- Linked Data Fragments (LDF)
- Linked Data Interface
- Search (Elastic)
- Action API
- REST API
- Dumps
- Recent Changes stream

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

- 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

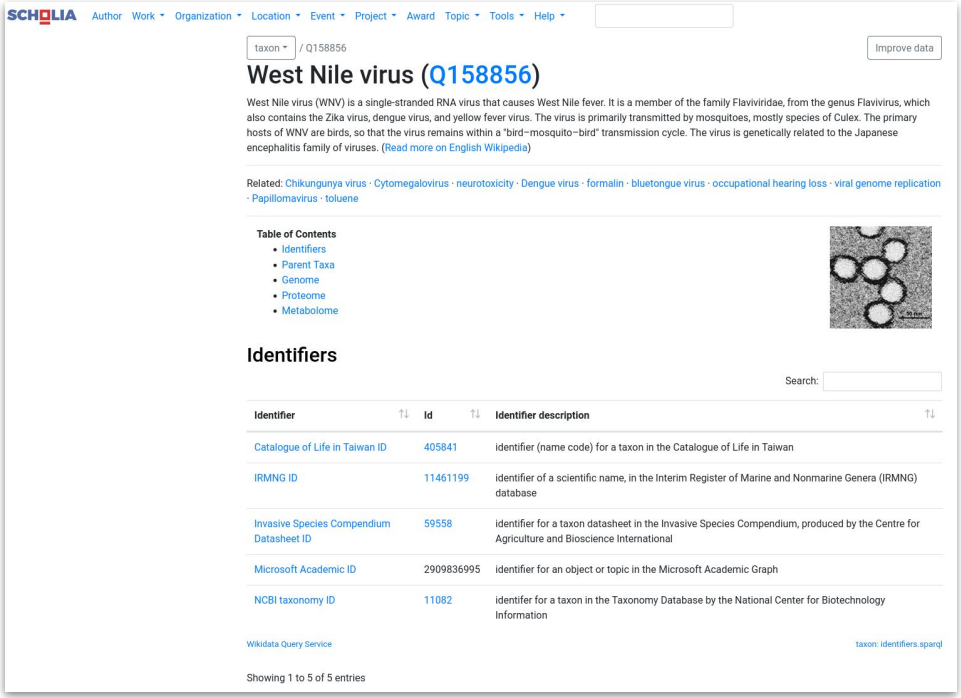
Policies and recommendations:

- 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

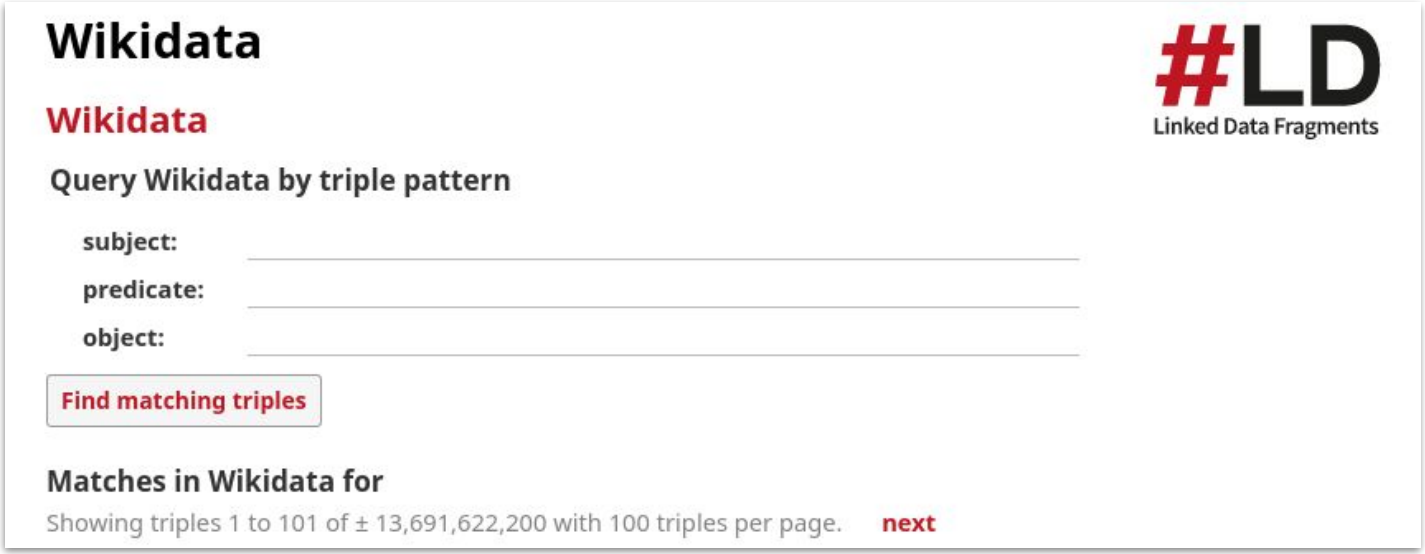
Wikidata Query Service

Used for example by:

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



Linked Data Fragments



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

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

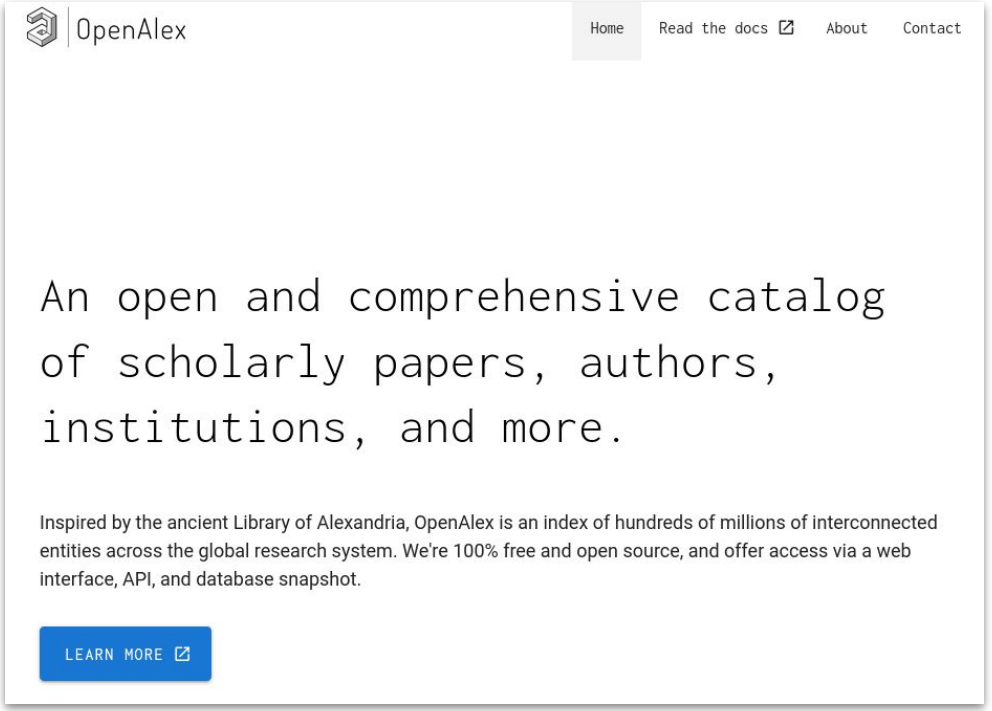
Policies and recommendations:

- 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

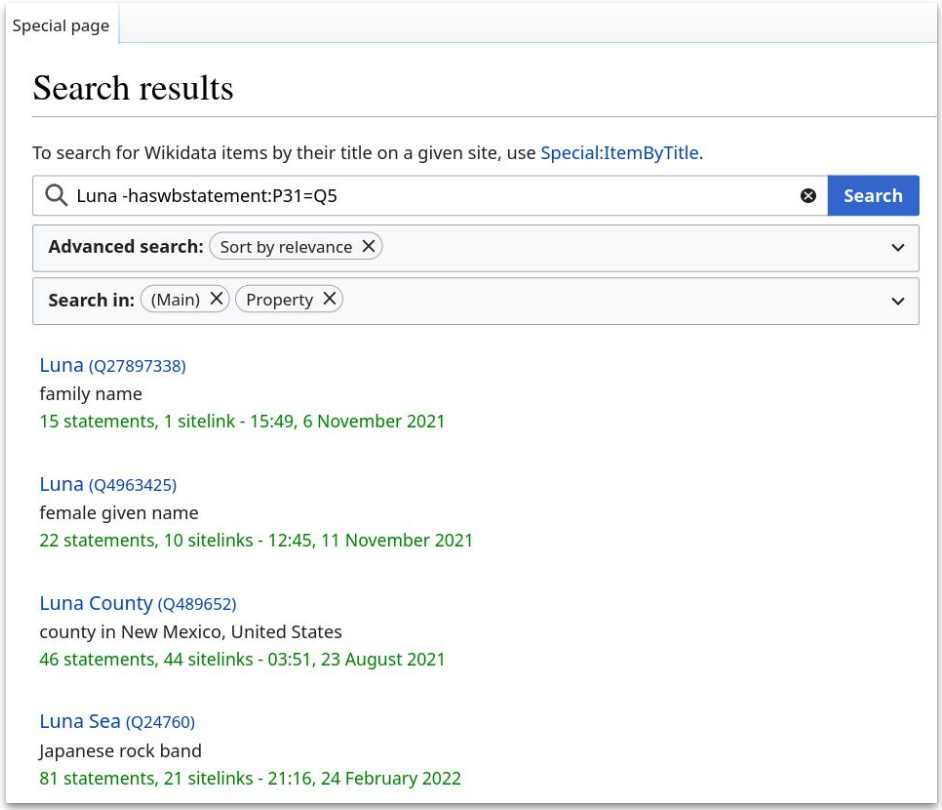
Linked Data Interface

Used for example by:

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



Search


Special page
Search results
To search for Wikidata items by their title on a given site, use Special:ItemByTitle.
Luna -haswbstatement:P31=Q5
Search
Advanced search: Sort by relevance
Search in: (Main) Property
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

- 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

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

Action API

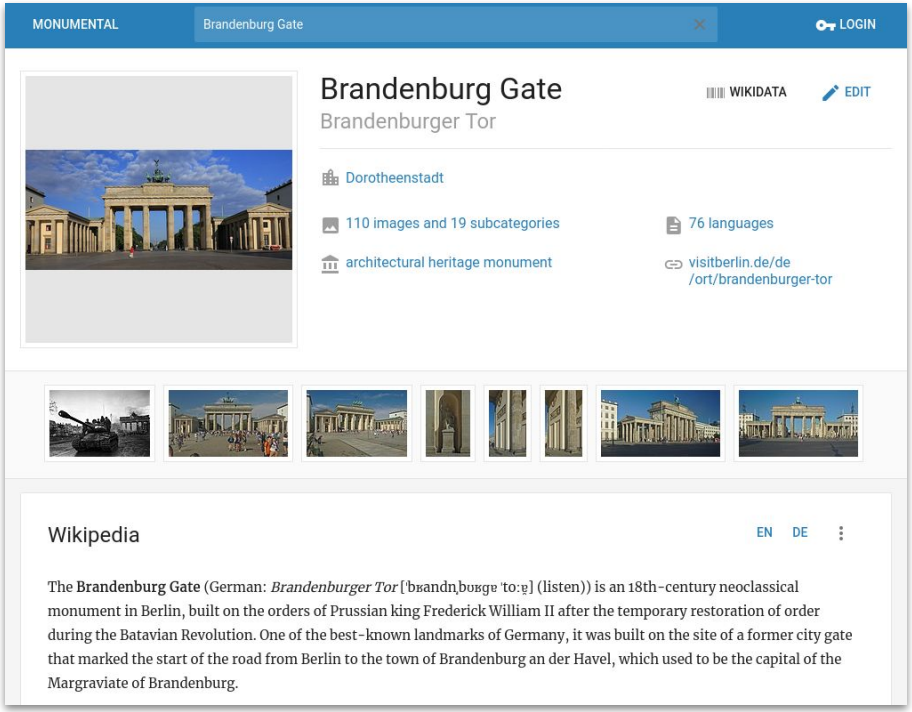
Policies and recommendations:

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

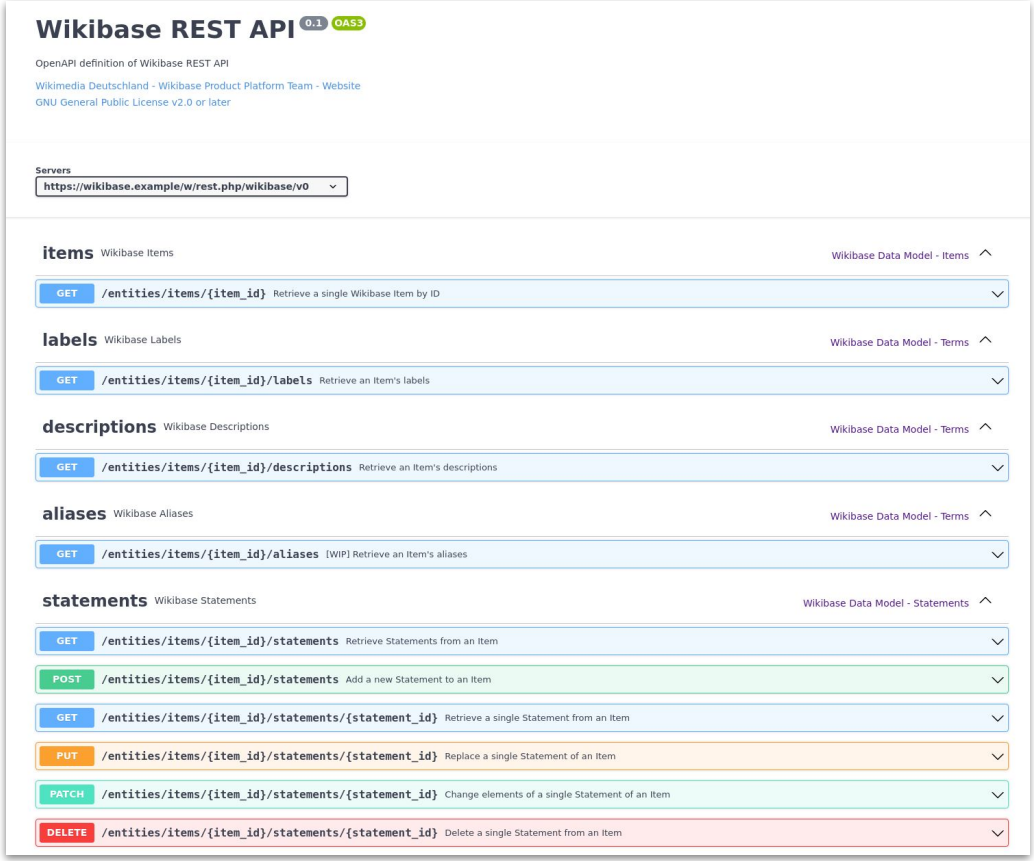
Action API

Used for example by:

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



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 [WIP] 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

- 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

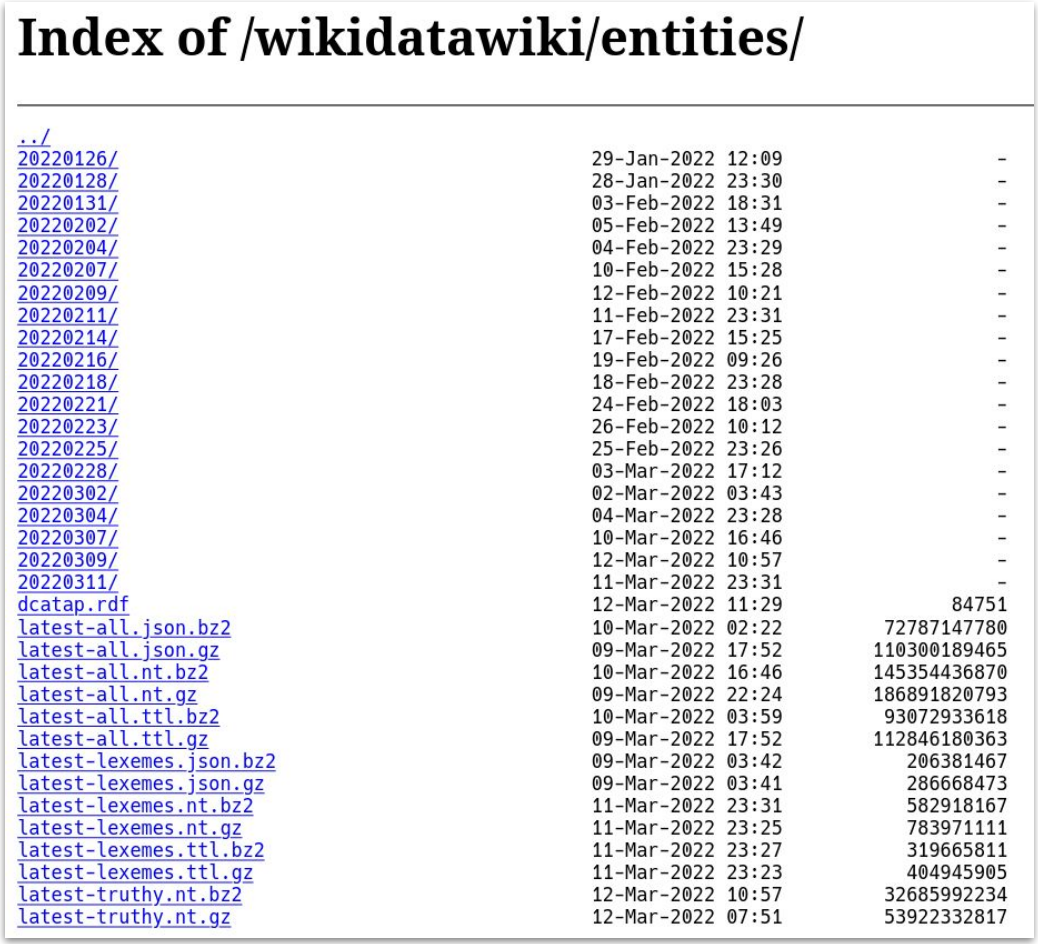
Use when:

- 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)

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

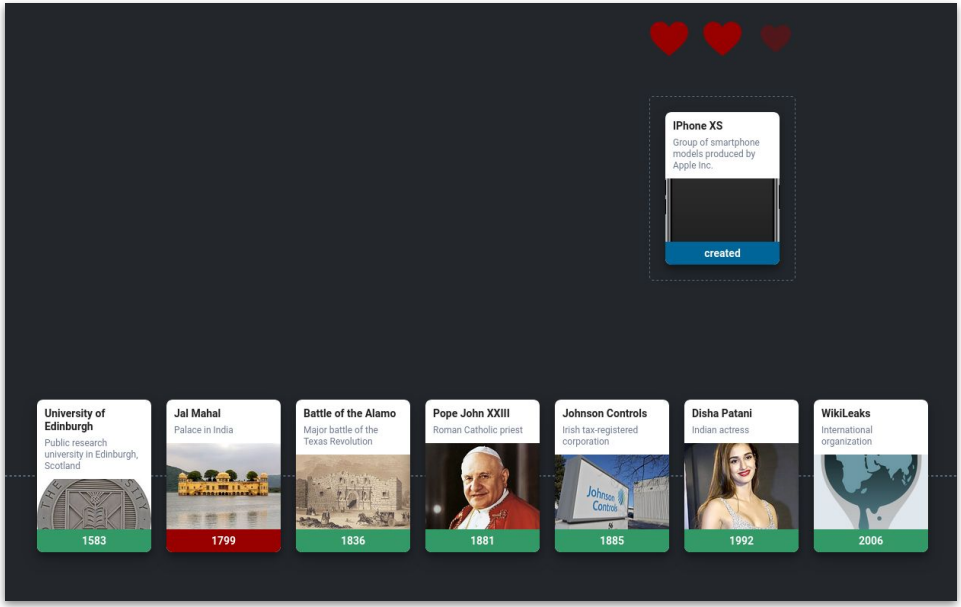
Policies and recommendations:

- 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

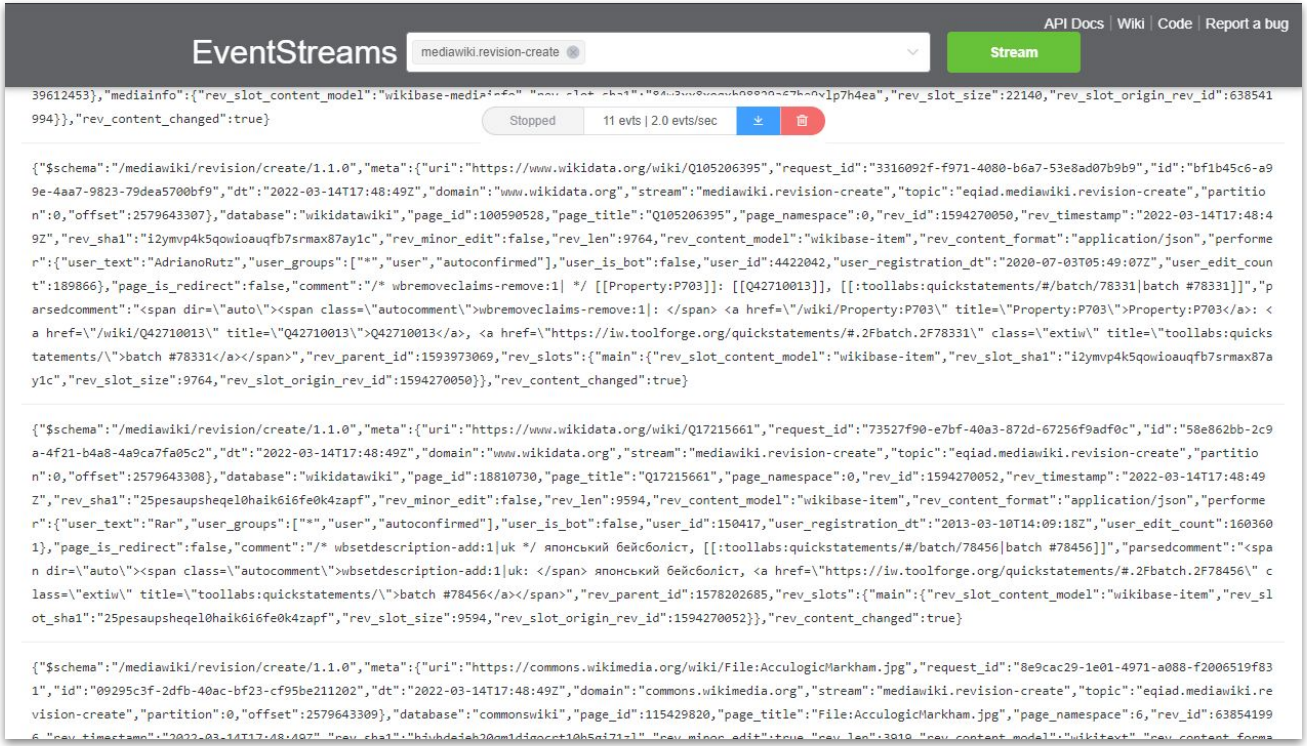
Dumps

Used for example by:

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



Recent changes stream



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

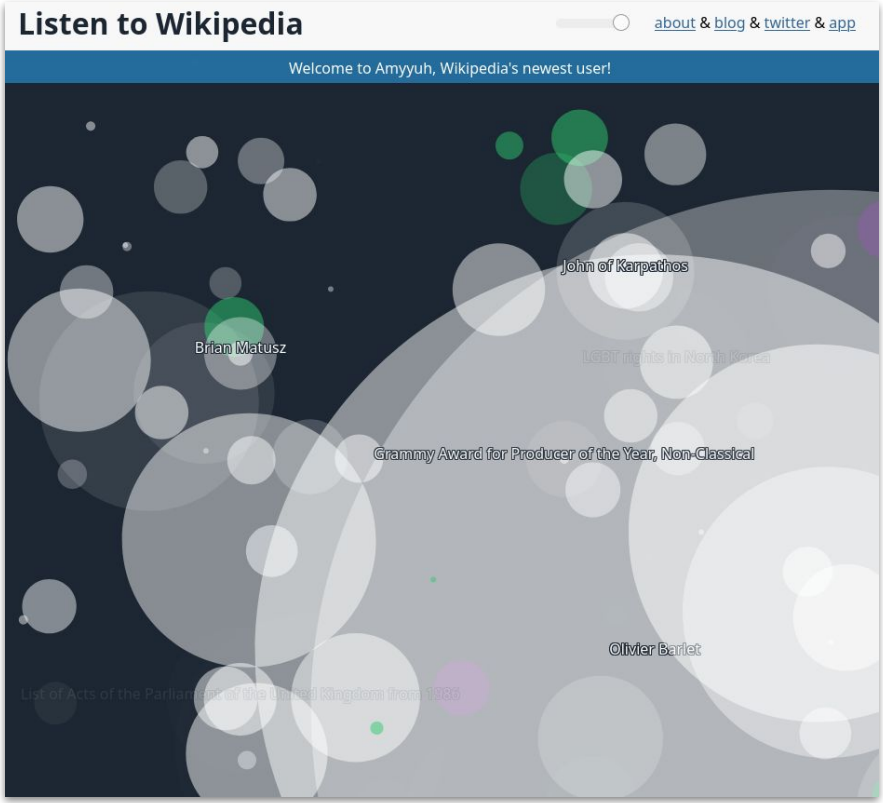
Use when:

- 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

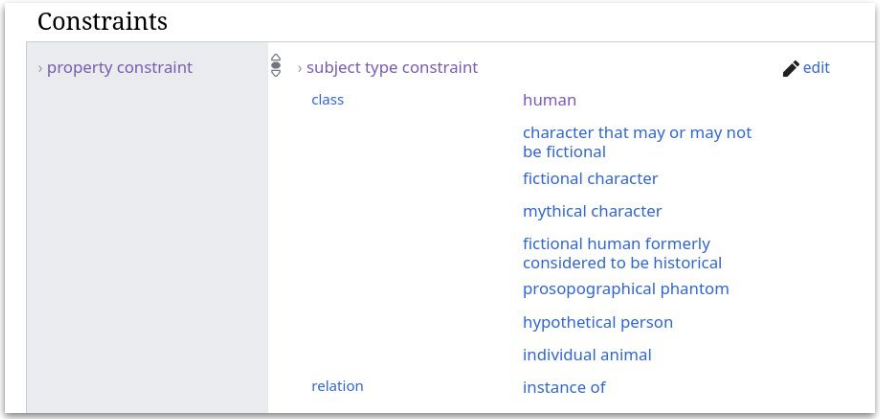
Used for example by:

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



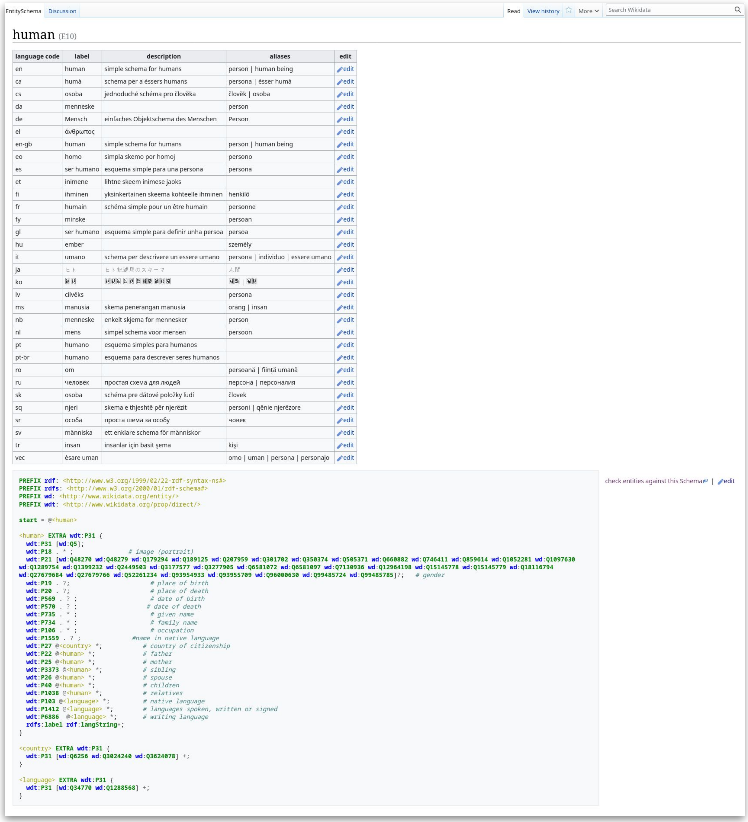
Useful tools to know

Constraints Checks



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

EntitySchemas



EntitySchema | Discussion | Read | View history | More | Search Wikidata

human (E10)

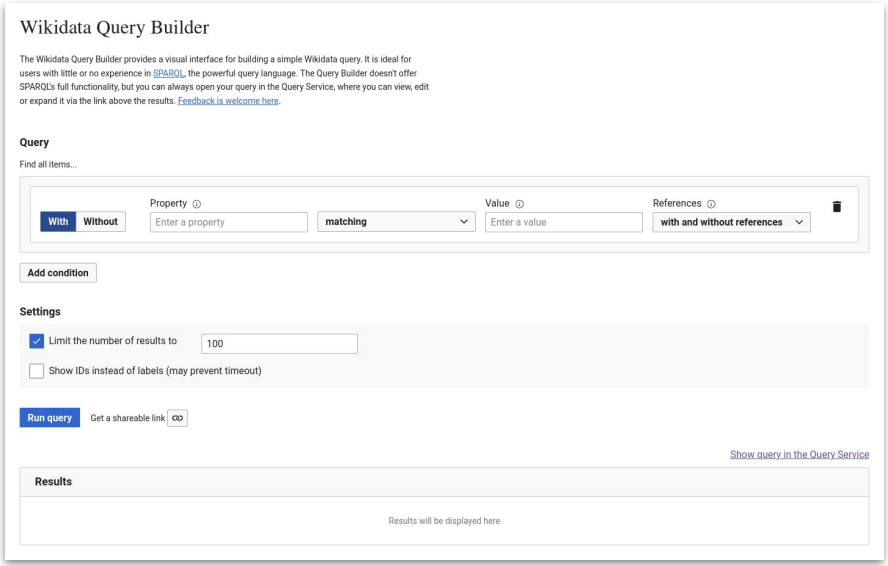
language code	label	description	aliases	edit
en	human	simple schema for humans	person | human being	edit
ca	humà	schema per a éssers humans	persona | ésser humà	edit
cs	osoba	jednoduché schéma pro člověka	člověk | osoba	edit
da	menneske		person	edit
de	Mensch	einfaches Objektschema des Menschen	Person	edit
el	άνθρωπος			edit
en-gb	human	simple schema for humans	person | human being	edit
eo	homo	simpla skemo por homoj	persono	edit
es	ser humano	esquema simple para una persona	persona	edit
et	inimene	lihtne skeem inimese jaoks		edit
fi	ihminen	yksinkertainen skeema kohteelle ihminen	henkilö	edit
fr	humain	schéma simple pour un être humain	personne	edit
fy	minske		persoan	edit
gl	ser humano	esquema simple para definir unha persoa	persoa	edit
hu	ember		személy	edit
it	umano	schema per descrivere un essere umano	persona | individuo | essere umano	edit
ja	ヒト	ヒト記述用のスキーマ	人間	edit
ko	[illegible]	[illegible]	[illegible] | [illegible]	edit
lv	cilvēks		persona	edit
ms	manusia	skema penerangan manusia	orang | insan	edit
nb	menneske	enkelt skjema for mennesker	person	edit
nl	mens	simpel schema voor mensen	persoon	edit
pt	humano	esquema simples para humanos		edit
pt-br	humano	esquema para descrever seres humanos		edit
ro	om		persoană | ființă umană	edit
ru	человек	простая схема для людей	персона | персоналия	edit
sk	osoba	schéma pre dátové položky ľudí	človek	edit
sq	njeri	skema e thjeshtë për njerëzit	personi | qënie njerëzore	edit
sr	особа	проста шема за особу	човек	edit
sv	människa	ett enklare schema för människor		edit
tr	insan	insanlar için basit şema	kişi	edit
vec	èsare uman		omo | uman | persona | personajo	edit

check entities against this Schema | edit

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>

start = @<human>

<human> EXTRA wdt:P31 {
  wdt:P31 [wd:Q5];
  wdt:P18 . * ;               # image (portrait)
  wdt:P21 [wd:Q48270 wd:Q48279 wd:Q179294 wd:Q189125 wd:Q207959 wd:Q301702 wd:Q350374 wd:Q505371 wd:Q660882 wd:Q746411 wd:Q859614 wd:Q1052281 wd:Q1097630
wd:Q1289754 wd:Q1399232 wd:Q2449503 wd:Q3177577 wd:Q3277905 wd:Q6581072 wd:Q6581097 wd:Q7130936 wd:Q12964198 wd:Q15145778 wd:Q15145779 wd:Q18116794
wd:Q27679684 wd:Q27679766 wd:Q52261234 wd:Q93954933 wd:Q93955709 wd:Q96000630 wd:Q99485724 wd:Q99485785]?;   # gender
  wdt:P19 . ?;                   # place of birth
  wdt:P20 . ?;                   # place of death
  wdt:P569 . ? ;                 # date of birth
  wdt:P570 . ? ;                # date of death
  wdt:P735 . * ;                 # given name
  wdt:P734 . * ;                 # family name
  wdt:P106 . * ;                 # occupation
  wdt:P1559 . ? ;            #name in native language
  wdt:P27 @<country> *;          # country of citizenship
  wdt:P22 @<human> *;            # father
  wdt:P25 @<human> *;            # mother
  wdt:P3373 @<human> *;          # sibling
  wdt:P26 @<human> *;            # spouse
  wdt:P40 @<human> *;            # children
  wdt:P1038 @<human> *;          # relatives
  wdt:P103 @<language> *;        # native language
  wdt:P1412 @<language> *;       # languages spoken, written or signed
  wdt:P6886  @<language> *;      # writing language
  rdfs:label rdf:langString+;
}

<country> EXTRA wdt:P31 {
  wdt:P31 [wd:Q6256 wd:Q3024240 wd:Q3624078] +;
}

<language> EXTRA wdt:P31 {
  wdt:P31 [wd:Q34770 wd:Q1288568] +;
}
```

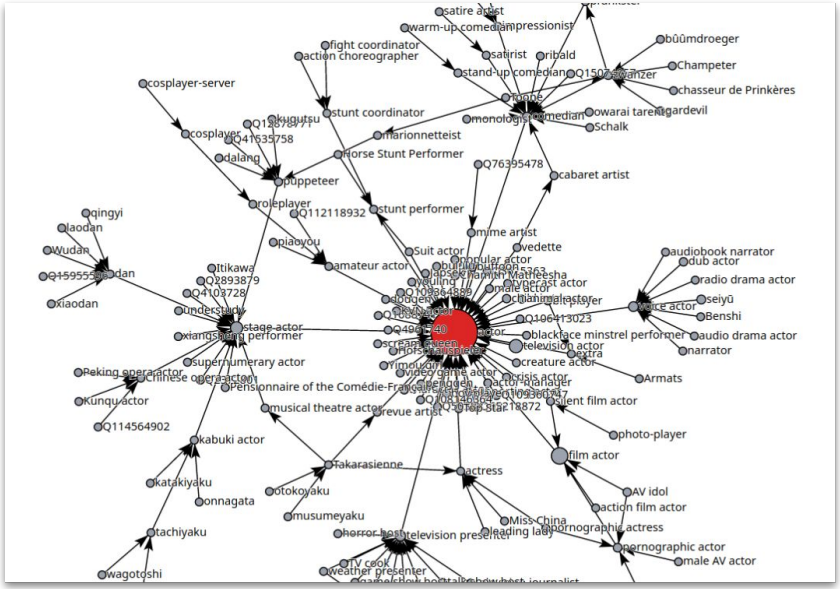
- Definition for how classes should be modeled
- Items can be automatically checked against the EntitySchema
- Using ShEx standard

Query Builder



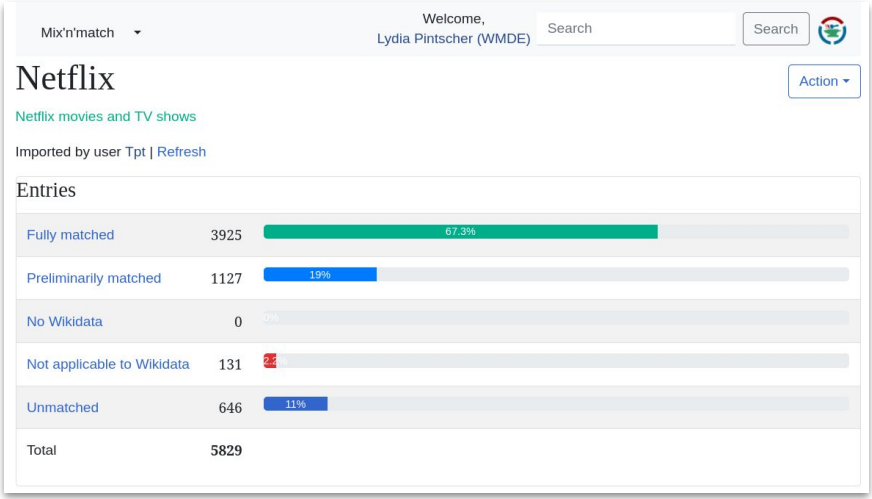
- query.wikidata.org/querybuilder
- Visual interface to create SPARQL queries for Wikidata

Wikidata Graph Builder



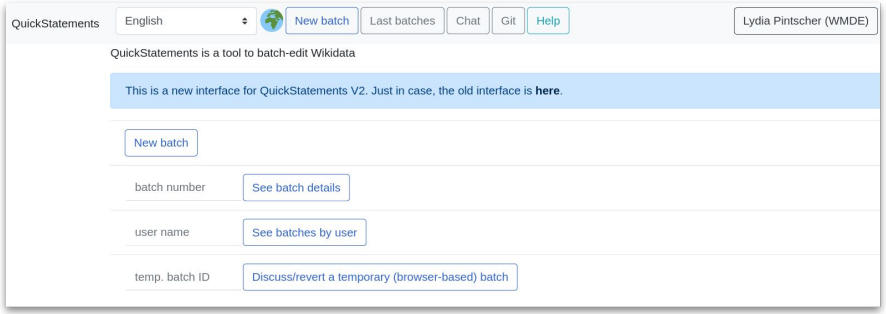
- angryloki.github.io/wikidata-graph-builder
- Visualize the relations going to or from a specific Item, class trees, etc.

Mix’n’Match



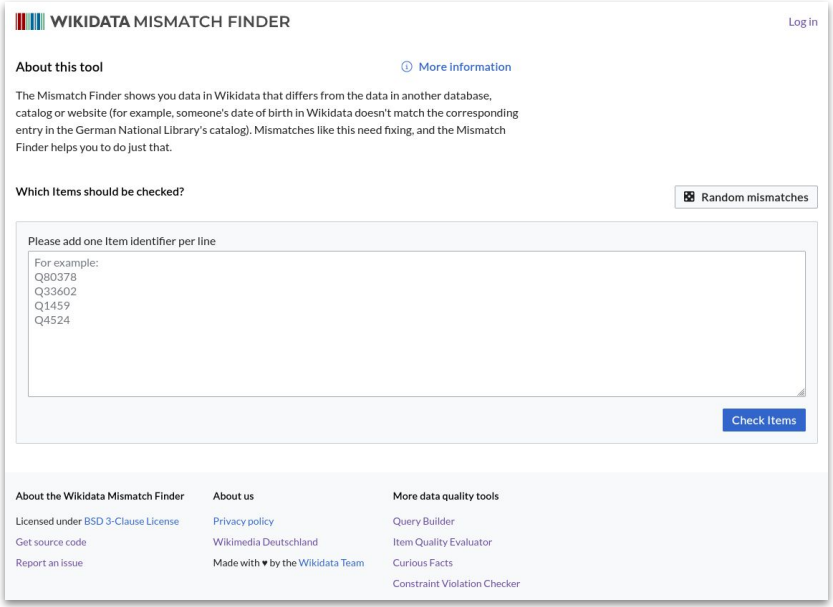
- mix-n-match.toolforge.org
- Tool for matching external catalogs to Wikidata

QuickStatements



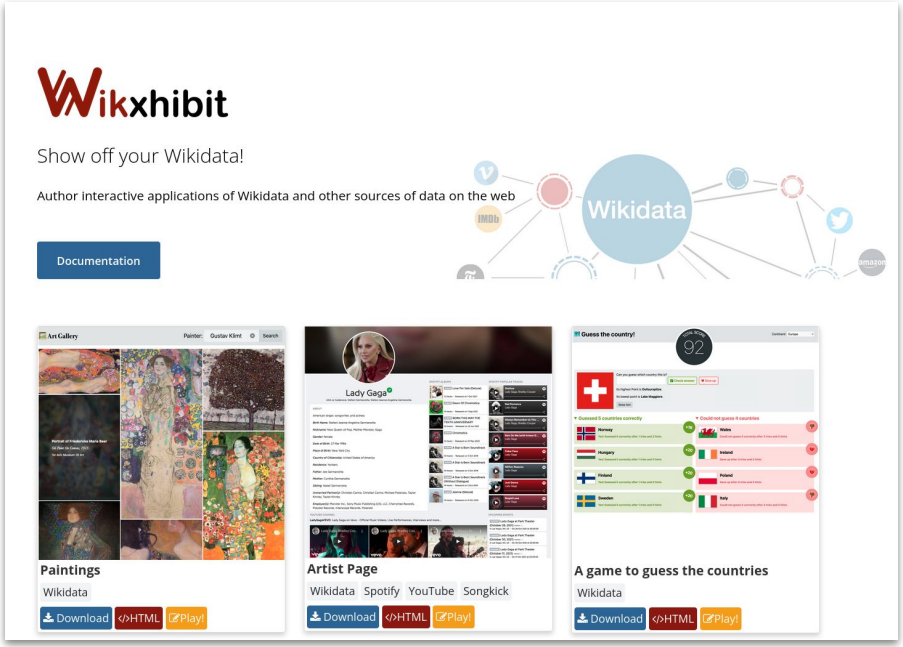
- quickstatements.toolforge.org
- Tool to import data into Wikidata and make other edits

Mismatch Finder



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

Snowman

- github.com/glaciers-in-archives/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.

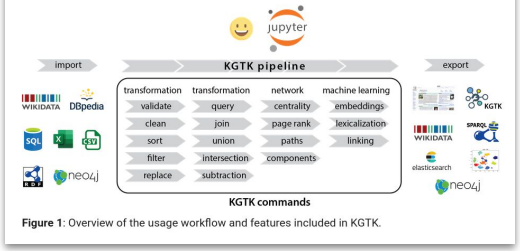


Figure 1: Overview of the usage workflow and features included in KGTK.

- 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

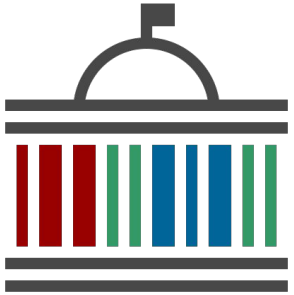
If your users know where the data they see is coming from, they have a chance to improve it for everyone and they will better understand that some mistakes are not on you.

Introduce yourself and your work on your user page

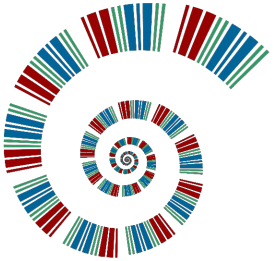
- Disclose if you are paid to edit Wikidata (required by the Terms of Use)
- Let others know who you are and what you do
- Be honest and upfront about your motives


WD:ROW


WIKIDATA




WIKIDATA
BOOKS




WIKIDATA
MEDICINE


WIKIDATA
music


WIKIDATA


WIKIDATA LGBT+


WIKIDATA
ASTRONOMY

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