

# Wikidata

A Gentle Introduction for Complete Beginners

Asaf Bartov <asaf@wikimedia.org>

Video of this talk:

<https://www.youtube.com/watch?v=eVrAx3AmUvA>

Wikimedia Foundation, San Francisco, February 2017

**You are lucky!**

# Two key problems

1. Dated data; lacking triggers; repetitive work
2. Inflexible ways of lateral queries of knowledge (Categories and their limitations)

~~Two key problems~~

**One solution**

An editable central storage for **structured and linked data**, on a wiki, under a free license.

(AKA **Wikidata**)

In short:

**Wikidata is <3**

Structured data, you say?

**Statement = Item --> Property --> Value**

**Item** is any topic (Wikipedia article, thing, person, place, concept, etc.)

**Property** is a specific single **kind** of data relevant to this item (e.g. height for mountains, capitals for countries, gender for humans)

**Value** is either a reference to another Item (the capital of Germany is Berlin) or a literal value (e.g. 8848 meters)

# Items and Properties

An **item** is described by a series of **statements**, each asserting a single datum or fact about it.

A **property** is a specific kind of information, like a field in a form, and is used to describe what the data means.

## Examples

An **item** about a **country** would have statements about the following **properties**: population, land area, official languages, borders-with, anthem, capital city, demononyms, Gross Domestic Product, etc. etc.



## Examples

An item about a **person** would have statements about: name, date of birth/death, place of birth/death, citizenship, occupation, languages spoken, religion, father, mother, children, notable works, place of education, etc. etc.

## Statement examples (Item --> Property --> Value)

Earth --> highest-point --> Mount Everest

Mount Everest --> elevation above sea level -->  
8848 meters

Earth --> deepest point --> Challenger Deep

Challenger Deep --> elevation above sea level -->  
-10,994±1 metre

## **Another way of looking at it**

### **Earth -->**

highest-point: Mount Everest

deepest point: Challenger Deep

### **Challenger Deep -->**

elevation above sea level:  $-10,994 \pm 1$  metre

### **Mount Everest -->**

elevation above sea level: 8848 meters

## Once more with ~~feeling~~ numeric IDs

Earth (**Q2**) --> highest-point (P610) --> Mount Everest (Q513)

Mount Everest (Q513) --> elevation above sea level (P2044) -->  
8848 meters

Earth (Q2) --> deepest point (P1589) --> Challenger Deep  
(Q459173)

Challenger Deep (Q459173) --> elevation above sea level  
(P2044) --> -10,994±1 metre

# Once more without squishy humanspeak

Q2 --> P610 --> Q513

Q513 --> P2044 --> 8848 meters

Q2 --> P1589 --> Q459173

Q459173 --> P2044 -->  $-10,994 \pm 1$  meters

# Interlude: Why numbers?

Labels are ambiguous: What is London?

City in England, in Canada, family name,  
personal name, movie company, hotel?

Language-neutral - not everybody speaks English!

Robot-friendly (robots love numbers ;)

Questions so far?

# Let's go explore Wikidata!

<https://wikidata.org>

<https://www.wikidata.org/wiki/Q17141>



# Questions?

(also: remember [\[\[Wikidata:Project chat\]\]](#))

# Let's teach Wikidata some things!

<https://www.wikidata.org/wiki/Q546374>

<https://www.wikidata.org/wiki/Special:NewItem>

**Contributing to Wikidata  
on your commute!**

# The Wikidata Game

Semi-automates micro-decisions and make incremental improvement of Wikidata super-fun.

<https://tools.wmflabs.org/wikidata-game>

<https://tools.wmflabs.org/wikidata-game/distribute>  
d

# Embedding Data from Wikidata

{{#property}} is magic! :)

[\[\[m:User:Ijon/Wikidata\\_embedding\\_SF\\_demo\]\]](#)

# Article Placeholder

<https://eo.wikipedia.org/w/index.php?search=Helen+DeWitt&title=Speciala%C4%B5o:Ser%C4%89i&go=Ek>

# Querying Wikidata

<https://query.wikidata.org/>

# Querying Wikidata

- \* query Wikidata using [\[\[SPARQL\]\]](#)
- \* Wikidata will tell you **everything it knows**, but **no more**
- \* The more data and the more links in Wikidata, the more useful it becomes (so contribute!)



[query.wikidata.org](https://query.wikidata.org) demos

# Links to sample queries

- [Notable cats](#)
- [Popular surnames among fictional characters](#)
- [Overall causes of death ranking \(presented as bubble chart\)](#)
- [Painters whose fathers were also painters](#)
- Count total articles in Wikipedia language XX by gender:  
<http://tinyurl.com/j6nedlb> (women) and <http://tinyurl.com/zdvghf> (men)
- [Largest cities with female mayor](#)
- [Paintings known to Wikidata, by location, on a map](#)
- [African women with articles in English but not in French](#)
  - [Same, but only good/featured articles in English \(i.e. "badges"\)](#)

Remember: “**good artists borrow, great artists steal.**” --attributed to Picasso

Questions?

Remember:

**Wikidata is <3**

# Extra Credit

For when you just can't get enough of Wikidata

## WD-FIST

- Free Image Search Tool for Wikidata

<https://tools.wmflabs.org/fist/wdfist/index.html>

## The Wikidata Reasonator

- Pretty-format data from Wikidata
- <https://tools.wmflabs.org/reasonator>

## Mix and Match

- Match Wikidata entities to external databases' entities

<https://tools.wmflabs.org/mix-n-match/>

<https://tools.wmflabs.org/mix-n-match/?mode=sitstats&catalog=92> (coverage of members of the Royal Society)

# Go edit Wikidata! :)

Now that you have the power,  
use it for good and not for evil.

Video of this talk:

<https://www.youtube.com/watch?v=eVrAx3AmUvA>  
asaf@wikimedia.org