Wikidata for Academics

Large-scale, interdisciplinary, multilingual, open data
Why should you be interested in Wikidata?

- A brief introduction to the largest open knowledgebase
  - Why you might draw data from it
  - Why you might deposit data into it

How to put data in

- Different ways to editing and create items
  - One at a time (direct manual editing)
  - Multiple at once (tool assisted batch editing)

How read data from

- Query information back out of Wikidata
  - SPARQL tool (and language)
  - A wide range of custom examples

Help, community and resources

- The hidden world behind Wikidata
WIKIPEDIA

Open access encyclopedia
300 language editions
Most-read source for most topics
Many pages read >100,000 times per year
Consulted by >95% of students
Consulted by >60% of clinicians
Consulted by >95% of researchers

WIKIDATA

Structured data knowledgebase
Machine-readable
Multilingual
Inter-disciplinary

e.g. Research output metadata
e.g. Scholar & researcher data
e.g. Conceptual relationships
Part of wider open knowledge ecosystem
WHY INTERACT WITH WIKIDATA

WHY DEPOSIT DATA INTO
  make sure AIs understand the world
- Ensure structured knowledge is FAIR
  vs proprietary databases and walled gardens
- There for the long term
  not susceptible to funding cycles
- Large existing community, tools, support

WHY DRAW INFO FROM
- Do interdisciplinary research
- Unique resource identifiers
  Common data language
  Inter-database dictionary
  Make your projects multilingual
- Transparent history for all items
WIKIPEDIA (EN)

Number of articles: 6M
Number of editors: 60K

WIKIDATA

Number of items: 80M
Number of editors: 40K

https://stats.wikimedia.org
EXAMPLE ITEMS AND PROPERTIES

Concepts
E.g. locations, events, diseases, genes, languages, processes

COVID-19 (Q84263196)

Has effect (P1542)
cytokine release syndrome (Q3961647)

Colchicine (Q326224)

Main subject (P921)

Research outputs
E.g. publications, journals, publishers, clinical trials

We Need Compassionate Leadership Management Based on Evidence to Defeat COVID-19 (Q96947691)

Main subject (P921)

Colchicine Coronavirus SARS-CoV2 Trial (Q89154576)

Research intervention (P4844)

People & orgs
E.g. Researchers, authors, universities, institutes

Agnes Binagwaho (Q4693015)

Author (P50)

Montreal Heart Institute (Q3151984)

Sponsor (P859)

etc
Granular disambiguation of concepts
All of Wikidata tied into a single network
90M items
>1.1B connections
Links data from wide variety of disciplines
OUTLINE

WHY SHOULD YOU BE INTERESTED IN WIKIDATA?
- A brief introduction to the largest open knowledgebase
  Why you might draw data from it
  Why you might deposit data into it

How to put data in
- Different ways to editing and create items
  One at a time (direct manual editing)
  Multiple at once (tool assisted batch editing)

How read data from
- Query information back out of Wikidata
  SPARQL tool (and language)
  A wide range of custom examples

HELP, COMMUNITY AND RESOURCES
- The hidden world behind Wikidata
**LETS MANUALLY EDIT AN ITEM**

### PhD theses that don’t list a publication date

1. Go to [https://w.wiki/ekc](https://w.wiki/ekc) and press
2. Find the thesis’s URL for the thesis
3. At the bottom of the Wikidata page, + add statement
   - Add “publication date” based on what’s listed at that URL
4. In that statement, + add reference
   - As the “reference URL”, use the thesis’s URL

### PhD theses that don’t list a main topic

1. Go to [https://w.wiki/ekg](https://w.wiki/ekg) and press
2. Find the thesis’s URL for the thesis
3. At the bottom of the Wikidata page, + add statement
   - Add the main subject(s) based on the keywords or title
4. In that statement, + add reference
   - As the “reference URL”, add the thesis’s URL
   - (or) As the reference, add “based on heuristic” “inferred from title”
**LET'S BULK EDIT SEVERAL ITEMS AT ONCE!**

- **Two main tools:** QuickStatements ([https://quickstatements.toolforge.org](https://quickstatements.toolforge.org)) and OpenRefine ([http://openrefine.org](http://openrefine.org))

- **Import tabular data into Wikidata**

- **Example:**

Create a new item
It’s an instance of a human
Its label in English is...
Its sex/gender is male
Its place of birth is Rome
Its place of death is Rome

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CREATE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>LAST</td>
<td>P31</td>
<td>Q5</td>
<td>S143</td>
<td>Q24731821</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>LAST</td>
<td>Len &quot;Giovanni Francesco Rossi&quot;</td>
<td></td>
<td>S143</td>
<td>Q24731821</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>LAST</td>
<td>P21</td>
<td>Q6581097</td>
<td>S143</td>
<td>Q24731821</td>
<td>+2017-10-04T00:00:00Z/11</td>
</tr>
<tr>
<td>5</td>
<td>LAST</td>
<td>P19</td>
<td>Q220</td>
<td>S143</td>
<td>Q24731821</td>
<td>+2017-10-04T00:00:00Z/11</td>
</tr>
<tr>
<td>6</td>
<td>LAST</td>
<td>P20</td>
<td>Q220</td>
<td>S143</td>
<td>Q24731821</td>
<td>+2017-10-04T00:00:00Z/11</td>
</tr>
</tbody>
</table>

Item to edit | Statements to add | Supporting reference
LETS BULK EDIT SEVERAL ITEMS AT ONCE!

PhD theses that don’t list a publication date
1. Go to Docs for the list https://tinyurl.com/QStable
   To see where this list came from, see https://w.wiki/f25
2. Go to https://quickstatements.toolforge.org
   Paste across a few random rows from the table
   “Import V1 commands” (if tab separated)
   (or) “Import CSV commands” (if comma separated)
3. Hit “Run” to add the information to Wikidata
Outline

Why should you be interested in Wikidata?
- A brief introduction to the largest open knowledgebase
  - Why you might draw data from it
  - Why you might deposit data into it

How to put data in
- Different ways to editing and create items
  - One at a time (direct manual editing)
  - Multiple at once (tool assisted batch editing)

How to read data from
- Query information back out of Wikidata
  - SPARQL tool (and language)
  - A wide range of custom examples

Help, Community and Resources
- The hidden world behind Wikidata
Find timeline of intervention studies on COVID-19
1. Go to query.wikidata.org
2. Filter for items that are interventional studies study type
3. Filter for items that are about COVID-19 as their main subject
4. Show their research intervention
5. Show their start time
6. Press and find something interesting

Based on SQL language
Isn’t as scary as you think
The code is human-readable(ish)
EXAMPLE QUERY RESULTS

Gendered age bias: actors vs academics

Adapted from Toby Hudson, Uni. Sydney, CC BY 4.0; https://w.wiki/Dw8; https://w.wiki/Dw9
SOME INTERESTING USAGE EXAMPLES FOR INSPIRATION

Map of witches in Scotland
witches.is.ed.ac.uk

Events timelines
histropedia.com/timeline

Scholia dashboard
scholia.toolforge.org

Chlamydia genome browser
chlambase.org

SO MANY TOOLS!

<table>
<thead>
<tr>
<th>General tools</th>
<th>wikidata.org/wiki/Wikidata:Tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Mix-n-match</td>
<td>tools.wmflabs.org/mix-n-match</td>
</tr>
<tr>
<td>- WD games</td>
<td>wikidata-game.toolforge.org/distributed</td>
</tr>
<tr>
<td>- Author disambiguator</td>
<td>author-disambiguator.toolforge.org</td>
</tr>
<tr>
<td>- Zotero importer</td>
<td>wikidata.org/wiki/Wikidata:Zotero</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Programmer tools</th>
<th>wikidata.org/wiki/Wikidata:Tools/For_programmers</th>
</tr>
</thead>
<tbody>
<tr>
<td>- API for WikiData</td>
<td>wikidata.org/w/api.php</td>
</tr>
<tr>
<td></td>
<td>for QuickStatements</td>
</tr>
<tr>
<td>- Python package</td>
<td>github.com/dahlia/wikidata</td>
</tr>
<tr>
<td>- R package</td>
<td>github.com/TS404/WikidataR</td>
</tr>
</tbody>
</table>
OUTLINE

WHY SHOULD YOU BE INTERESTED IN WIKIDATA?
- A brief introduction to the largest open knowledgebase
  Why you might draw data from it
  Why you might deposit data into it

How to put data in
- Different ways to editing and create items
  One at a time (direct manual editing)
  Multiple at once (tool assisted batch editing)

How read data from
- Query information back out of Wikidata
  SPARQL tool (and language)
  A wide range of custom examples

HELP, COMMUNITY AND RESOURCES
- The hidden world behind Wikidata
WHERE TO GET HELP?

GENERAL INFORMATION

- Interactive tours
  wikidata.org/wiki/Wikidata:Tours

- Academic introductory publications

ONE-TO-ONE

- General help and discussion
  wikidata.org/wiki/Wikidata:Project_chat

- Requesting SPARQL queries
  wikidata.org/wiki/Wikidata:Request_a_query

- Special interest Wikiprojects
  Examples: Chemistry Philosophy
  Source Metadata Covid19
  Open Access Ontology

List: (wikidata.org/wiki/Wikidata:WikiProjects)
Contact

Email Thomas.Shafee@gmail.com
Google Scholar Thomas Shafee
ResearchGate Thomas Shafee
LinkedIn Thomas Shafee
Username Search [[user:tshafee]]

Journals

WikiJournal of Medicine (WikiJMed.org)
WikiJournal of Science (WikiJSci.org)
WikiJournal of Humanities (WikiJHum.org)
PLOS (TopicPagesWiki.plos.org)


[END]