Getting the most out of Wikibase4Lib

LD4 Wikibase Working Hour

Jim Hahn (University of Pennsylvania), Esther Jackson (Columbia University), Timothy Ryan Mendenhall (Columbia University)

Code4Lib 2022 Pre-conference

23 May 2022

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

About the LD4 Wikibase Working Hour

The LD4 Wikibase Working Hour seeks to create a space for GLAM professionals experimenting with Wikibase/WBStack implementation, the software that Wikidata is based on, to learn collaboratively and share tips, tools, and resources. The working group will facilitate identification of areas for collaboration among institutions experimenting with separate institutional Wikibase instances. Topics considered are:

- Shared data modeling of properties common to many GLAM collections
- The development of documentation for the GLAM Wikibase community
- Learning about each other's Wikibase projects, workflows, and tools.
- Explore the development of mechanisms for channeling community feedback to the developers of Wikibase and the broader Wikimedia community.
- Complement, (or supplement) where possible, the work from existing Wikidata groups

More details on LD4 Wikibase Working Hour wiki page

WBStack and wikibase.cloud: background

- “Wikibase as a service”--shared hosting, installation, and maintenance for Wikibase instances
- “easily curate data in your own Wikibase instance, without worrying about the technical side of things”--https://www.wbstack.com/
- WBStack is an alpha release, and wikibase.cloud will replace it
- Wikibase.cloud will be managed by Wikimedia Deutschland
- Stated goal of keeping wikibase.cloud a free service
- Ties in to the Wikimedia 2021 Linked Open Data Strategy “Wikibase Ecosystem”

Previous LD4 Wikibase Working Hour sessions

- Two demonstrations on setting up Wikibase instances
- Two sessions on WBStack [soon to be migrated to Wikibase Cloud]
- Presentations on Wikibase instances hosted at Bibliotheque nationale de France, University of Toronto, Rhizome, Enslaved.org, Digital Scriptorium, and Pratt Semantic Lab
- Interactive series of workshops to build a WBStack instance tracking usage of alternative labels for LCSH [October-December 2021]--see next slide
- For more information, please consult our LD4 Wikibase Working Hour wiki page

2021 LD4 WBStack experiment

- Overall goal: use WBStack, a free Wikibase sandbox, to create a repository of local labels being used for LCSH headings: https://ld4-wbs-test.wiki.opencura.com/wiki/Main_Page
- No sustainability plan in place:intended as an exercise or experiment
- Draft data model
  - Centered on labels, not concepts. Think of the RDA entity Nomen and its affiliated Appellation properties
  - No attempt to create new alternative labels, just to document alternative labels currently in use
  - No attempt to replicate or import the entirety of LCSH, just the aspects needed to document the use of alternative labels for LCSH
  - On planning sheet, organized around “shapes” or primary entities being described by the model. Some properties repeat across entities, like Label, Alias, Description
- Draft data model was developed by planning group over 6 months
- Over three monthly sessions, participants work on data modeling, property creation, and item creation

2021 WBStack experiment : Primary entities

- Preferred local label
- Vocabulary
- Institution
- External label
- System
- Policy
- Proposed entities:
  - Source citation
  - Revision proposal
  - Replacement method

Demo: adding and editing items and properties

Using Columbia University Libraries' WBStack sandbox, I will do the following:

- Show the WBStack Dashboard
- Briefly show user management features
- Add statements to an existing item
- Add a property
- Add an item for University of Buffalo Libraries
- Look at the query service and run a sample query
- Look at the mappings and federated properties features

WBStack: Moving beyond defaults: Dashboard

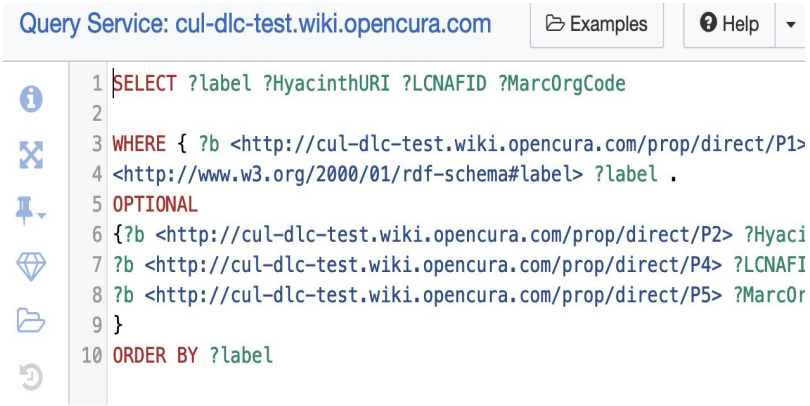
- Basic branding (logo, theme)
- Users: free-for-all vs. approval required
- Length for values and “multilangs” (aka Labels, Descriptions, Aliases)
- Each WBStack instance is blank after you create it via the dashboard: your WBStack.com login will not transfer over to each WBStack instance!
- General principle: “social” infrastructure of Wikidata is not in place

WBStack: User management

- All instances
  - “PlatformReservedUser” to enable the WBStack administrators to access and update the instance as needed
  - User pages are a default feature, but need to be created, as in Wikipedia / Wikidata
  - Special pages→ User list
  - Group privileges, password policies can be customized to your instance
  - DEFAULT: All users can create items AND properties without any approval process!
- “Closed” instances:
  - First account created should be for the bureaucrat and instance administrator
  - Instance admin will have to approve all users (until additional admins and bureaucrats are appointed!)
  - Special pages→ Confirm account requests
- For open instances: optionally assign users to groups based on needed permissions

Tools and querying

- Query service: prefixes do not carry over from Wikidata
  - PROPERTY: <http://SITE-HANDLE.wiki.opencura.com/prop/direct/P#>
  - ITEM: <http://SITE-HANDLE.wiki.opencura.com/entity/Q#>
- QuickStatements
  - As with above, property and item numbers will be different; otherwise behaves as with Wikidata
- OpenRefine: Custom reconciliation service is possible: https://openrefine-wikibase.readthedocs.io/en/latest/
- API:
  - https://SITE-HANDLE.wiki.opencura.com/w/api.php
  - https://SITE-HANDLE.wiki.opencura.com/w/rest.php
  - Sandbox also available at Special pages→ API Sandbox



Mappings and federation

- Only accessible to WBStack admin, not to general users
- Federated properties: “same as” Wikidata, but not possible to create instance-specific properties
- Mappings: map properties and items to Wikidata. Does not “bring over” statements, dependent properties, etc.

Summary: WBStack vs. Wikidata

- Social infrastructure: by default, users have much more power in WBStack (e.g. unfettered property creation) than in Wikidata; this can be adjusted, though
- WBStack can be kept “closed,” in the sense that accounts must be requested
- Federated properties: all or none but direct match to Wikidata; otherwise mappings can be used
- Using default features like the Query Service and Quick Statements require some minor adaptations to accommodate different P and Q numbers, different URIs
- Other tools like OpenRefine can be customized to work with WBStack instances, but this requires more effort
- Unlike a Wikibase instance, WBStack can be accessed by ANYONE with the link