

# Beyond the basics: Intermediate Wikidata techniques

**Asaf Bartov** [asaf@wikimedia.org](mailto:asaf@wikimedia.org)  
[@abartov](https://twitter.com/abartov)



**WIKIMEDIA**  
FOUNDATION

Slides: [http://bit.ly/wikidata\\_tools](http://bit.ly/wikidata_tools)

Jan 2020, Porto

# 7-minute catch-up



**WIKIMEDIA**  
FOUNDATION

# What is Wikidata?

- A central editable repository of **structured and linked data**, on a wiki, under a free license (CC0)
  - Pure free software; pure free data
  - Open and editable by anyone, including the ontology
- A sister project of Wikipedia, by Wikimedia
- Increasingly, where the cool kids play!

# Structured and linked data?

- Structured data: every piece of information is described as a **triple** -- subject, predicate, object, or as Wikidata says: **item, property, value**.
  - Each such triple can itself be further described/qualified with a series of triples (e.g. time range, context, aspect, provenance)
- Data statements **link** to each other, enabling **discovery and reasoning**.



# Item: Earth (Q2)

instance of

property

terrestrial planet

0 references

value

# Item: Earth (Q2)

start date

4540 million years ago

[edit]

determination method

lead-lead dating

subject of

age of the Earth

Qualifiers

▼ 1 reference

stated in

The age of the Earth in the  
twentieth century: a problem  
(mostly) solved

[edit]

References

[add reference]

[add]







# **(Just) A handful of Wikidata uses**

1. Describe data with flexible level of detail or abstraction
2. Centralize data and avoid duplication of effort
3. Avoid multiple copies falling out of sync
4. Connect disparate IDs and align datasets
5. Serve as discovery platform for data elsewhere
6. Massively multilingual aligned vocabularies
7. Lateral querying across arbitrary vectors



**Hmm... got examples?**

**Who is the most famous  
Portuguese in the world?**

# Famous Portuguese

by # of languages in which they are covered on Wikipedia

person	personLabel	personDescription	sitelinks
<a href="#">Q1496</a>	Ferdinand Magellan	Portuguese explorer in the service of Spain	160
<a href="#">Q7328</a>	Vasco da Gama	Portuguese explorer	152
<a href="#">Q11571</a>	Cristiano Ronaldo	Portuguese association football player	150
<a href="#">Q37060</a>	José Saramago	Portuguese writer and recipient of the 1998 Nobel Prize in Literature	120
<a href="#">Q311440</a>	António Guterres	Secretary-General of the United Nations	106
<a href="#">Q79983</a>	José Mourinho	Portuguese association football player and manager	97
<a href="#">Q590</a>	Luís de Camões	Portuguese poet	96
<a href="#">Q173481</a>	Fernando Pessoa	Portuguese poet, writer, literary critic, translator, publisher and philosopher	94

Wikidata Query = ❤️ Run this query yourself: <https://w.wiki/Fe7>

# José Saramago

...by any other ID...

## Identifiers

VIAF ID	114473675	edit	✓
	▶ 2 references		
		+ add value	

Library of Congress authority ID	n85130797	edit	✓
	▶ 2 references		
		+ add value	

GND ID	118964038	edit	✓
	▶ 2 references		
		+ add value	

# José Saramago

...by any other ID...

Open Library ID

OL3112635A

edit

0 references

+ add reference

+ add value

FAST ID

166806

edit

0 references

+ add reference

+ add value

Biblioteca Nacional de España ID

XX998466

edit

0 references

+ add reference

+ add value

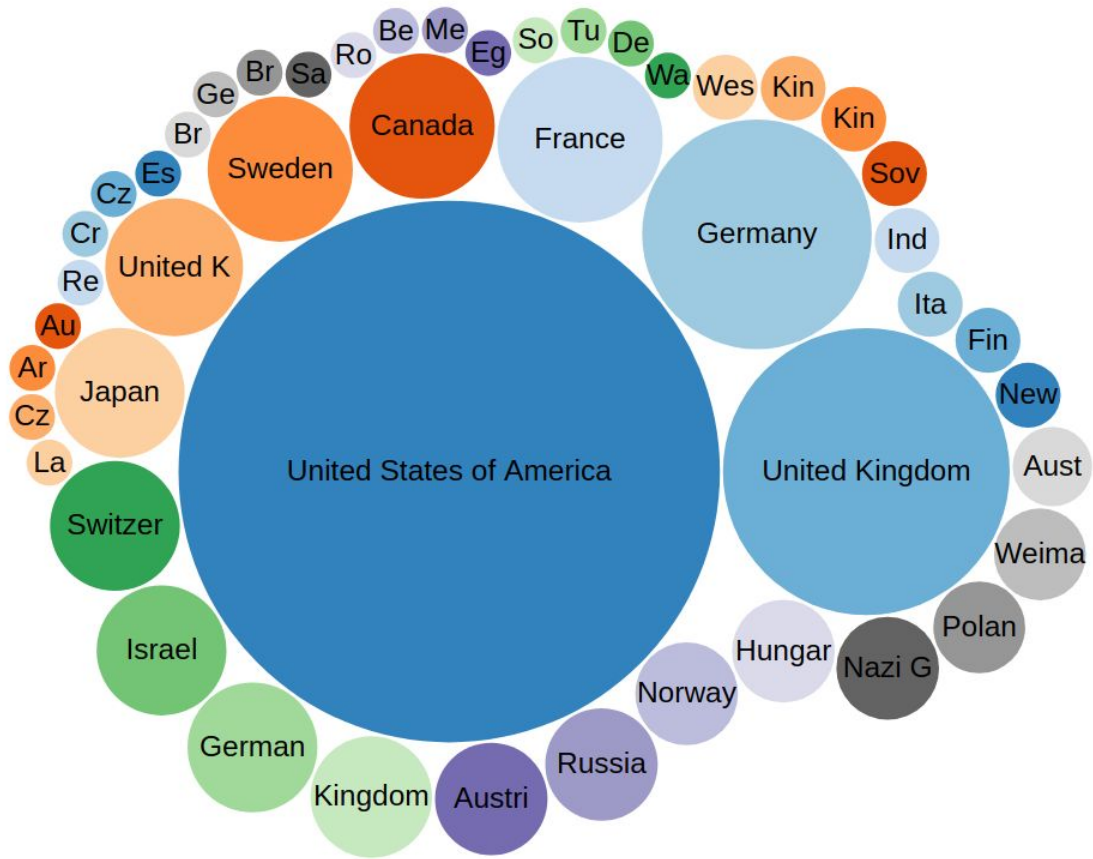
ISFDB author ID

4793

edit

...and 66 other identifiers!

# Nobel prizes by country

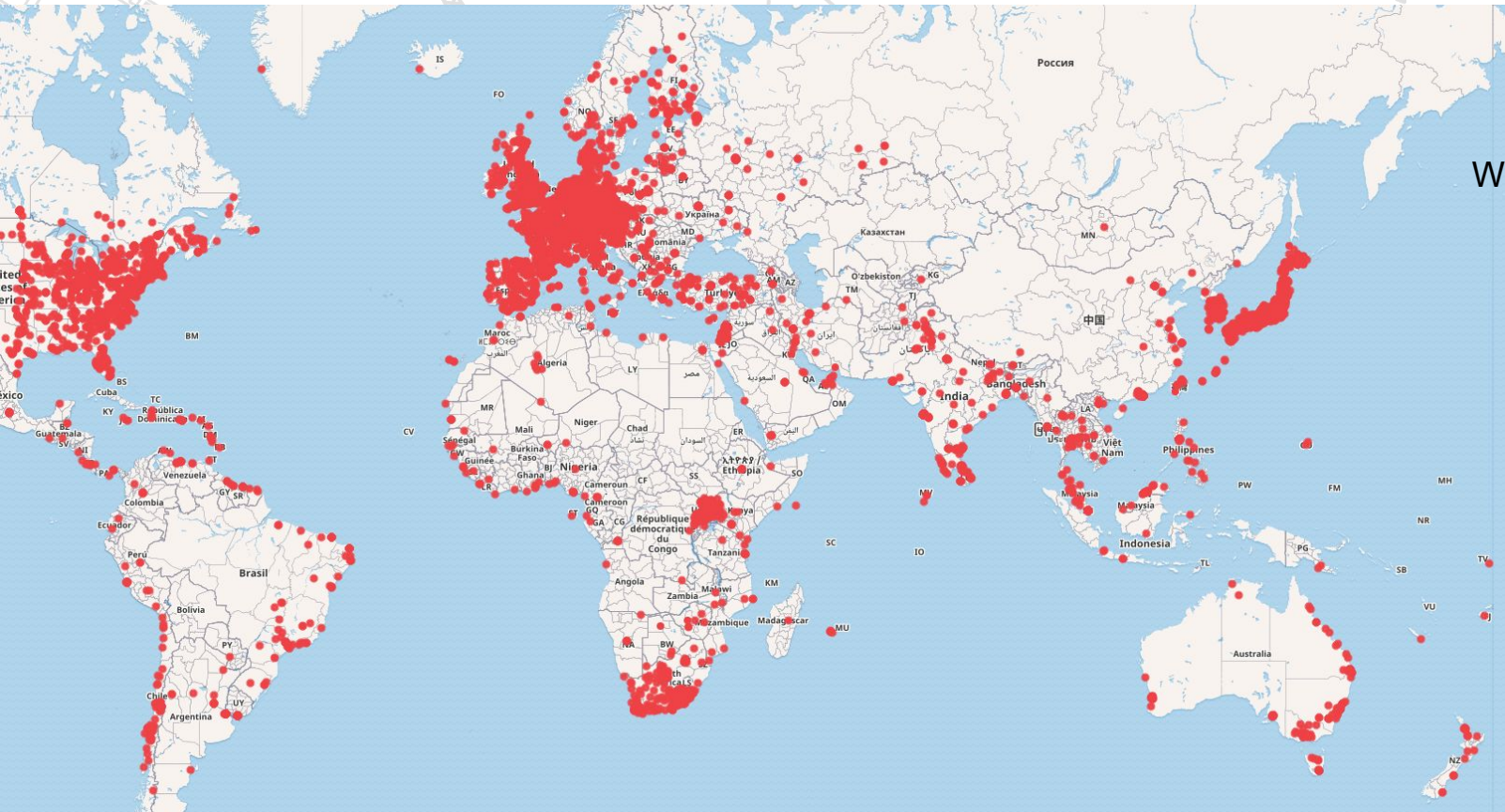


Wikidata Query = ❤️  
- run this query yourself:  
<https://w.wiki/9sZ>



# Hospitals in the world

SELECT \* WHERE { ?item wdt:P31\*/wdt:P279\* wd:Q16917; wdt:P625 ?geo . }



Wikidata Query = ❤️

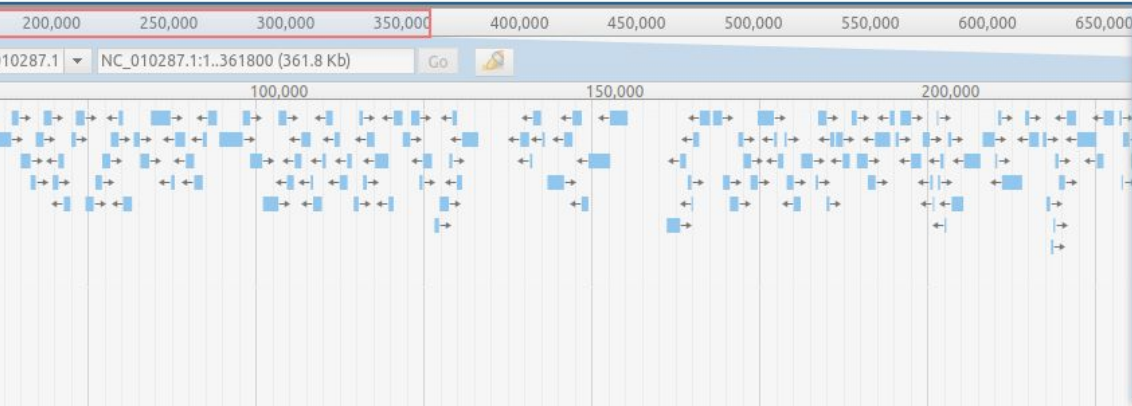
**Mind  
the  
gap!**

# Amazing things built on top of Wikidata

# WikiGenomes.org

Gene Search:

72




**CTL0096**

### Attributes

<b>description</b>	metal transporting ATPase CTL0096
<b>entrezGeneID</b>	5858085
<b>refSeq</b>	NC_010287.1
<b>uri</b>	<a href="http://www.wikidata.org/entity/Q21171887">http://www.wikidata.org/entity/Q21171887</a>

### Region sequence

 FASTA

```
>NC_010287.1 NC_010287.1:117833..119811 (- strand)
length=1979
TGTC CCGCAACTTTTTCTCCTCTCCATTTTCTCGAGAATTACTCAGCGATTTTTTTGAA
TCAGGGATGGCGGAAGAGAATAGCCCTCTGCTCTC CAAAAAACCGTCGTTTAAGCCA
AAATTTAACCTTAAATCTGCCTATATCTCCTTGGCTTTATACTTAGGATCGCTTCTTT
CTCATTGGATAGGATATCAGGCTCTCTCTCTTTACTTCTCAITTCATTTTTTTCTTA
```

# Crotos art browser



Crotos

Cosmos  
Callisto  
Lab

139 Results

- Everything
- Everything
- Painting
- Sculpture
- Drawing
- Print
- Photograph
- Mosaic
- Installation
- Handscroll
- Tapestry
- Stained glass

Creator : **Michelangelo** ✓

-40000

1 2 3 4 5 ... 7



Pitti Tondo  
Michelangelo - Bargello



Young Slave  
Michelangelo - Galleria dell'Accademia



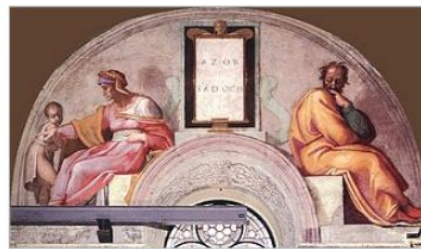
The Drunkenness of Noah  
Michelangelo - Sistine Chapel



Atlas Slave  
Michelangelo - Galleria dell'Accademia



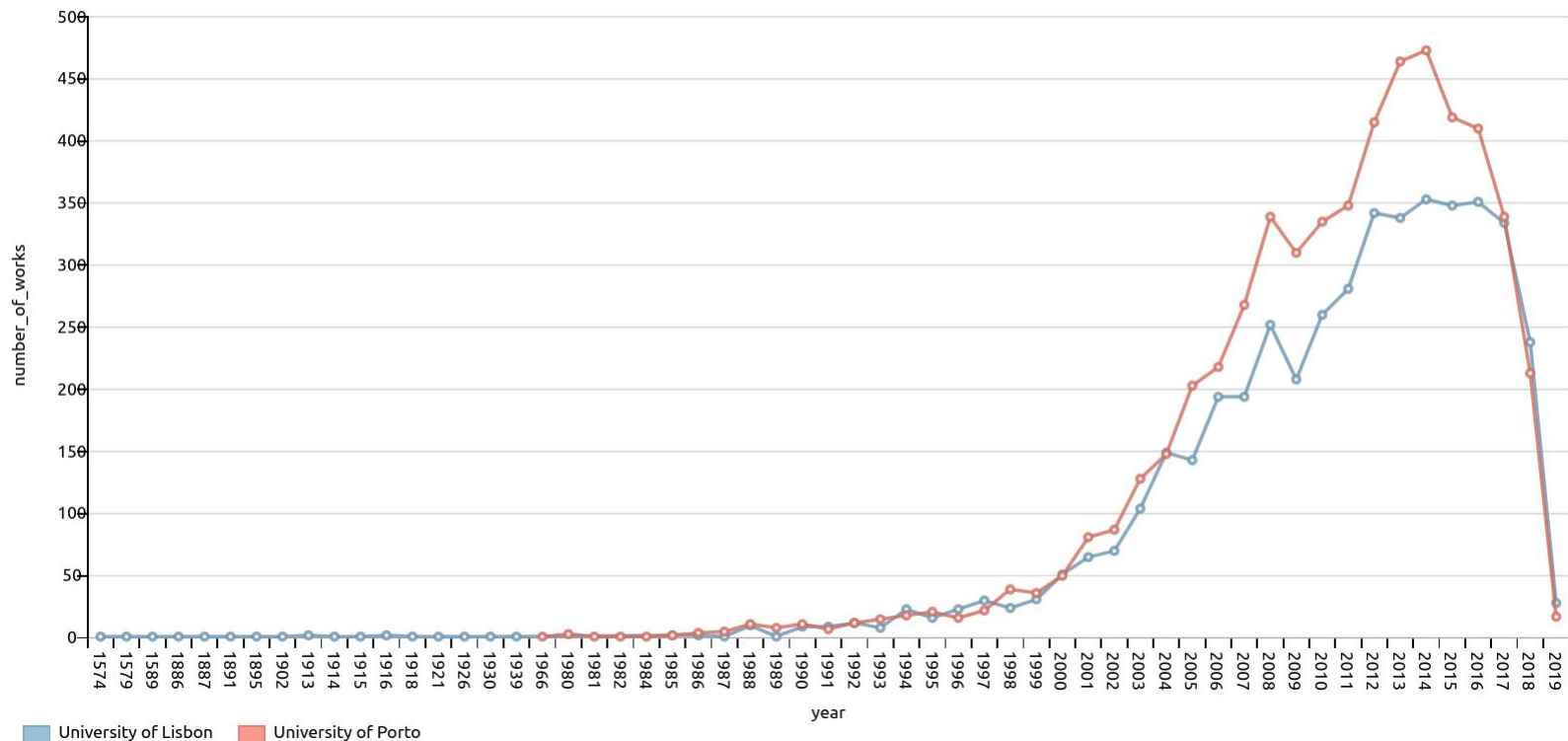
New Sacristy  
Michelangelo - Basilica of San Lorenzo



# Scholia

Works per year

Explore and  
compare  
scholars,  
journals,  
institutions



# Scholia

Number of works

Topic

Organization

**Most  
common  
topics  
written  
about, by  
scholars at  
each  
university**

240	Portugal	University of Porto
195	Portugal	University of Lisbon
70	amyotrophic lateral sclerosis	University of Lisbon
56	malaria	University of Lisbon
51	aquatic science	University of Lisbon
50	mitochondrion	University of Porto
47	nanoparticle	University of Porto
40	phylogenetics	University of Porto
39	Mozambique	University of Porto

# Quick survey

1. What work are you already doing with Wikidata, where?
2. What particular scenarios do you need more skills or help with?
3. Any particular tools or techniques you'd like to learn today?



**WIKIMEDIA**  
FOUNDATION

# Querying and exploring Wikidata



**WIKIMEDIA**  
FOUNDATION



# Wikipedia/Wikidata Tools for Google Sheets



**WIKIMEDIA**  
FOUNDATION

# Why:

To align tabular data against Wikidata or Wikipedia; to get automated reports without coding (Wikidata, Quarry, Pageviews!)

# How:

Open Google Sheets, from the "Add-ons" menu, select "Get add-ons...", and search for "Wikipedia Tools"



**Example:**

[http://bit.ly/wdsheet\\_pt](http://bit.ly/wdsheet_pt)

# Useful maintenance and to-do queries



**WIKIMEDIA**  
FOUNDATION

# Why:

[Wikidata Query](#) can shed light on content gaps, visualize them, suggest articles for editathons, and much more

# How:

Watch [a video tutorial](#); ~~steal~~ adapt [other people's queries](#)

# Examples:

people born in Portugal without ENWP article

<https://w.wiki/FeA>

best-described items in National Museum

Soares dos Reis:

<https://w.wiki/FeC>

alumni of University of Porto:

<https://w.wiki/FeE>



WIKIMEDIA  
FOUNDATION

# Wikimedia Cultural Diversity Observatory



**WIKIMEDIA**  
FOUNDATION

# Why:

[Wikimedia Cultural Diversity Observatory \(WCDO\)](#) can shed light on content gaps, visualize them, suggest lists for editathons, ...

# How:

Explore the interface above. @marcmiquel can help.



# Examples:

Top Portuguese cultural content missing from Spanish Wikipedia:

<http://bit.ly/2TtbDMZ>

Spread of cultural content across Wikipedias:

[https://wcdo.wmflabs.org/ccc\\_spread/](https://wcdo.wmflabs.org/ccc_spread/)

Coverage of other cultures in given Wikipedia:

[https://wcdo.wmflabs.org/ccc\\_coverage/](https://wcdo.wmflabs.org/ccc_coverage/)

# {{Wikidata list}} ("Listeria")



WIKIMEDIA  
FOUNDATION

# Why:

Regularly-updated Wikidata-generated lists are useful to assess the status of content on Wikidata, and as to-do lists (redlink lists) for creating articles.

# How:

Read documentation in {{Wikidata list}};  
~~steal~~adapt [other people's lists](#)

# Examples:

Women in Red's redlists:

[https://en.wikipedia.org/wiki/Wikipedia:WikiProject\\_Women\\_in\\_Red/Redlist\\_index](https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Women_in_Red/Redlist_index)

# Contributing to Wikidata



**WIKIMEDIA**  
FOUNDATION

# Wikidata Editing Framework



**WIKIMEDIA**  
FOUNDATION

# Why:

Enrich Wikidata items based on Wikipedia articles, right from Wikipedia!

# How:

to install, add to [\[\[m:Special:My page/global.js\]\]](#) the following:

```
if ( mw.config.get("wgDBname") !== 'ruwiki' )
    mediaWiki.loader.load(
'//ru.wikipedia.org/w/index.php?title=MediaWiki:WEF_AllEditors.js&action=r
aw&ctype=text/javascript' );
```

# Examples:

Let's look at [\[\[Artur Ivens Ferraz\]\]](#) ([Q16519927](#))



# Cradle



**WIKIMEDIA**  
FOUNDATION

# Why:

Create easy input forms for creating new items, with only the relevant fields.

# How:

<https://tools.wmflabs.org/wikidata-todo/cradle/#/>

# Example:

Let's look at the template for "soldier"

[https://tools.wmflabs.org/wikidata-todo/cradle/  
#/subject/soldier](https://tools.wmflabs.org/wikidata-todo/cradle/#/subject/soldier)

# Tabernacle



**WIKIMEDIA**  
FOUNDATION

# Why:

Edit and improve Wikidata content for a given set of items (from a query) in a table, without clicking in and out of items.

# How:

Make a query (or a PagePile) for the set of items you want to edit, feed it to [Tabernacle](#), choose the columns you want (label/desc, properties), and go.

# Examples:

Notable Portuguese without an English label

Human settlements in Portugal

# Mix 'n' Match



**WIKIMEDIA**  
FOUNDATION

# Why:

Align an external dataset with Wikidata;  
discover missing items in Wikidata;  
INTERCONNECT ALL THINGS

# How:

Go to [Mix'n'match](#), select a dataset ("catalog"), and either review automatic matches or make manual matches.

To add a dataset, consider first proposing a new Wikidata property for IDs from your data set, then either import by pasting tab-delimited (from a spreadsheet?) data [here](#), or [configure a scraper](#) here.



# Examples:

Heritage of Portuguese Influence

Enciclopédia Itaú Cultural ID

# WikiShootMe



**WIKIMEDIA**  
FOUNDATION

# Why:

Discover missing photographic coverage by location. Plan photo trips. Upload on the go.

# How:

Go to [WikiShootMe](https://wikishootme.org), enable location on your device or search for a location. Red dots need photos!



**Example:**

[WikiShootMe](#)

# Batch editing



**WIKIMEDIA**  
FOUNDATION

# QuickStatements



**WIKIMEDIA**  
FOUNDATION

# Why:

Make large numbers of edits to Wikidata based on tabular data. Largest scale tool for contributing without coding.

# How:

Prepare your data **just so** (this is the time-consuming part), then paste or upload to [QuickStatements](#), run it, and have an espresso.

# QuickStatements format v1

Each statement *must* consist of an **entity**, a **property**, and a **value**.

An **entity** can be an **item**, a **property**, a **lexeme**, a **form**, or a **sense**.

A **value** can be another **entity**, a **string**, a **time**, a **location**, or a **quantity**, depending on the property type, or one of the special values *somevalue* or *novalue*.

Ex:

Qxxx <tab> P31 <tab> Q5 (= add "instance-of human" to Qxxx)

To add labels/descs, specify an **entity**, L<lang> or D<lang>, and a **string**.

Ex:

Q179959 <tab> Lmi <tab> "kākāpō"

Much more at: <https://www.wikidata.org/wiki/Help:QuickStatements>



# QuickStatements format v2

Paste a CSV (not TSV!). Each line *must* consist of an entity, and any number of values, for operations specified by columns at the header row.

Ex:

qid, P31

Qxxx, Q5 (= add "instance-of human" to Qxxx)

To add labels/descs, specify an entity, L<lang> or D<lang>, and a string.

Ex:

qid, Lmi, Dmi

Q179959, "kākāpō", "manu o Aotearoa"

Much more at: <https://www.wikidata.org/wiki/Help:QuickStatements>

# How do I get lists of QIDs?

- (Remember to prep for QS on a spreadsheet)
- **Wikidata Query**
- **Wikipedia Tools for Sheets**
- **PetScan**
- **Tabernacle**

# Bonus material



**WIKIMEDIA**  
FOUNDATION

# Contributing to Wikipedia using Wikidata



**WIKIMEDIA**  
FOUNDATION

**What if we could "seed" a  
Wikipedia article with an  
appropriate section  
template and some data  
from Wikidata?**

**{{Mbabel|<qid>}}**

example

- **SQID** (ex. literary work)
- **Wikidata Graph Builder**
- **SPARQL-rc**

**In conclusion...**



**WIKIMEDIA**  
FOUNDATION



**Remember:**

**Use your power for  
good and not for evil**

**Thank you for listening**

**(share this!)**

**asaf@wikimedia.org**

**@abartov**



**WIKIMEDIA**  
FOUNDATION