Developing a linked data workflow using Wikidata

Will Kent
Wiki Education

SWIB - September 2023

Wiki Edu

Agenda

- **Program background**
- **Course details**
- **Sharing outcomes**

Wiki
Edu

Background

- **Wiki Education courses started in 2019***
  Intro course; no prior experience necessary
  Six sessions, each an hour long
  Seven slide-based training sessions and some exercises as homework^
  ~3 hour commitment/session
  All meetings virtual
  Slack channel to supplement
  Particular emphasis on community integration, querying, manual editing, tools, and batch edits

*https://meta.wikimedia.org/wiki/Wiki_Education_Foundation/Wikidata_Program_Evaluation
^https://dashboard.wikiedu.org/training/wikidata-professional

Wiki Edu

Outcomes

- **Based on survey results we know the course is a success!**
Participants are satisfied with the curriculum
Understanding of Wikidata fundamentals
There was one common piece of feedback…

Let's keep going

- **Is there an advanced course?**
Can we have space to practice with these tools?
How do we convince stakeholders to run Wikidata projects?
Complex questions about more queries, tools, and policies
How can we develop a linked data workflow with Wikidata?

Some challenges

- **Different goals**
Metadata roundtripping
Metadata enrichment
Modeling/schemas
Translation
Data visualization
Querying
Property creation
Wikipedia work
Intuition can only take you so far

Sample questions

- **What are best practices for translating labels, descriptions, and aliases?**
How do I highlight the work our researchers do in Scholia?
Everything about batch edits - how do I make them? How can I be precise? How can I build them into a workflow for my staff?
How do I create a property? How do hierarchies work on Wikidata? Do they work?
How do I develop visualizations with Wikidata?
How do you create a workflow?

Wiki Edu

A project-based course

- **In November 2022 we piloted a new Wikidata curriculum, requiring a project and accompanying dataset to participate**
To date we have run two of these courses with another two on the horizon
Twenty-five participants so far
Sharing the outcomes of this these courses is important as Wikidata becomes more common in libraries, museums, and across the internet

Course details

Same setup as the beginner course - six sessions, hour-long, all virtual

No training modules, more live demos

Aim for 10-20 participants per course

Emphasis on the cohort learning together

1:1 meetings encouraged

Specific case studies and resources

Meetings loosely themed, but greater emphasis on participant-driven sessions

Wiki Edu

Curriculum

Some experience with Wikidata is assumed

Meetings themed around data modeling, querying/data visualization, general Wikidata tools, batch editing, documentation and community integration

Specific tools include Cradle, Tabernacle, SQID/Reasonator, Mix'n'match, SourceMD, Merge tools, Graph builder, Wikidata/Wikipedia spreadsheet extension, PAWS, PetScan, Listeria, etc.

Batch tools include Quickstatements, Open Refine, Author Disambiguator

Wikibase - how and when you may want to use it

Ample time left for discussion, troubleshooting, and individual questions/participant-led tutorials

Wiki Edu

Wikidata:Linked open data workflow

Other languages: English ⊞   español ⊞   français ⊞   hrvatski ⊞   português ⊞   中文 ⊞   日本語 ⊞

There are many considerations when contributing data, media or other assets to Wikimedia projects. This chart attempts to list some of them and track tools and scripts in the linked data workflow. This is based on the data and media partnerships chart on Outreach Wiki.

PREPARE and tidy source data and images	RECONCILE with Wikimedia modeling and coverage	INGEST data and media	ANALYZE, correct, and enrich	RE-USE content intra-wiki and externally	REPORT and measure
• Wikidata Data Donation • Creative Commons • Flickr image hosting • Google Spreadsheets, Wikipedia and Wikidata tools for Google Spreadsheets	• OpenRefine • Wikidata Mix'n'match • Wikidata Query • WikiCommons Query Service • Wikidata Graph Builder • SQID • Petscan • PAWS • Mediawiki API	• QuickStatements • OpenRefine • Wikidata-CLI • Pattypan • url2commons • PyWikibot / upload.py • Mediawiki API • Commons Upload Wizard (from Flickr) • GLAM Wiki Toolset (deprecated)	• Wikidata Query • TABernacle • Listeria • Maintenance query scripts • Wikidata Distributed Game • Cat-a-lot • VisualFileChange.js • SPARQL RC	• Wikidata infobox / Infobox templates • Creator templates • Template:Artwork, Template:Book, etc • Template:Label, Template:WrapWD, etc • Listeria / Template:Wikidata_list	• GLAMorgan • BaGLAMa • GLAM Wiki Dashboard
Notes [edit] Try finding a similar project or collection set on Wikidata or Commons to see how it has been done in the past.	**Notes** [edit] For Wikidata, usually a "crosswalk database" is needed to map terms from the uploading data set (a CSV file or records from an API) to	**Notes** [edit] Try uploading small test batches before doing large data sets. When ingesting collection	**Notes** [edit] Depending on the success of the import and uploading process, you may need to deal with duplicates or conflicts with other editors	**Notes** [edit] Scripts and templates can generate on-wiki content such as tables and infoboxes from Wikidata.	**Notes** [edit] Show the impact of contributions by tracking metrics on files used or impressions over time. For partnerships, this can help

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

Impact

1.69K	**2.27K**	**5.84K**	**11**	**18.7K**	**20.6K**	**1**
Items Created	Items Edited	Total Edits	Editors	References Added	Item Views	Commons Uploads

www.wikidata.org

General

5.84K	Total revisions
8	Merged
3	Interwiki links added

Items

1.69K	Created
6	Cleared

Claims

17.9K	Created
984	Changed
314	Removed

Labels

1.74K	Added
513	Changed
23	Removed

Descriptions

1.74K	Added
60	Changed
9	Removed

Aliases

1.94K	Added
268	Changed
6	Removed

Other

1.58K	Qualifiers added	0	Reverts performed
11.4K	References added	0	Restorations performed
8	Redirects created	30	Other updates

Lexeme

0	Created

https://dashboard.wikiedu.org/courses/Wiki_Education/Wikidata_Project_Course_(Spring_2023)/home

Wiki Edu

Why is this important?

Structured courses with live instructors remain rare

Without continued work, projects stop

Creating a space for sharing allows for more support, exchange of ideas, and recommendations for best practices

Since there are so many moving parts with a workflow, if one person leaves, projects can slow or stop outright

The impact project-based courses have on Wikidata can be BIG!

Not to mention impact on local collections

New tools and systems will appear - having courses, spaces, and people familiar with this will improve a lot

Building a workflow is complicated - there is no manual for any of this

Wiki
Edu

Course feedback

People want more! Longer courses, sustained contact, more time to spend on projects with mentors, coaches, other practitioners

There is no one-size-fits-all course and that's okay

Content that was review for some was helpful and not annoying or a waste of time

The size was particularly helpful - there are many Wikidata offerings - videos, slide tutorials, etc, but the combination of size, live sessions, and timing worked well for many

Wiki Edu

Pedagogical style

Participant-driven (when possible)

Inquiry and action-based

Open pacing, leave room for conversation to develop

Curriculum pushes editing and becoming integrated into the community; encouraging these is central to the fabric of the course

Assume Wikidata is right and wrong; frame as opportunities for improvement

Wiki
Edu

What's next?

Build the curriculum out to better engage with more specific needs

For popular topics, create training modules

Document everything - failures, successes, and workflows

Wikidata in the classroom

Share findings at conferences to encourage more Wikidata work and convince a broader audience to let their staff run Wikidata projects

Thank you!

Learn more /
stay in touch

will@wikiedu.org
learn.wikiedu.org

Wiki
Edu