Wikibase as an IR authority file

IR ← Wikibase ↔ Wikidata
Presenters

Joe Cera
Digital initiatives and scholarly communications librarian
Berkeley Law Library
cerax@berkeley.edu

Michael Lindsey
Web developer, magician
Berkeley Law Library
mlindsey@berkeley.edu
Starting idea

Maintain repository authority files without conflicting with ILS authorities
Make this authority information linked data friendly
Minimize duplication of effort across projects and tasks
Do this with 2 people who have lots of other things to do
Background

- Law library with 150+ faculty to track
- Repository contains faculty publications, publications from 13 student-run journals, and other miscellany
- Wikidata QID as author ID, the hook into multiple other projects
- Operating in a scholarly environment where DOIs are not the norm
- Many peers manage a repository platform that uses email addresses as identifiers (?!)
- Process(es) currently managed through a collection of spreadsheets, scripts, and manual work
Technology

Repository: TIND IR in an environment that also uses TIND ILS (MARC-based)

Wikibase: ir-authorities hosted on wikibase.cloud

APIs and scripts:

- cURL (wrapped in PHP logic)
- Wikibase Action API
- Wikidata REST API
# Wikibase Action API

**action** = **wbcreateclaim** (as an example)

<table>
<thead>
<tr>
<th><strong>action=wbcreateclaim</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>**(main</td>
</tr>
<tr>
<td>Creates Wikibase claims.</td>
</tr>
<tr>
<td><strong>Parameters:</strong></td>
</tr>
<tr>
<td><strong>entity:</strong></td>
</tr>
<tr>
<td>This parameter is required.</td>
</tr>
<tr>
<td><strong>snaktype:</strong></td>
</tr>
<tr>
<td>This parameter is required.</td>
</tr>
<tr>
<td>One of the following values: nocvalue, somevalue, value</td>
</tr>
<tr>
<td><strong>property:</strong></td>
</tr>
<tr>
<td>This parameter is required.</td>
</tr>
<tr>
<td><strong>value:</strong></td>
</tr>
<tr>
<td><strong>summary:</strong></td>
</tr>
<tr>
<td><strong>tags:</strong></td>
</tr>
<tr>
<td>Values (separate with</td>
</tr>
<tr>
<td><strong>token:</strong></td>
</tr>
<tr>
<td>This parameter is required.</td>
</tr>
<tr>
<td><strong>baserevid:</strong></td>
</tr>
<tr>
<td>Type: integer</td>
</tr>
<tr>
<td><strong>bot:</strong></td>
</tr>
<tr>
<td>Type: boolean (details)</td>
</tr>
</tbody>
</table>
Wikidata REST API

Just a couple from this API while appreciating the potential
Institutional Repositories Authority Files

This WikiBase was created with the intent of exploring the potential for using the platform as a method of storing and making use of the data for repository authorities. The initial population of data is based on Berkeley Law individuals. As this is in the exploratory phase, the data (and platform) could disappear at any moment and rebuilt in another way.

Data models

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
<th>Value may be new entity?</th>
<th>Usage note</th>
</tr>
</thead>
<tbody>
<tr>
<td>instance of (P20)</td>
<td>human</td>
<td>no</td>
<td>this may include middle names and initials</td>
</tr>
<tr>
<td>Given Name (P14)</td>
<td>string value</td>
<td>yes</td>
<td>in the form of Last Name, First Name</td>
</tr>
<tr>
<td>Family Name</td>
<td>string value</td>
<td>yes</td>
<td>in the form of Last Name, First Name</td>
</tr>
<tr>
<td>IR Authority Name Form (P13)</td>
<td>value</td>
<td>yes</td>
<td>this should be just the QID from Wikidata and the link will auto-format</td>
</tr>
<tr>
<td>wikidata identifier (P2)</td>
<td>Wikidata QID</td>
<td>yes</td>
<td>this is the Library of Congress Name Authority number - the link will auto-format</td>
</tr>
<tr>
<td>LCNA (P6)</td>
<td>LCNAF number</td>
<td>yes</td>
<td>this is the associated ORCID identifier - the link will auto-format</td>
</tr>
<tr>
<td>ORCID (P22)</td>
<td>ORCID identifier</td>
<td>yes</td>
<td>this is an internal institutional identifier</td>
</tr>
<tr>
<td>person id (P7)</td>
<td>value</td>
<td>yes</td>
<td>this is the link to a person's faculty profile</td>
</tr>
<tr>
<td>SSRN author id (P24)</td>
<td>numeric SSRN</td>
<td>yes</td>
<td>this is the link to a person's CV</td>
</tr>
<tr>
<td>Faculty Profile (P18)</td>
<td>URL</td>
<td>yes</td>
<td>this is the link to a person's faculty profile</td>
</tr>
<tr>
<td>Curriculum Vitae (P19)</td>
<td>URL</td>
<td>yes</td>
<td>this is the position of the person within the institution</td>
</tr>
</tbody>
</table>

Query Examples

To explore the data in this wikiBase, you can start with some of these examples to see what exists as
Repository authority records

Fields included in the authority record for the repository:

7001_a: IR authority name form (P15 in ir-authorities)
7001_0: wikidata QID (P2 in ir-authorities)
7001_1: LCNA (P8 in ir-authorities <-> P244 in Wikidata)
7001_1: ORCiD (P22 in ir-authorities <-> P496 in Wikidata)
Repository authority records
ir-authorities wikibase to/from Wikidata

LCNA (P8 in ir-authorities <-> P244 in Wikidata)

ORCiD (P22 in ir-authorities <-> P496 in Wikidata)

SSRN author ID (P24 in ir-authorities <-> P3747 in Wikidata)

Faculty Profile (P18 in ir-authorities <-> P973 in Wikidata - described at URL)

Curriculum Vitae (P19 in ir-authorities <-> P8214 in Wikidata - curriculum vitae URL)
irAuthorities wikibase not shared

Position held (P23) - eventually shared as P39 in Wikidata, needs more thought due to the politics of position

Person ID (P7) - local identifier

IR Authority Name Form (P15) - name form used as authority in IR

Given name (P14) and Family Name (P13) - local version of names used in Berkeley Law materials
Process: New faculty member joins Berkeley Law

[**wikidata**] Create wikidata QID if it doesn’t already exist
Add affiliation, start time, orcid, lcnaf (the first two before the bots come)

[**ir auth**] Create record including QID, profile url, cv url

[**script**] Push and/or Pull between Wikidata and ir auth:
orcid, lcnaf identifiers, cv url, and profile url
update TIND IR authority file with new authority from ir-auth

[**implement**] depending on title/position: add to faculty publications list
Process: When existing faculty information changes

[ir-auth] Update changes (name, identifiers, cv url, profile url)

[script] Check Wikidata for conflicts
Pull relevant changes from Wikidata to ir-auth
Push relevant changes to Wikidata from ir-auth (caution with conflicts)

[script] Update TIND IR authority file if changes are relevant to content

[manual] Batch update IR, if changes are relevant

[to implement] on title/position change, update faculty publications list
Process and output examples

Updating Wikidata/ir-authorities/repository authorities in one script for one author:

Alan J. Auerbach (Wikidata / ir-authorities / repository authority)

>php syncAuthority.php Q36

Or multiple authors:

>for q in Q36 Q68 Q118 Q164; do php syncAuthority.php $q; done
First lines give an overview/summary of the operations:

- Logged in to https://ir-authorities.wikibase.cloud
- HTTP request to www.wikidata.org
- Pushing IR-Auth ORCID (P22) to WD as P496
- Lawcat link https://lawcat.berkeley.edu
- Updating existing LawCat authority record
More console output

Summary report of links and changes for each QID followed by a XML view of any updated repository authority as it was added.
Refinements made along the way

Shorter output to console (ability to view verbose version)

Verbose console output to log file

Clear links to Wikidata / ir-authorities / repository authority (unless New)

Clear list of properties being updated

Generalized process provides for easier properties to add/remove
Challenges

What to do about affiliation?
- Currently hard-coded but this presents other issues: faculty departure, deceased faculty, affiliation at time of publication

What to do about deceased faculty?
- Adding Wayback Machine archive of profile page

Latency issues
- Changes are not immediately available for API/SPARQL
The future of this project

[to implement]

Thank you!