

# Scaling Wikidata Query Service

unlimited access to all the world's  
knowledge for everyone is hard

**WIKIMANIA  
SINGAPORE**

Lydia Pintscher - @nightrose



# Wikidata Query Service?!

**WIKIMANIA  
SINGAPORE**



# The basics

The Wikidata Query Service is...

- A critical part of Wikidata
- Querying relations that unlock the true power of Wikidata
- Running Blazegraph

# Who uses the Query Service?

**WIKIMANIA  
SINGAPORE**



# Wikidata Editors

- Understand and maintain parts of Wikidata
- Advocacy work and workshops
- Show the world what they worked on



This list is periodically updated by a bot. Manual changes to the list will be reversed in the next update!

2025-04-15 10:00:00 UTC (UTC+08:00) (UTC+08:00) (UTC+08:00)

Item	Image	Person	Born	Died	Description	Occupation	Position held	Stars	Stems
Q113298		Sergei Puskepalis	1966-04-15	2022-09-20	Russian actor	actor theatrical director stage actor film actor film director film screenwriter		20	57
Q115264881		Barbara Witten	1953-10-23	2022-09-20		journalist		0	16
Q1212127		Enrie Mányasi	1954-08-14	2022-09-20	Hungarian journalist			1	20
Q1659935		Marcel Sánchez	1929-03-24	2022-09-20	Spanish association football player	association football player		1	16
Q1748276		Virginio Rognoni	1924-08-05	2022-09-20	Italian politician	political lawyer	Italian Minister of the Interior Italian Minister of Defence Italian Minister of Justice vice president of the Council supérieur de la magistrature member of the Chamber of Deputies of the Italian Republic	14	55
Q216376		Greg Akow	1923-12-11	2022-09-20		author		5	24

# Knowledge seekers/sharers

- Use queries to satisfy curiosity
- Use queries to share something curious with the world (e.g. via social networks)



Larissa Borck  
@Larissa\_Borck

Thanks to [@wikidata](#), here comes an overview on all Swedish citizens who dies in 1949 - and whose works will enter the [#PublicDomain](#) on 1 January 2020. [#PublicDomainDay](#) (h/t [@JolanWuyts](#))  
[query.wikidata.org/#%23%20Swedish...](https://query.wikidata.org/#%23%20Swedish...)

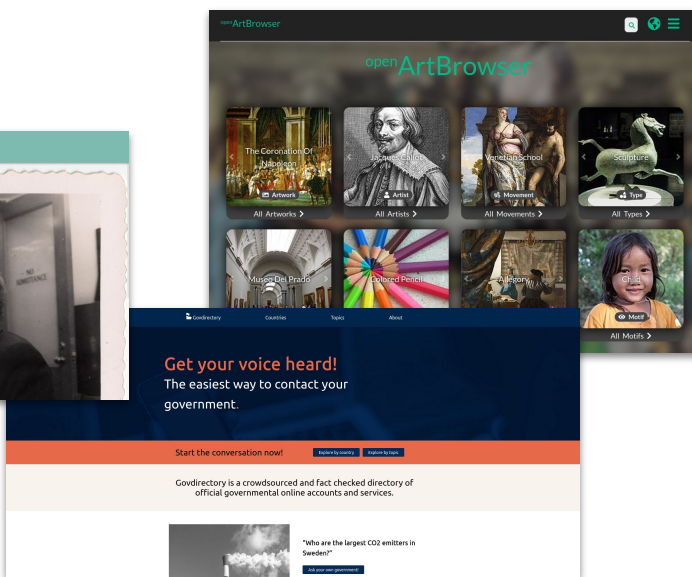
**WikidataFacts**  
@WikidataFacts

popes who were children of other popes:  
[query.wikidata.org/embed.html#%23...](https://query.wikidata.org/embed.html#%23...)

parent	Sergius III
child	John XI
parent	Anastasius I
child	Innocent I
parent	Hormisdas
child	Silverius

# Small and medium sized re-users

- Use queries to power their applications and services



**WIKIMANIA  
SINGAPORE**

# Wikimedia projects

- Structure work in wiki projects and power campaigns, especially around missing content (e.g. Women in Red)
- Better understand the content of the project

WIKIMANIA  
SINGAPORE

Wikipedia:WikiProject Women in Red/Missing articles by occupation/Archaeologists 9 languages

Project page Talk Read Edit View history Pages Tools

From Wikipedia, the free encyclopedia

→ Wikipedia:WikiProject Women in Red / Missing articles by occupation

**Wikirealist index:** **Archaeologists**

Welcome to WikiProject Women in Red (WIR). Our objective is to turn **red links** into **blue ones**. Our scope is women's biographies, women's works, and women's issues, broadly construed.

**This list of red links** is intended to serve as a basis for creating new articles on the English Wikipedia. All new articles must satisfy Wikipedia's *notability* *noting* *noting* red links on this list may or may not qualify.

This table of missing women biographies was generated using Wikidata for Wikipedia:WikiProject Women/Women in Red. See Template:Women in Red for other lists by focus area or by country. The list will be refreshed roughly daily to remove **blue links** - no manual editing is required.

The table can be sorted by sliding on column headers. The 'skelinks' column indicates the number of links to information about the women on other wikis (including commons, source and quote) and is often an indicator of a biography article on another language wikipedia. Skelinks can be accessed via the wikidata link in the 'term' column.

This list is automatically generated from data in Wikidata and is periodically updated by Listgenerator.

**Links made within the list area will be removed on the next update!** Update the list now! [PFRQL?] [Find images?]

name	image	description	country of citizenship	date of birth	date of death	place of birth	place of death	wikidata term	site tasks
<span>Hajnalka Herold</span>	<span></span>	archaeologist	<span>Hungary</span>	<span>Not/unknown value</span>	<span></span>	<span></span>	<span></span>	<span>Q43709818</span>	<span>0</span>
<span>Danka Pavšič</span>	<span></span>	Slovene archaeologist	<span>Slovenia</span> <span> Socialist Federal Republic of Yugoslavia</span>	<span>Not/unknown value</span>	<span></span>	<span></span>	<span></span>	<span>Q4829874</span>	<span>0</span>
<span>Darja Pirhayer</span>	<span></span>	Slovene archaeologist	<span>Slovenia</span> <span> Socialist Federal Republic of Yugoslavia</span>	<span>Not/unknown value</span>	<span></span>	<span></span>	<span></span>	<span>Q48291144</span>	<span>0</span>
<span>Annette Kirsh</span>	<span></span>	German classical archaeologist	<span></span>	<span>Not/unknown value</span>	<span></span>	<span>Not/unknown value</span>	<span></span>	<span>Q55089942</span>	<span>0</span>



# Wikimedia development teams and tool builders

- Queries power tools (Listeria, Item Quality Evaluator, Integrality), often by getting lists of Items

Pandas by place of residence [ edit | edit source ]  
 at entities | control | all groupings | without grouping property | below threshold (1) | Querying about Admins:Kings (24190) | Grouping by residence (40)

Top groupings (Minimum 1 items)		Top Properties (used at least 1 times per grouping)				
Name	Count	sex or gender (P21)	place of birth (P19)	date of birth (P569)	place of death (P20)	date of death (P570)
Hamilton Zoo (Q211584)	10	100.0% (1) #	80.0% (8) #	90.0% (9) #	0% (0) %	0% (0) %
Edmonton Valley Zoo (Q539105)	8	100.0% (8) #	75.0% (6) #	87.5% (7) #	0% (0) %	0% (0) %
Chattanooga Zoo at Warner Park (Q268790)	6	100.0% (6) #	0% (0) %	100.0% (6) #	83.33% (5) #	83.33% (5) #
Chincinelli Zoo and Botanical Garden (Q423333)	6	100.0% (6) #	33.33% (2) #	33.33% (2) #	16.67% (1) %	16.67% (1) %
Rosamond Gifford Zoo (Q2166524)	5	100.0% (5) #	100.0% (5) #	100.0% (5) #	0% (0) %	0% (0) %
Parco Natura Viva (Q236466)	4	100.0% (4) #	75.0% (3) #	75.0% (3) #	0% (0) %	0% (0) %
Belfast Zoo (Q211348)	4	100.0% (4) #	25.0% (1) %	25.0% (1) %	0% (0) %	0% (0) %
Assiniboine Park Zoo (Q468813)	4	100.0% (4) #	50.0% (2) %	50.0% (2) %	0% (0) %	0% (0) %
Adelaide Zoo (Q244627)	3	100.0% (3) #	66.67% (2) %	100.0% (3) #	0% (0) %	0% (0) %
Ménagerie du jardin des Plantes (Q2197762)	3	100.0% (3) #	66.67% (2) %	66.67% (2) %	0% (0) %	0% (0) %
Liberec Zoo (Q231518)	3	100.0% (3) #	33.33% (1) %	100.0% (3) #	0% (0) %	0% (0) %
Parc animalier d'Auvergne (Q166889)	3	100.0% (3) #	0% (0) %	33.33% (1) %	0% (0) %	33.33% (1) %
National Zoo & Aquarium (Q49593)	3	100.0% (3) #	66.67% (2) %	100.0% (3) #	0% (0) %	0% (0) %
Tarangza Zoo (Q1944813)	3	100.0% (3) #	100.0% (3) #	100.0% (3) #	0% (0) %	0% (0) %
Jardin zoologique de la ville de Lyon (Q2387471)	2	100.0% (2) #	50.0% (1) %	100.0% (2) #	0% (0) %	0% (0) %
Zoo parc de Trégomeur (Q2576057)	2	100.0% (2) #	50.0% (1) %	100.0% (2) #	0% (0) %	50.0% (1) %
Auckland Zoo (Q266235)	2	100.0% (2) #	50.0% (1) %	50.0% (1) %	0% (0) %	0% (0) %
Chiba Zoological Park (Q1148599)	1	100.0% (1) #	100.0% (1) #	100.0% (1) #	0% (0) %	0% (0) %
Turtle Back Zoo (Q795349)	1	100.0% (1) #	0% (0) %	100.0% (1) #	100.0% (1) #	100.0% (1) #
Perth Zoo (Q304383)	1	100.0% (1) #	100.0% (1) #	100.0% (1) #	0% (0) %	0% (0) %
Zoo Aquarium de Madrid (Q199480)	1	100.0% (1) #	100.0% (1) #	100.0% (1) #	0% (0) %	0% (0) %

### Item Quality Evaluator

Gets the quality scores of Wikidata items and calculates their average value.  
 How does this work?

For which Items should the average score be calculated?

Item list  SPARQL

```

1 SELECT ?director ?directorLabel WHERE {
2   ?director wdt:P31 wd:Q5.           # instance of human
3   ?director wdt:P21 wd:Q6581072.    # gender female
4   ?director wdt:P186 wd:Q2526255.  # occupation film director
5   SERVICE wikibase:label { bd:serviceParam wikibase:language "(AUTO_LANGUAGE, en)." }
6 }
7 LIMIT 100
  
```

You can also prepare your SPARQL query in the [QueryService](#) and copy it here

[Assess Item Quality](#)

WIKIMANIA  
SINGAPORE



# Current (interconnected) challenges

**WIKIMANIA  
SINGAPORE**



# Keeping up with data size

**WIKIMANIA  
SINGAPORE**

- Wikidata keeps growing
- Blazegraph has no sharding support -> larger disks and memory size required
- Internal Blazegraph limitations for number of allocators



# Keeping up with query and write load

**WIKIMANIA  
SINGAPORE**

- The query load can overload the system, leading to high response times and timeouts
- WDQS write loads are lagging behind Wikidata

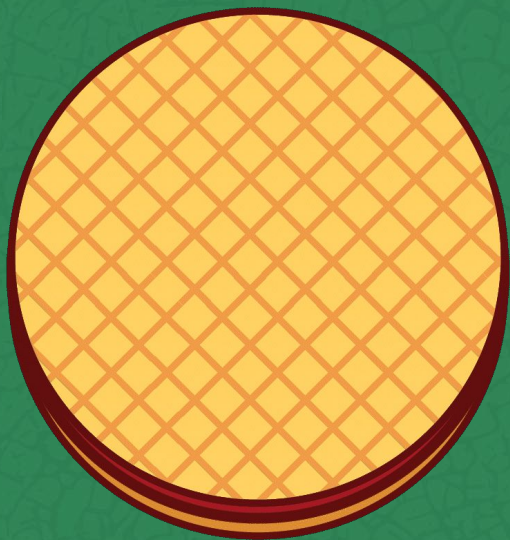


# Keeping it stable and secure

**WIKIMANIA  
SINGAPORE**

- Servers crashing leading to limited capacity and overload
- Blazegraph no longer actively developed / maintained





# Consequences

- Legitimate queries time out
- As the graph grows, queries that worked before, now no longer work
- Editors are restricting their editing work
- Editors and reusers are not getting useful new functionality

# How have we addressed the problem?

**WIKIMANIA  
SINGAPORE**





# What we've done

- Introduced a new streaming updater to cope with more edits/min
- Made a disaster mitigation plan
- Got an overview of alternative backends
- Took pressure off the system
  - Built out the Wikibase Ecosystem and especially Wikibase Cloud
  - Developed the Wikibase REST API
  - Improved documentation for the different ways to access Wikidata's data to help developers chose the right one for their usecase

# What we are doing

- Thinking through splitting off a part of the graph into a separate Blazegraph instance while keeping the data in Wikidata
- Discussing the future of the scientific article corpus in Wikidata with the WikiCite community
- Reducing redundant data by introducing a new language code for multilingual content

**And in the  
future?**

**WIKIMANIA  
SINGAPORE**



**We need to  
continue  
addressing the  
different aspects of  
the problem**

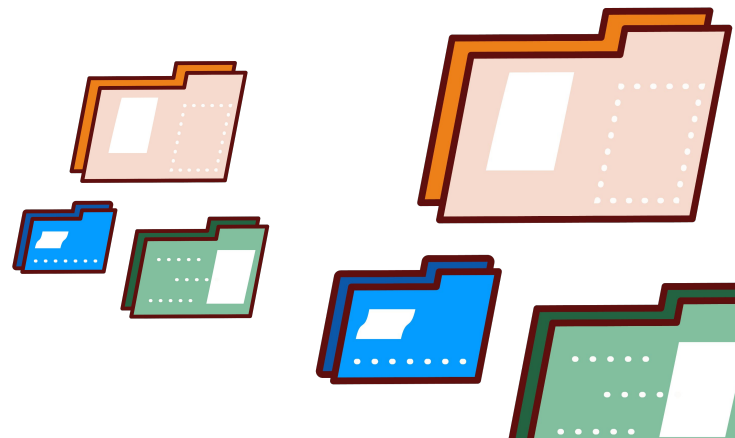
**WIKIMANIA  
SINGAPORE**

- A lot of data
- A lot of queries
- A lot of edits
- Unmaintained Blazegraph



# A lot of data

- Move large specialized data out of Wikidata (e.g. Wikibase Ecosystem)
- Continue to reduce redundant data (e.g. mul language code, automated descriptions, Lua modules for inverse relations)
- Split the graph



# A lot of queries



- **Move people to more appropriate access methods to reduce the load**
  - **Make it easier to work with other existing access methods (e.g. subset dumps, Query Service on cloud providers)**
  - **Provide additional access methods**
  - **Automatically reroute queries to more appropriate systems**
  - **Automatically rewrite inefficient queries**
  - **Increase the incentives and pressure to move to other access methods**

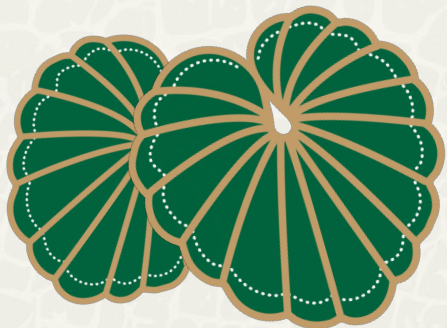
# A lot of edits

- Reduce redundant data and the unnecessary edits that come with them



**WIKIMANIA  
SINGAPORE**

# Unmaintained Blazegraph



**WIKIMANIA  
SINGAPORE**

- Evaluate alternatives and move away from Blazegraph
- Evangelize for the development of new graph backends





# More questions? Want to stay up-to-date?

[Wikidata's weekly newsletter](#)

Search Platform team office hours

lydia.pintscher@wikimedia.de

@nightrose

[User:Lydia Pintscher \(WMDE\)](#)

**WIKIMANIA  
SINGAPORE**





**WIKIMANIA  
SINGAPORE**