A splash of Wikidata
WIKIMEDIA
FOUNDATION

I work at the Wikimedia foundation a non-profit, which supports the Wikimedia movement
What is the wikimedia movement you ask?

A JOKE
The movement is a like the hydra, a multi-headed beast of tooth and fang
A JOKE
Ok, that sounds a little too scary for an introduction. It’s actually...

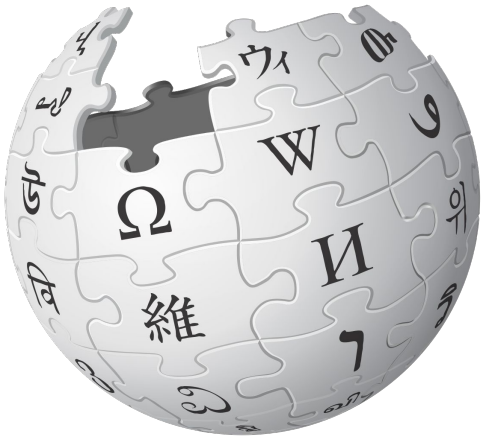
Very nice people! But they are a “beast” in what they’ve been able to accomplish.

Here is a group of us in Cape Town South Africa at the yearly gathering of Wikimedians (people who volunteer time in support of the movement) called Wikimania.

We sure do love the word wiki.

Not just Wikipedia

Wiktionary
Wikiquote
Wikibooks
Wikisource
Wikinews
Wikiversity
Wikispecies
Wikivoyage
Wikidata
Wikimedia Commons
MediaWiki



This includes projects like Wikipedia, but not just Wikipedia - the most popular project in the movement.

We support all these projects under the mission - “[To] empower and engage people around the world to collect and develop educational content under a free license or in the public domain, and to disseminate it effectively and globally.”

All free, all collaborative. From a free encyclopedia to a free dictionary, to travel guide, and media repository
And others services and projects like our cloud services for open-source projects

To be clear; free as free to access, free to reuse - even commercially, free to contribute, etc. Not free and we’ll collect data on your habits, not free and we’ll shovel intrusive ads at you, not free but you are restricted in how you use it. Truly, genuinely, free.

How we support

Advancement
Product
Technology
Community
Engagement
Legal

WIKIMEDIA
FOUNDATION

We provide support for these projects and communities through...
Advancement - Fundraising and partnerships
Product - user facing software - contributors and readers
Technology - core infrastructure, APIs, etc
Community - support communities around the world. Chapters (incorporated independent non-profit, some with their own staff and programs), thematic organizations, and user groups.
We help with grant-making, training, event planning etc.
Legal - advocacy and defending communities
Comms, finance, Talent and culture

https://wikimediafoundation.org
https://meta.wikimedia.org/wiki/Wikimedia_movement_affiliates

Obligatory cat gif

WIKIDATA

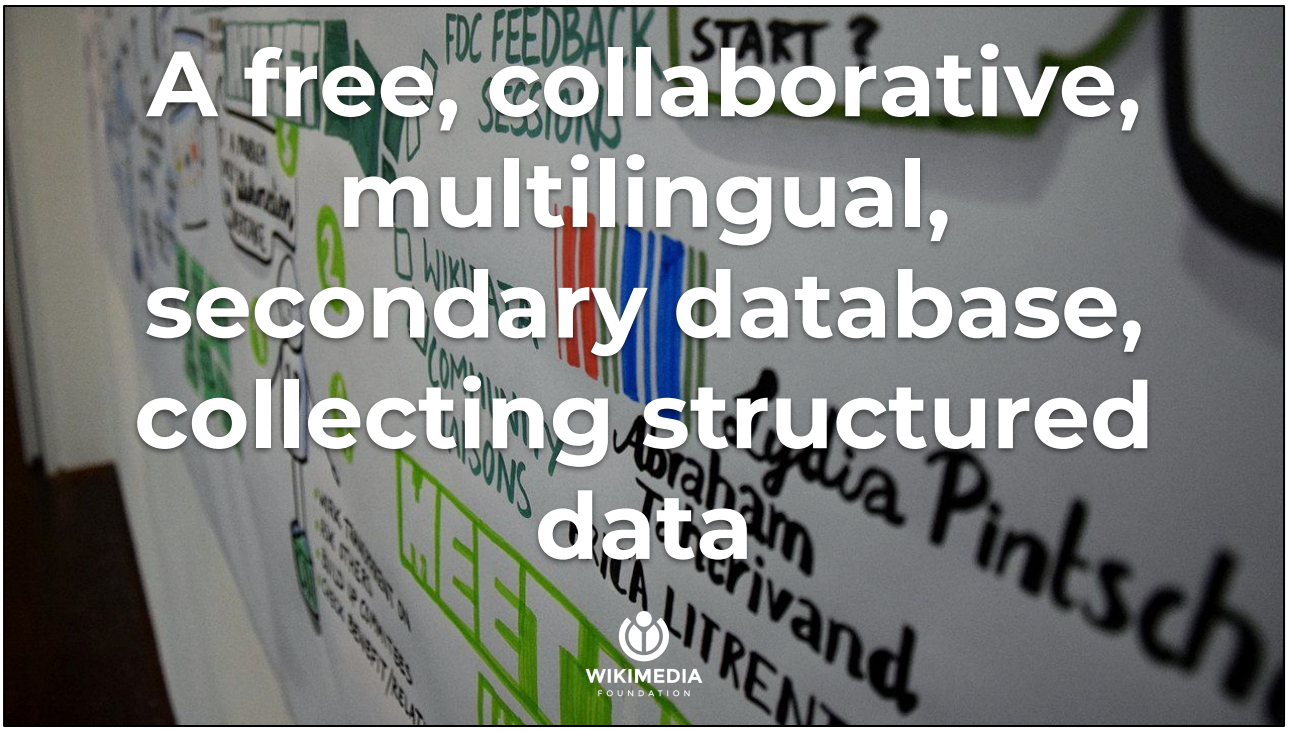
So, that's a little background on what we do. Now let's talk about one of our projects; Wikidata

Wikidata is becoming an important research tool

It's also the coolest thing ever. As my colleague Asaf likes to say, "You are lucky to be learning about it now. You are a Wikidata hipster"

"You'll be able to say you knew about Wikidata before it was cool."

Asaf's introduction to Wikidata video (cued to quote) :
https://youtu.be/eVrAx3AmUvA?t=1m3s



“Ok, Chris, I know I’m cool and all, but what is it?” - https://www.wikidata.org/wiki/Wikidata:Introduction

A free, collaborative, multilingual, secondary database, collecting structured data

Not quite six year old project started in late 2012

Little history
Who needs this? Well, at first it was just us.

Wikimedia projects like Wikipedia are available in close to 300 languages. Data - from everything between a birth date and the GDP of a country - was duplicated, wrong, and missing across articles on the same subject. Starting in 2013 the movement set out to centralize data in a single repository, connect it to projects and their articles, and keep the "wiki way" of contributing. It was called Wikidata.

- Wikimedians wanting to keep links to other versions of articles about the same topic in different languages.
- Infobox data (like birthdates, census data, GDP, etc.
- List articles based on data in Wikidata

data goes out of date or is updated - census data as an example. Volunteers would have to manually run around and update that in multiple locations. Who knows where the value appears on-wiki! So this was time-consuming and difficult.

All data is licensed under CC0 or public domain.

“All structured data from the main, property and lexeme namespaces is available under the Creative Commons CC0 License;”

I already explained free, let’s talk Collaborative

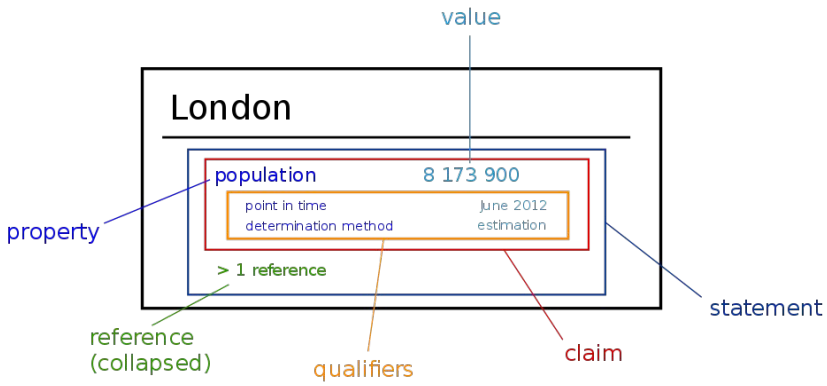
Data is entered and maintained by Wikidata editors, who decide on the rules of content creation and management. Automated bots also enter data into Wikidata. Anyone can edit, don’t even need an account.

There are multiple ways to contribute, from individual edits, running a bot, or bulk importing.

Multilingual

Editing, consuming, browsing, and reusing the data is fully multilingual. Data entered in any language is immediately available in all other languages. Editing in any of over 400 languages is possible and encouraged.

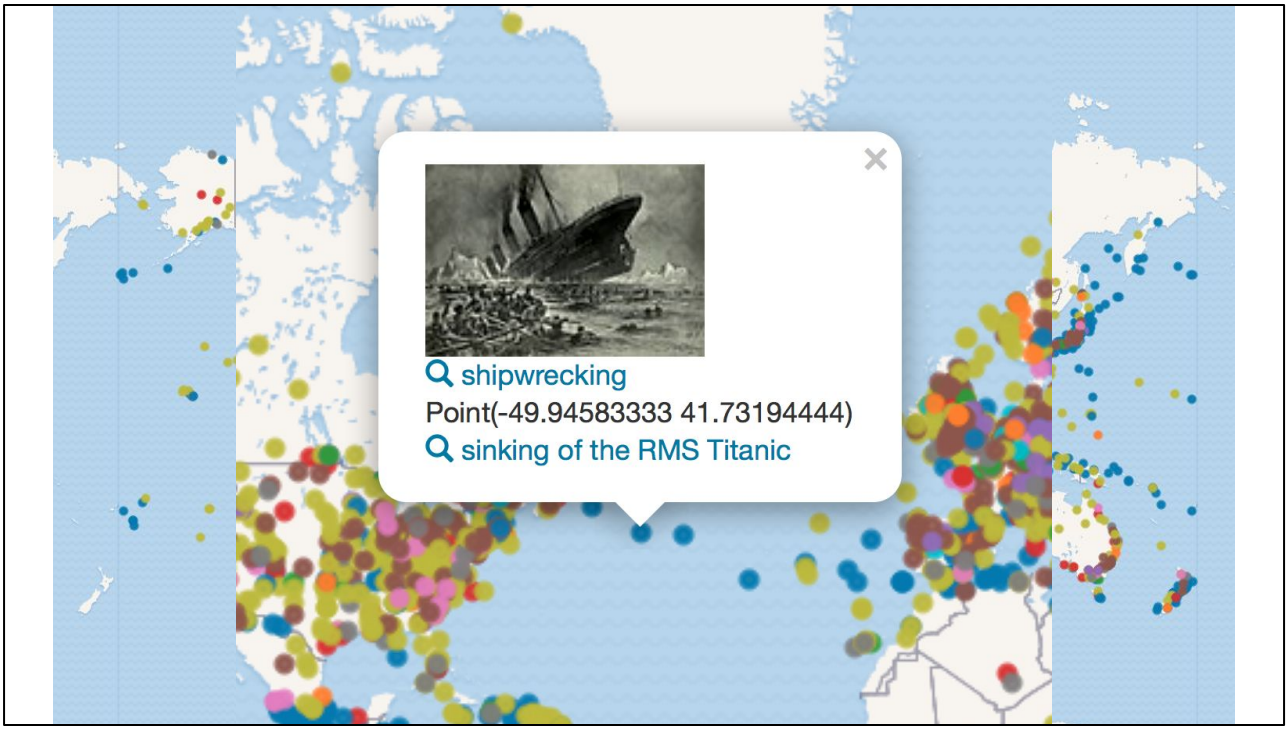
Secondary



secondary database

Wikidata records not just statements, but also their sources, and connections to other databases. This reflects the diversity of knowledge available and supports the notion of verifiability.

I'll explain this more in a minute.



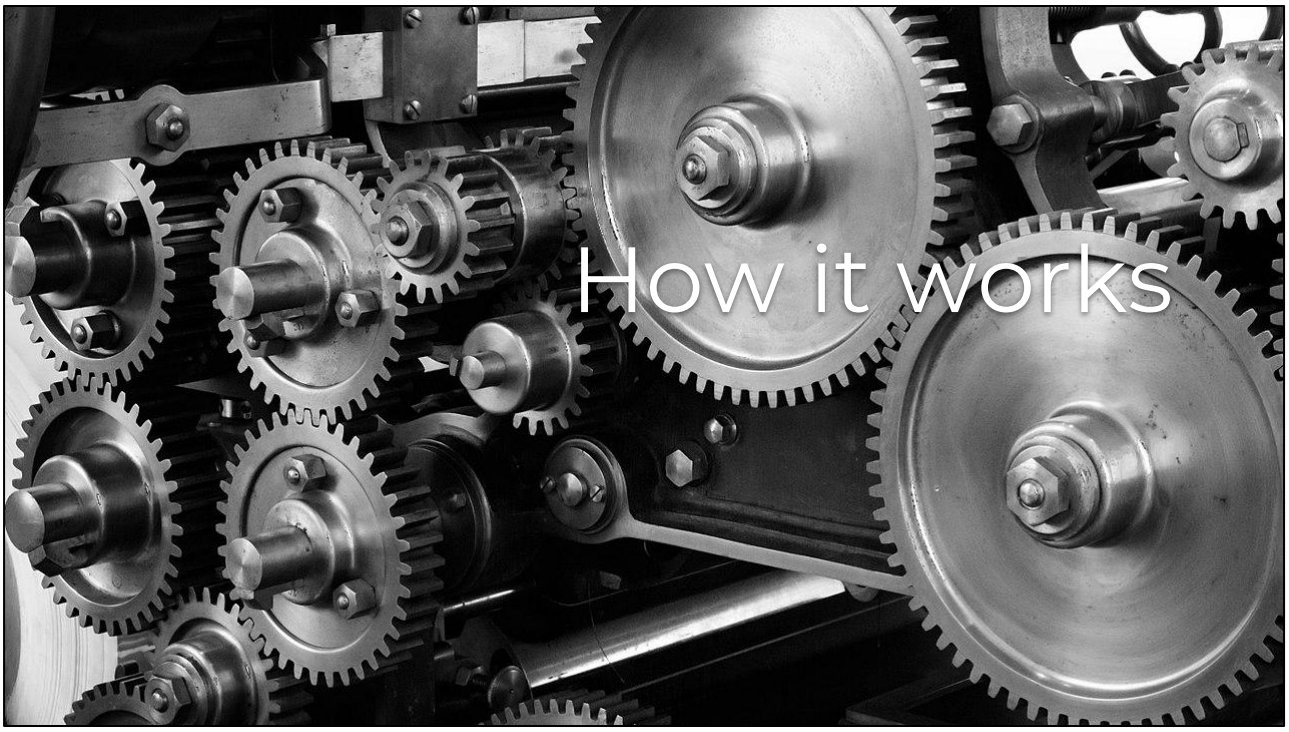
collecting structured data

Imposing a high degree of structured organization allows for easy reuse of data by Wikimedia projects and third parties, and enables computers to process and “understand” it.

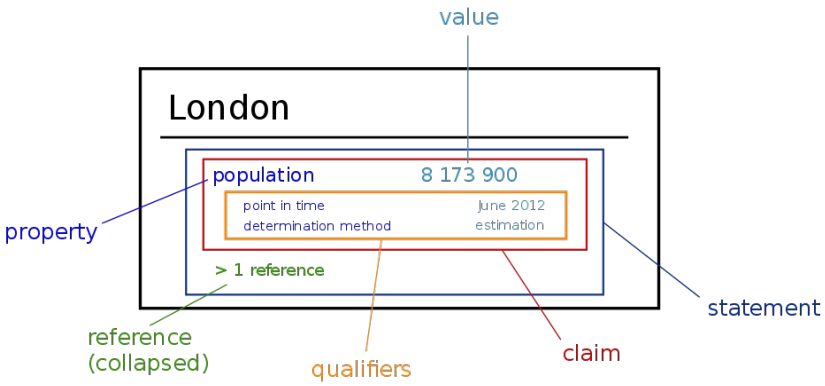
...support for Wikimedia wikis, and anyone in the world. This means you!

Here’s an example query. Don’t need to read this. I’ll show you more in a minute. Map of disasters, color-coded by disaster type

Source: Wikidata Query Services - http://tinyurl.com/ya54mv3w



So that is what Wikidats is, now let’s see how it works
Let’s take a look at an example page - https://www.wikidata.org/wiki/Q3100174



Statements consist of:

- claim (a property-value pair such as population(property): 8,173,900(value), point in time, determination method, etc)
- References - origin of a statement
- rank - quality factor of a statement when an item has multiple claims (say multiple occupations for an individual)

.

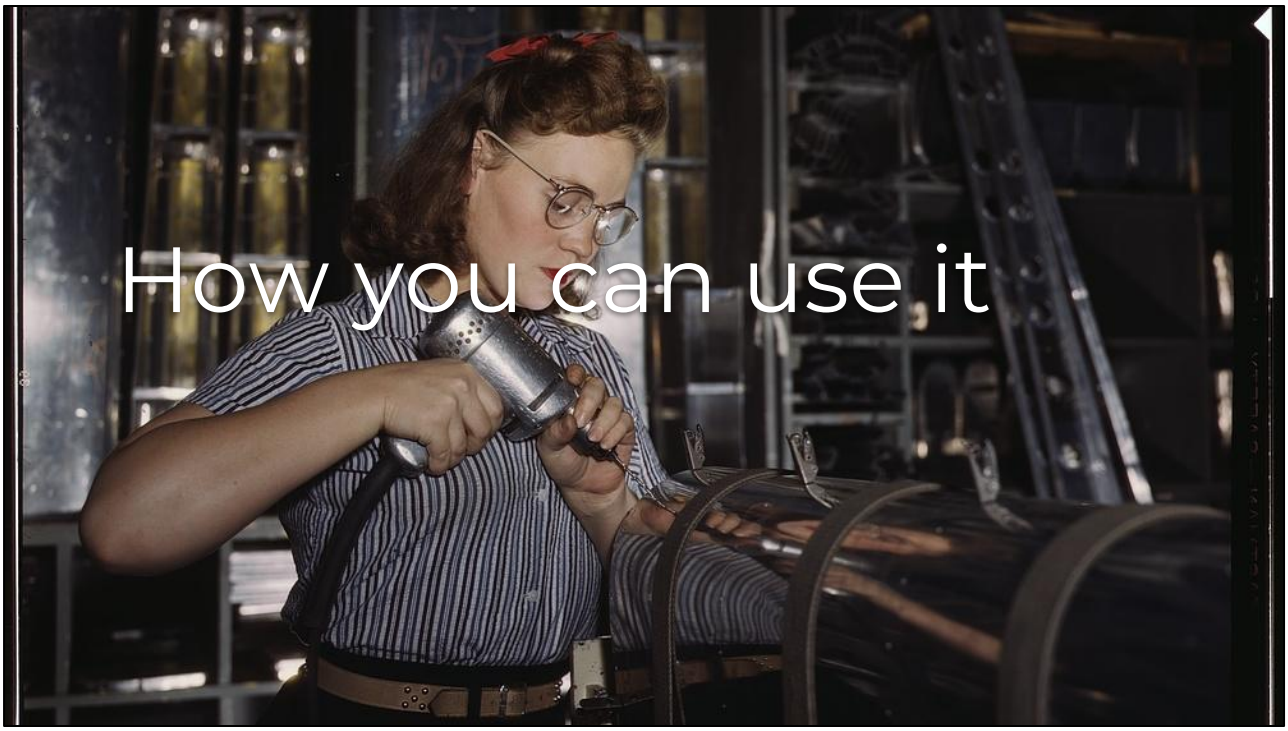
London is an item (Q84) that includes a statement about it's population.

Population is a property (P1082) and a value of 8,173,900

wikidata.org/wiki/Q3100174

Show how an entry is structured
Show how they are edited
Show history
Show recent changes or abuse filters

Live updates - https://www.wikidata.org/wiki/Special:RecentChanges?hidebots=1&hidecategorization=1&limit=50&days=30&enhanced=1&damaging__likelybad_color=c4&damaging__verylikelybad_color=c5&urlversion=2



Great Chris I get it. Your 5 minute demonstration has convinced me. Now how do I use it? :)

how you can use it - https://www.wikidata.org/wiki/Wikidata:Data_access

Linked Data interface - https://www.mediawiki.org/wiki/Wikibase/EntityData

https://www.wikidata.org/wiki/Q42

JSON - https://www.wikidata.org/wiki/Special:EntityData/Q42.json

P31 “instance of” property Q5 value (human) - seems mostly harmless

api - https://www.mediawiki.org/wiki/Wikibase/API

sparql - best for getting your feet wet (DEMO TIME)- https://query.wikidata.org

birthday today

Baseball players born in St. Louis - http://tinyurl.com/y7skhzf2

For other Wiki data: https://quarry.wmflabs.org/

Also has an API: https://www.mediawiki.org/wiki/Wikidata_Query_Service

R wrapper for WDQS API; send a SPARQL query, get back a data.frame - https://cran.r-project.org/package=WikidataQueryServiceR

Dumps - https://www.wikidata.org/wiki/Wikidata:Database_download - useful for offline use (JSON, RDF, XML)

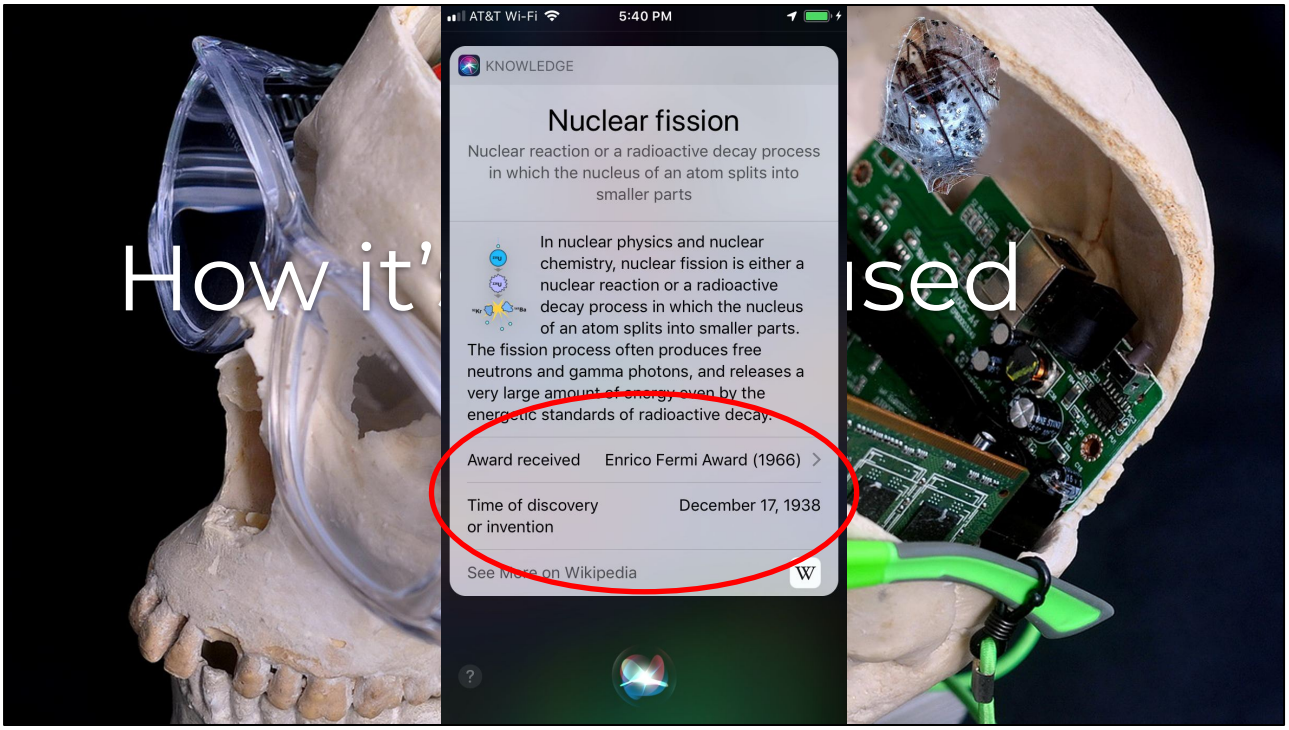
Bots - https://www.wikidata.org/wiki/Wikidata:Bots

Donate data - https://www.wikidata.org/wiki/Wikidata:Data_donation

query.wikidata.org

sparql - https://query.wikidata.org
birthday today
Baseball players born in St. Louis - http://tinyurl.com/y7skhzf2
(For other Wiki data: https://quarry.wmflabs.org/)

A similar tool for visualizing Wikidatal: https://tools.wmflabs.org/reasonator/?



How it's being used

Wikipedia itself!- https://www.wikidata.org/wiki/Wikidata:How_to_use_data_on_Wikimedia_projects#Examples_of_use_cases

https://fr.wikipedia.org/wiki/Modèle:Infobox_Fromage

Placehodler aritcles - beta softare - generates placeholder articles where one doesn’t exist - https://www.mediawiki.org/wiki/Extension:ArticlePlaceholder

Example on the Gujarati Wikipedia (a language native to the western state of Gujarati - 56 million speakers in India) - an entry on Theodore Roosevelt https://gu.wikipedia.org/wiki/વિશેષ:AboutTopic/Q33866

Google - knowledge graph - https://www.google.com/search?q=Urdu+language - information is sources from Wikidata among others

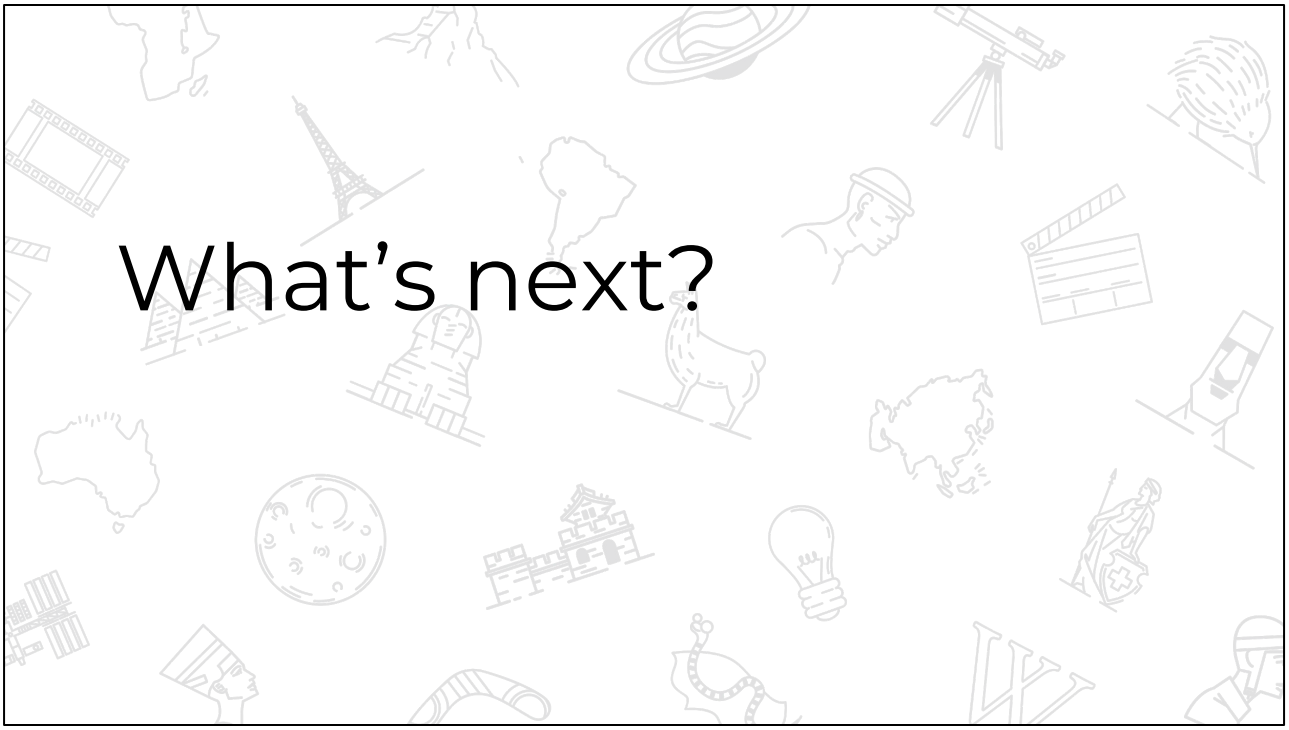
Apple - Siri knowledge panel - (CLICK) Here’s a query asking “What is Nuclear fission?” The “Award recieved” and “Time of discovery or invention” is sourced from wikidata - https://www.wikidata.org/wiki/Q11429

Amazon - Alexa

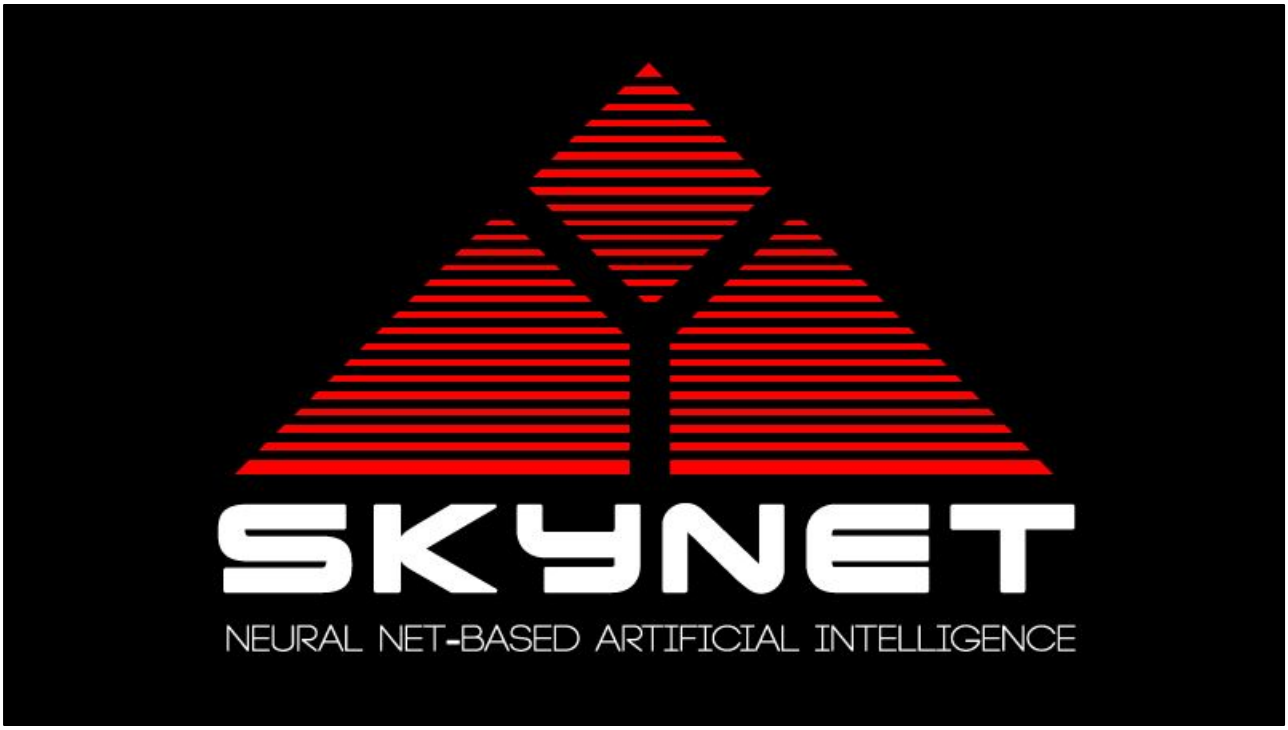
https://blog.quora.com/Announcing-Wikidata-References-on-Topics

https://www.wikidata.org/wiki/Wikidata:Wikidata_for_authority_control

https://twitter.com/pigsonthewing/status/867270004017909760

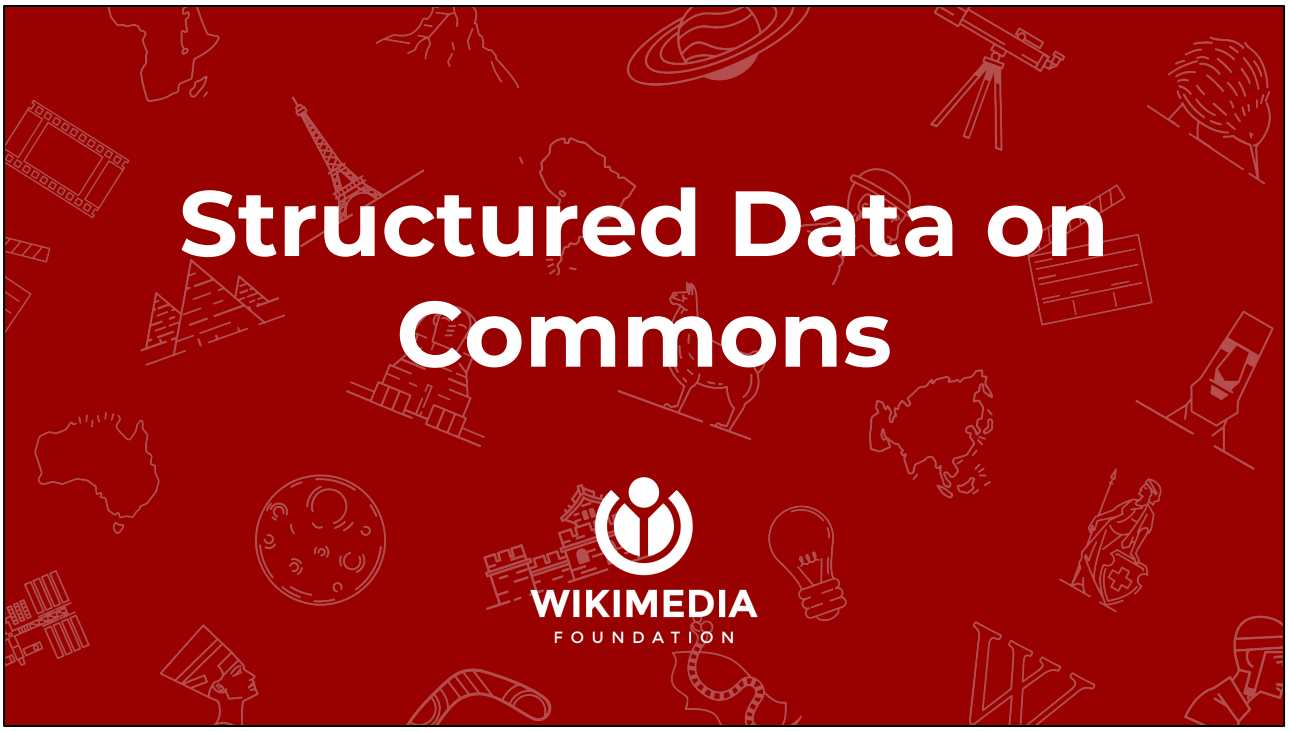


Covered the history, the current, and now, the future!
What's next?



JOKE
No, wait, sorry that was an accident!
JOKE

Structured data on Commons
Lexeme data - https://www.wikidata.org/wiki/Wikidata:Lexicographical_data
Your contributions! - https://commons.wikimedia.org/wiki/Commons:Welcome



Structured data on commons

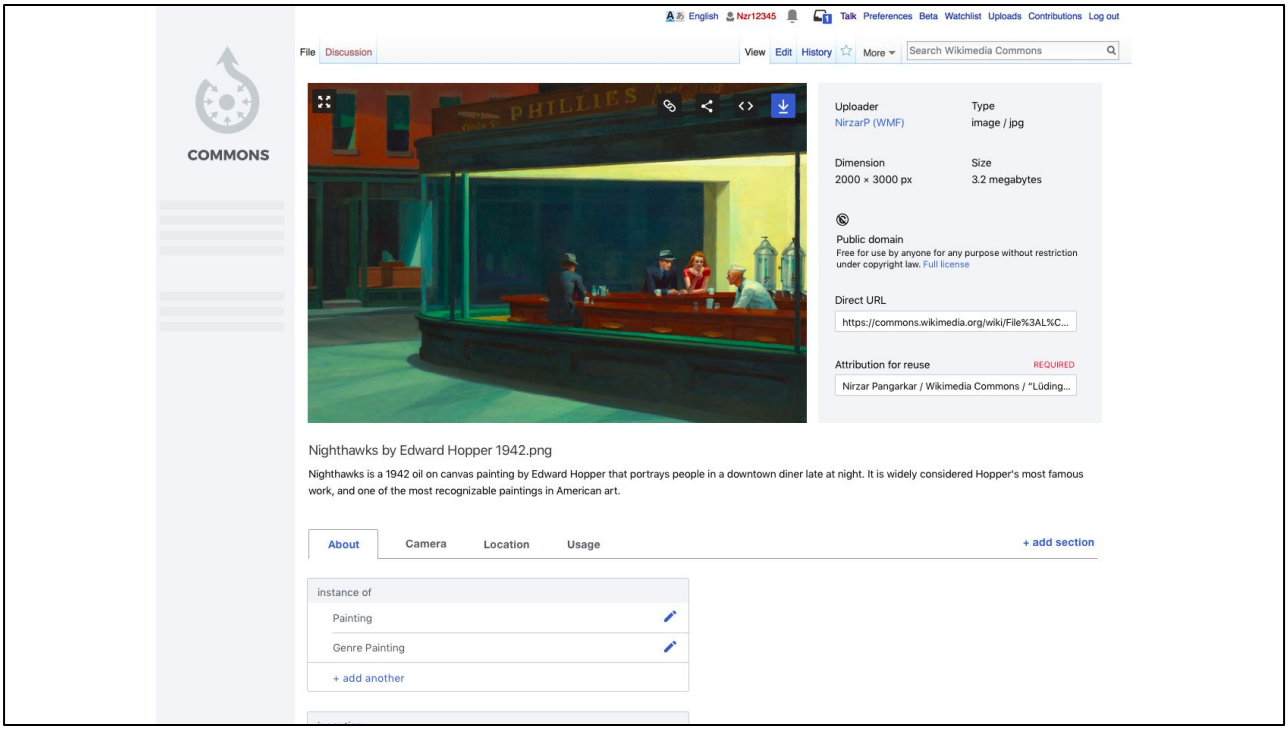
Commons is our media repository. Millions of files - images, animations, video, sound. Most unstructured.
All information about a file is wikitext - an open field of whatever. Ok, there's some structure with templates and categories.

https://commons.wikimedia.org/wiki/Commons:Structured_data

SDoC converts the free media files on Wikimedia Commons to a structured and machine-readable format, so that they become easier to view, search, edit, organize and re-use. To achieve that, the Commons backend is migrated to Wikibase, the same technology as used for Wikida

Idea is that you could define the metadata about a media (copyright, lens used, date taken, location, etc) and the contents (what does the media depict)

Great presentation to learn more:
https://upload.wikimedia.org/wikipedia/commons/6/60/Structured_Data_on_Commons_-_EuropeanaTech_2018.pdf



Example, note the structured form at the bottom with the option to add additional sections and fields.

Lexeme data - https://www.wikidata.org/wiki/Wikidata:Lexicographical_data

Structured dictionaries

It will allow editors to describe precisely all words in all languages, and will be reusable, just like the whole content of Wikidata

Structuring Wikipedia, Commons, and now data in a dictionary

More data types to come

Your contributions!
How can you improve the data? Tools? Feedback? Review? Visualizations? Identify gaps?
How can you contribute to the data? What can you add?
What tools could you build using the data? Tools for reuse, review?
Who can you share this cool new thing with?

- https://commons.wikimedia.org/wiki/Commons:Welcome

After all, you’re now a Wikidata hipster

Thank you for your time and listening to my terrible humor.

Time for questions?

Credits

- https://commons.wikimedia.org/wiki/File:Hercules_Killing_the_Lernean_Hydra.jpg
- https://commons.wikimedia.org/wiki/File:Group_photo,_Wikimania_2018,_Cape_Town_(_1050886).jpg
- https://commons.wikimedia.org/wiki/File:Wikimedia_Foundation_logo_-_vertical.svg
- https://commons.wikimedia.org/wiki/File:Wikidata-logo-en.svg
- https://commons.wikimedia.org/wiki/File:Feet-hipster-longboard-skateboard_(24031327000).jpg
- https://commons.wikimedia.org/wiki/File:Hipster_Professional_(Unsplash).jpg
- https://commons.wikimedia.org/wiki/File:Gears_(22711070429).jpg
- https://commons.wikimedia.org/wiki/File:Operating_a_hand_drill_at_the_North_American_Aviation_1a35312v.jpg
- https://commons.wikimedia.org/wiki/File:Intelligence_artificielle.jpg
- https://commons.wikimedia.org/wiki/File:Terminator-skynet.png
- https://commons.wikimedia.org/wiki/File:Focus_Graphic_Recordings,_wmcon14_berlin-003.jpg
- https://commons.wikimedia.org/wiki/File:Crowd_gathering_for_a_group_photo_at_Wikimania_2014.jpg
- https://commons.wikimedia.org/wiki/File:Library_of_Beijing_Foreign_Studies_University,_multilingual_decorative_wall,_Nov_2016.jpg
- https://commons.wikimedia.org/wiki/File:Wikidata_statement.svg
- https://commons.wikimedia.org/wiki/File:Lexeme_data_model.svg
- https://commons.wikimedia.org/wiki/File:Radio_City_Music_Hall_neon_(Unsplash).jpg
- https://commons.wikimedia.org/wiki/File:Structured_Commons_-_Design_exploration_Wikimania_2017_-_File_page_01.png