

WIKIDATA

i/o



Repository Fringe 2017
at the University of Edinburgh
3-4 August 2017.

Wikidata

The free and open knowledge base



Part 1: How to add data to Wikidata

Repo Fringe 2017

Ewan McAndrew - @emcandre

Navino Evans - @NavinoEvans

<https://tinyurl.com/WikidataRepo1>

Please create an account on Wikidata (You can use your Wikipedia login if you already have one).

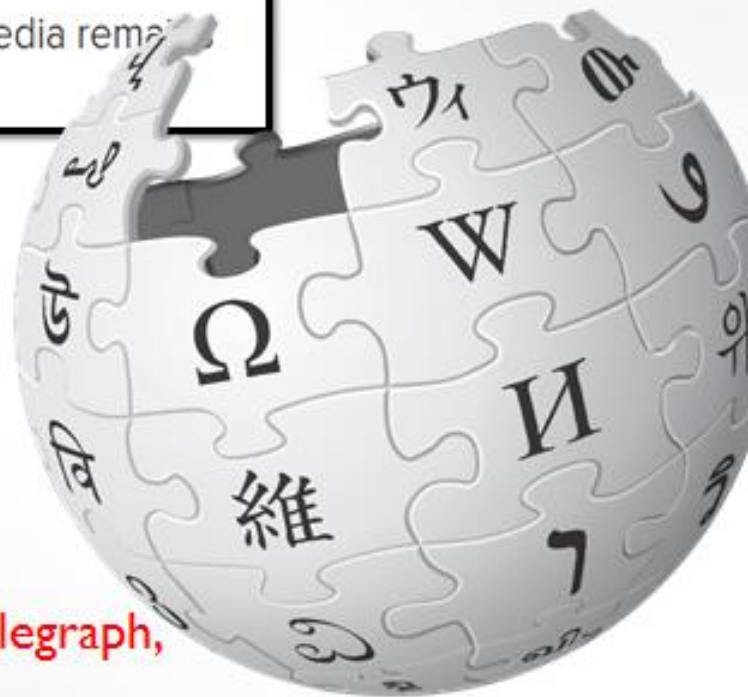
The image shows a screenshot of the Wikidata homepage. At the top right, there are navigation links: "English", "Not logged in", "Talk", "Contributions", "Create account", and "Login". Below these is a search bar with a magnifying glass icon. A large blue arrow points from the right side of the page towards the search bar. The main content area features a central "Welcome to Wikidata" message, stating it is "the free knowledge base with 15,511,314 data items that anyone can edit". Below this message are three main sections: "Welcome!", "Learn about data", and "Get involved". The "Learn about data" section includes three images: "New Earth view", "properly aligned paint cans", and "cotton vials, Mount Everest sign". The left sidebar contains various navigation links such as "Main page", "Community portal", "Project chat", "Create a new item", "Item by title", "Recent changes", "Random item", "Query Service", "Nearby", "Help", "Donate", "Preferences", "Create a book", "Download as PDF", "Portable version", "In other projects", "Wikimedia Commons", "Meta-Wiki", "Wikispecies", "Wikisource", "Wikivoyage", "Wikidata", "Wikinews", "Wikiquote", "Wikisource", "Wikiversity", and "Wikivoyage". At the bottom, there are "Tools" links: "What links here", "Recent changes", "Special pages", "Permanent link", and "Page information".

<https://www.wikidata.org>

The Internet's Favorite Website

As web traffic shifts toward mobile, a new study finds Wikipedia remains the most popular informational site around.

- 80,000 regular contributors (of which only 3000 or so are considered 'very active' Wikipedians).
- 500 million visitors per month
- 1.5 billion monthly unique devices
- The fifth most popular website in the world.
- Trusted more than the BBC, ITV, the Times, the Telegraph, The Guardian and more according to Yougov survey (2014).



**DATA IS THE
NEW BACON**

What is Wikidata?

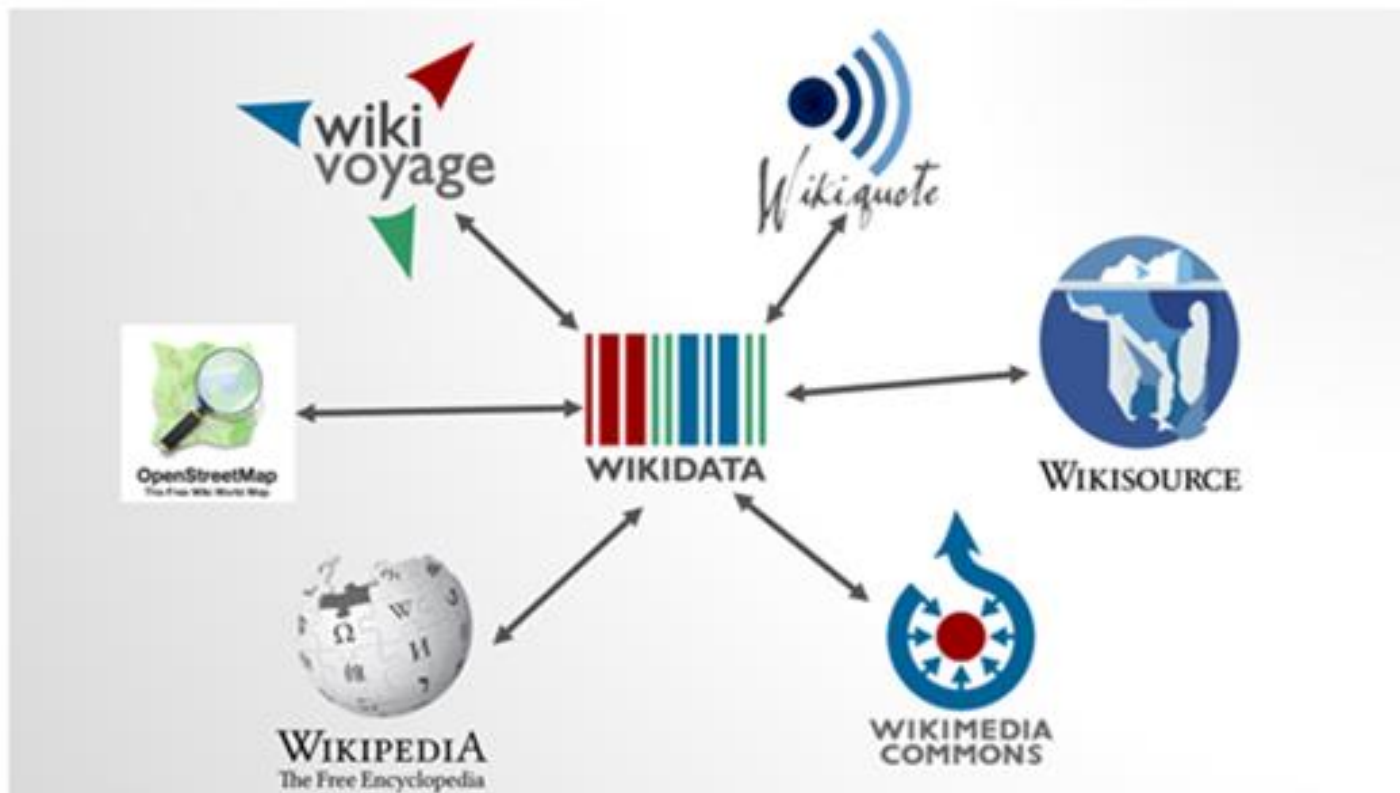


Wikidata is a free linked database of secondary data that can be read and edited by both humans and machines.

Wikidata acts as central storage for the structured data of its Wikimedia sister projects including Wikipedia, Wikivoyage, Wikisource, and others.

- Bibliographic
- Biographic
- Biomedical
- Geographic
- Taxonomic
- Authority file

A centralised hub of structured data for all the Wikimedia projects



What is Wikidata actually?

- repository of the world's knowledge
- database anyone can read and edit
- multi-lingual
- designed to deal with the reality Wikipedia has to deal with
- free and open source software.
- All data on Wikidata is CC-0 licenced.

English Wikipedia “only includes only includes 30% of the items entered in the other 287 languages.”*

Language	Language (local)	Wiki	Articles	Total	Edits	Admins	Users	Active Users	Images	Depth
English	English	en	5,205,057	39,859,108	841,187,469	1,297	28,749,525	112,301	851,064	935.45
Swedish	svenska	sv	3,361,100	6,752,452	36,314,210	71	500,503	2,223	0	5.48
Cebuano	Cebuano	ceb	2,668,059	4,872,925	9,794,491	5	28,432	107	66	1.37
German	Deutsch	de	1,958,190	5,667,148	161,040,878	236	2,446,855	18,341	131,917	101.94
Dutch	Nederlands	nl	1,874,463	3,633,438	48,119,572	49	778,155	3,901	19	11.66
French	français	fr	1,773,804	8,183,404	130,180,351	167	2,558,400	14,307	40,111	207.00

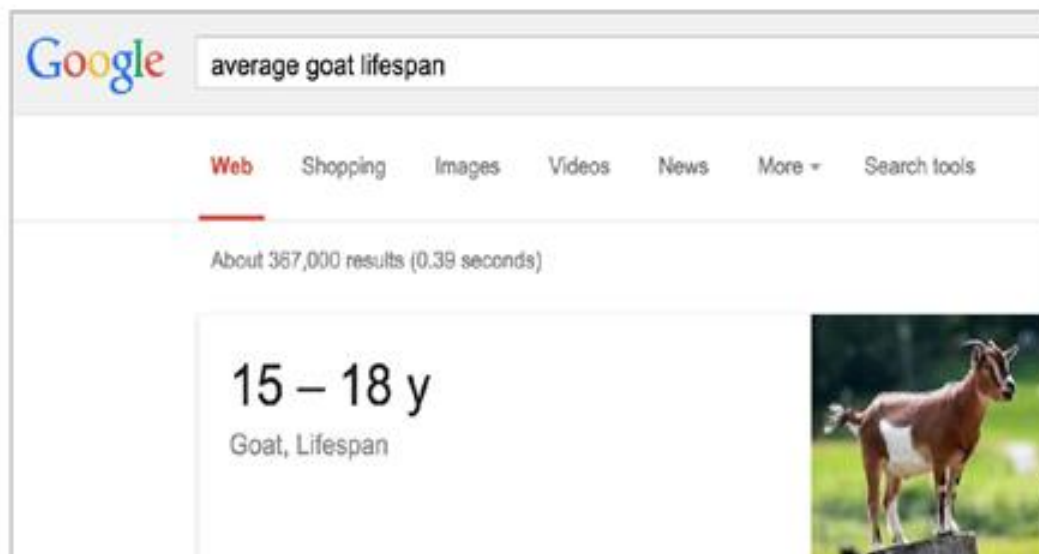
Grand total [edit | edit source]

Articles	Total	Edits	Admins	Users	Active Users	Images
40,742,921	151,823,044	2,214,808,661	3,998	62,976,506	263,618	2,410,128

Vietnamese	Tiếng Việt	vi	1,146,809	3,211,773	24,059,192	23	501,251	1,400	21,222	24.29
Japanese	日本語	ja	1,022,325	2,966,811	61,403,824	49	1,109,224	11,787	83,431	74.88
Portuguese	português	pt	928,143	4,135,001	47,141,295	70	1,793,088	6,793	38,558	136.1

“The promise of linked open data seems to have finally arrived.”

Okay Google..... what's the average lifespan of a goat?



The screenshot shows a Google search interface. At the top left is the Google logo. To its right is a search bar containing the text "average goat lifespan". Below the search bar are navigation tabs for "Web", "Shopping", "Images", "Videos", "News", "More", and "Search tools". Underneath these tabs, it says "About 367,000 results (0.39 seconds)". The main result is a knowledge panel with the text "15 – 18 y" in large font, and "Goat, Lifespan" in smaller font below it. To the right of the text is a photograph of a brown and white goat standing on a rock in a grassy field.



“The primary issue with Google’s knowledge panels is that they aren’t terribly knowledgeable: They provide information but often leave out any context on where that information came....”

Caitlin Dewey, The Washington Post – 11 May 2016



English Not logged in

Item Discussion Read View history Search Wikidata

Wikidata

dog breed

[In more languages](#) [Configure](#)

Language	Label	Description	Also known as
English	Skye Terrier	dog breed	
French	Skye terrier	race de chiens	
German	Skye Terrier	britische Hunderrasse	
Scots	No label defined	No description defined	

[All entered languages](#)

Statements

instance of	dog breed	edit
	+ 1 reference	

Greyfriar's Bobby (1855-1872)
Sat by his master's grave for 14 years.
Died aged 16.

Kennel club say Skye Terrier's have a life expectancy of "over 12 years".

External source:

▼ Species data

- The BayScience Foundation, Inc. *ZipcodeZoo.com*. »
- Cornell Lab of Ornithology. *All about Birds*. »
- Flindt, R. *Amazing Numbers in Biology*. Springer, 2006.
- National Geographic Society. "Animals." *National Geographic*. »
- Species 2000 and Integrated Taxonomic Information System. *Catalogue of Life*. »

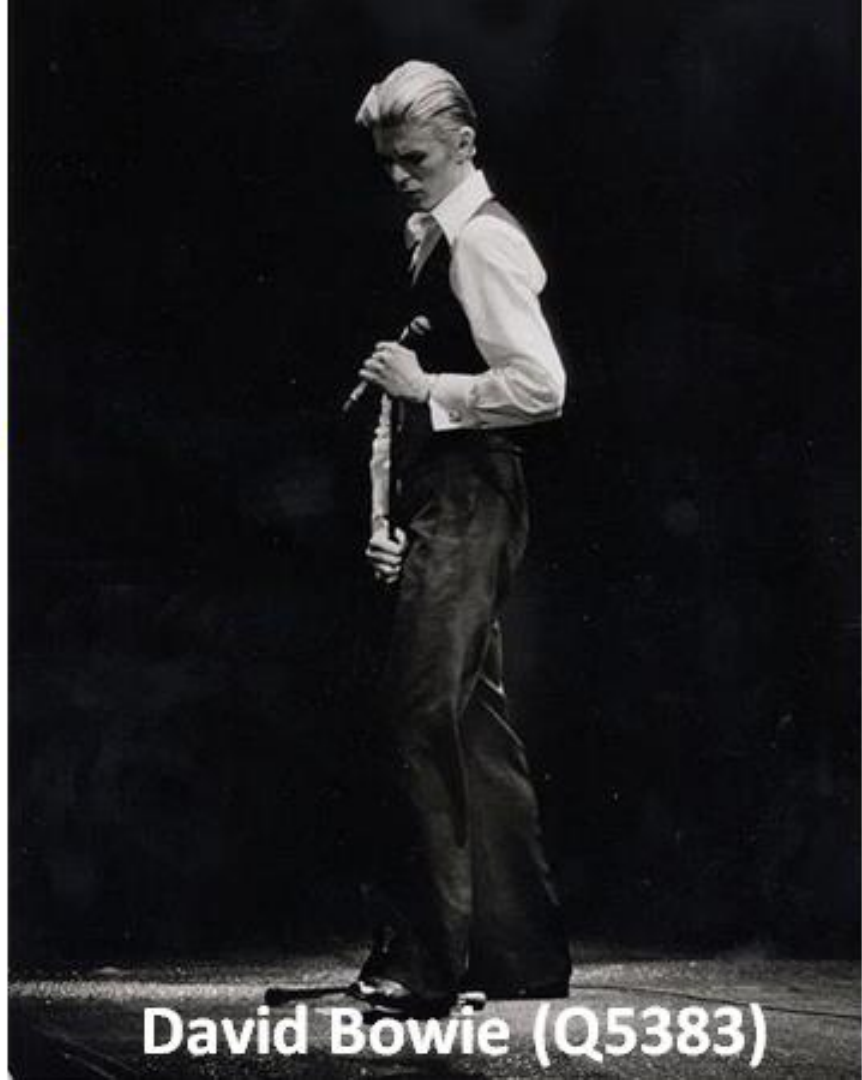
▼ Note

This list is intended as a guide to sources of further information. The inclusion of an item in this list does not necessarily mean that its content was used as the basis for any specific Wolfram|Alpha result.

Requests by researchers for detailed information on the sources for individual Wolfram|Alpha results can be directed [here](#). Requests will be handled on the basis of priority, importance and available resources.

date of birth (P569)

8 January 1947



David Bowie (Q5383)

Sweden (Q34)

**population
(P1082)**

9,747,355

according to: Statistics Sweden

method: estimation

point in time: 31 Dec 2014



picture by Hector Melo A.

A provenance-preserving Answer Engine



The sum of all **human knowledge**

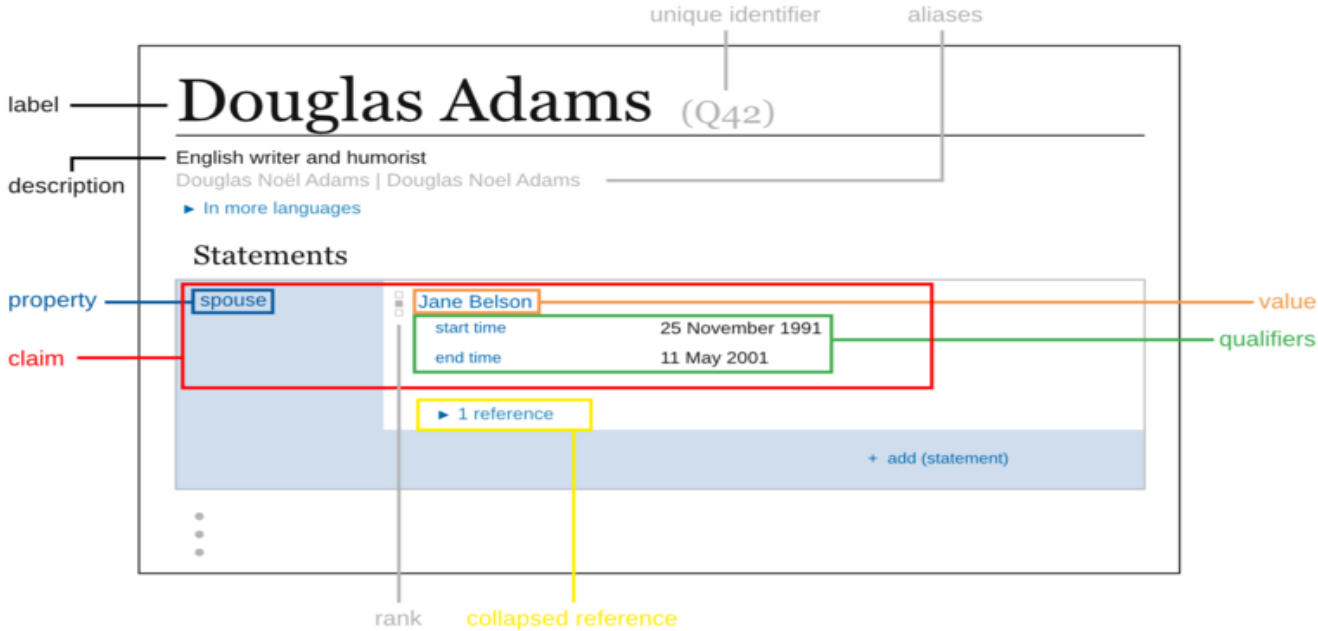
+



WIKIDATA

The sum of all **data and sources** backing human knowledge

Example Wikidata item & statement



wikidata.org/wiki/Q42

Douglas Adams

Wikidata key stats – data and community

- 30.2 million items
- 160 million statements
- 530 million edits
- 18,695 active editors
- 11 million items added since August 2016.
- 2000+ more active editors since August 2016.



“Data is beautiful. Data is information.”

[Official Wikidata stats](#)

[More stats](#)

SECTION

MANUAL EDITING

Practical session - Adding data



- Open you selected batch from the [batches spreadsheet](#)
- Show how to add '*award received*' (P166) statement + reference.
- Show how to create an item from scratch
 - 'Instance of' = 'human'
 - 'Gender' = 'female'
 - 'Award received' = 'Fellowship of the Royal Society c'
 - **Qualifier:** Point in time = year of election
 - **Reference:** Reference URL = Url of page on RSE website



Extension exercise: Adding referencing and additional data

To add a reference:

1. Click 'Add reference', Type 'reference URL' as the Property for the reference field.
2. Copy the URL web address from the Batch spreadsheet and paste into the value field for the reference in Wikidata. Click Save.

EXTENSION (IF TIME):

Add member (P463) of Royal Society of Edinburgh (Q117467) if it is not there as a statement.
Add [Date of birth \(P569\)](#) data if it can be found.
Add [Place of birth \(P19\)](#) data if it can be found.
Add references to backup claims.



[CC-BY Stewart Butterfield via Flickr](#)

SECTION

MASS EDITING

Essential tools for mass editing Wikidata

[Quickstatements v.2](#)

For importing data from a spreadsheet into Wikidata. The syntax you need to use is explained in [QuickStatements v.1](#)

[Wikipedia and Wikidata Tools for Google sheets](#) ([Demo](#))

Google sheets add-on for pulling data from Wikidata and Wikipedia directly into a spreadsheet (Note: you need a Google account to install this)

Practical - mass editing using QuickStatements

1. Go to the [batches spreadsheet](#), then **click the link with your selected batch number**
2. **Select all cells highlighted orange** (the QuickStatements commands), then **copy them to your clipboard** (click edit then copy)
3. Go to [QuickStatements](#) and click
4. Click 'Import commands' -> 'Version 1 format'
5. **Paste in the commands copied in step 2, then click 'import'**
6. **Check a selection of the commands** to make sure they have imported correctly
7. **Click the "RUN" button** at the bottom to launch your first mass edit!

Demo results

Using the [Wikidata Query Service](#)

- [Bubble chart - countries with the most UNESCO endangered languages](#)

Using [Listeria](#) to generate a Wikipedia list

- [List of UNESCO endangered languages](#)
- [List of female fellows of the Royal Society of Edinburgh](#)

End of practical - let's see the improved results!

Map query - <http://tinyurl.com/RSEmap>

Link straight to map - <http://bit.ly/RSEmap2>

Timeline - <http://tinyurl.com/RSEtimeline>

Wikidata query itself - <http://bit.ly/RSEquery>

Listeria list - <http://tinyurl.com/RSEListeria>



Links and further reading

- [https://www.wikidata.org/wiki/Wikidata:Data Import Guide](https://www.wikidata.org/wiki/Wikidata:Data_Import_Guide)
- [https://www.wikidata.org/wiki/Wikidata:Database download](https://www.wikidata.org/wiki/Wikidata:Database_download)
- <https://www.wikidata.org/wiki/Property:P2966> – National Library of Wales ID for collection items.
- <https://www.wikidata.org/w/api.php>
- <https://tools.wmflabs.org/reasonator/> - Wikidata made 'pretty'
- <https://tools.wmflabs.org/reasonator/?q=Q42> – Reasonator page for Douglas Adams (Q42) by way of example.
- <https://www.mediawiki.org/wiki/Extension:ArticlePlaceholder> – another way of placeholdering articles using structured data from wikidata to populate information in the meantime until an article can be created.
- [Wikidata: Current trends and priorities](#) (May 2017 presentation with current

Developer links

#wikidata on chat.freenode.net

wikidata-l@lists.wikimedia.org

[Wikidata – The New Rosetta Stone](#) (article).

[Google closes Freebase](#) (article).

[Google's sketchy attempt to control the world's knowledge](#) (article).

api @ wikidata.org/w/api.php

sandbox @ wikidata.org/wiki/Special:ApiSandbox

The Wikidata Game: <https://tools.wmflabs.org/wikidata-game/distributed/>

PHP

Wikibase API Library: github.com/addwiki/wikibase-api

SPARQL abstraction: github.com/Benestar/asparagus

Python

Wiki bot Framework: mediawiki.org/wiki/Manual:Pywikibot/Wikidata

C# .NET

Wikibase API Library: github.com/Benestar/wikibase.net