

---

---

# 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)

### action=wbcreateclaim

[\(main | wbcreateclaim\)](#)

Creates Wikibase claims.

#### Parameters:

<b>entity:</b>	ID of the entity the claim is being added to This parameter is required.
<b>snaktype:</b>	The type of the snak This parameter is required. One of the following values: novalue, somevalue, value
<b>property:</b>	ID of the snaks property This parameter is required.
<b>value:</b>	Value of the snak when creating a claim with a snak that has a value
<b>summary:</b>	Summary for the edit. Will be prepended by an automatically generated comment. The length limit of the autocomment together with the summary is 260 characters. Be aware that everything above that limit will be cut off.
<b>tags:</b>	Change tags to apply to the revision. Values (separate with   or <a href="#">alternative</a> ): <a href="#">#temporary_batch_1652635906227</a> , <a href="#">Connecting-Senses</a> , <a href="#">Czech-Authorities-Sync</a> , <a href="#">InfoboxExport gadget</a> , <a href="#">WE-Framework gadget</a> , <a href="#">WikibaseJS-cli</a> , <a href="#">WikibaseJS-edit</a> , <a href="#">client-automatic-update</a> , <a href="#">client-linkitem-change</a> , <a href="#">data-bridge</a> , <a href="#">dataDrainer</a> , <a href="#">distributed-game</a> , <a href="#">gadget-addsenseimage</a> , <a href="#">gadget-draggablesitelinks</a> , <a href="#">gadget-dragndrop</a> , <a href="#">gadget-merge</a> , <a href="#">gadget-move</a> , <a href="#">gadget-moveclaim</a> , <a href="#">gadget-rearrangevalues</a> , <a href="#">openrefine-3.4</a> , <a href="#">openrefine-3.5</a> , <a href="#">openrefine-3.6</a> , <a href="#">openrefine-3.7</a> , <a href="#">openrefine-3.8</a> , <a href="#">openrefine-4.0</a> , <a href="#">reference-game</a> , <a href="#">svwiki-DescriptionGadget</a> , <a href="#">termbox</a> , <a href="#">view-it</a> , <a href="#">wikibase-patcher-v1</a> , <a href="#">wikidata-for-firefox</a> , <a href="#">wikidata-for-web</a> , <a href="#">wikidata-ui</a> , <a href="#">wikimedia-commons-app</a> , <a href="#">wwwyzzerrdd</a>
<b>token:</b>	A "csrf" token retrieved from <a href="#">action=query&amp;meta=tokens</a> This parameter is required.
<b>baserevid:</b>	The numeric identifier for the revision to base the modification on. This is used for detecting conflicts during save. Type: integer
<b>bot:</b>	Mark this edit as bot. This URL flag will only be respected if the user belongs to the group "Bots". Type: boolean ( <a href="#">details</a> )

- This module requires read rights.
- This module requires write rights.
- This module only accepts POST requests.
- Source: WikibaseClient
- License: [GPL-2.0-or-later](#)

# Wikidata REST API

Just a couple from [this API](#) while appreciating the potential

## items Wikibase Items

Wikibase Data Model - Items ^

**GET** /entities/items/{item\_id} Retrieve a single Wikibase Item by ID

## statements Wikibase Statements

Wikibase Data Model - Statements ^

**GET** /entities/items/{item\_id}/statements Retrieve Statements from an Item

**POST** /entities/items/{item\_id}/statements Add a new Statement to an Item

# ir-authorities

Item [Discussion](#) [Read](#) [View history](#) [More](#)

## Charles Weisselberg (Q164)

No description defined [edit](#)

Weisselberg, Charles D. | Weisselberg, Charles | Weisselberg, Charles D | Charles D Weisselberg | Charles D. Weisselberg

[In more languages](#)  
Configure

Language	Label	Description	Also known as
English	Charles Weisselberg	No description defined	Weisselberg, Charles D. Weisselberg, Charles Weisselberg, Charles D Charles D Weisselberg Charles D. Weisselberg

### Statements

Given Name	<span>Q164</span> Charles <a href="#">edit</a>
	<a href="#">- 0 references</a>
	<a href="#">+ add reference</a>
	<a href="#">+ add value</a>

Family Name	<span>Q164</span> Weisselberg <a href="#">edit</a>
	<a href="#">- 0 references</a>
	<a href="#">+ add reference</a>
	<a href="#">+ add value</a>

IR Authority Name Form	<span>Q164</span> Weisselberg, Charles D. <a href="#">edit</a>
	<a href="#">- 0 references</a>
	<a href="#">+ add reference</a>
	<a href="#">+ add value</a>

wikidata identifier	<span>Q164</span> <a href="#">Q63616752</a> <a href="#">edit</a>
	<a href="#">- 0 references</a>

Main Page [Discussion](#) [Read](#) [Edit](#) [View history](#) [More](#)

## Main Page

### Institutional Repositories Authority Files [\[edit\]](#)

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 [\[edit\]](#)

#### Data model for humans

Property	Value	Value may be new entity?	Usage note
instance of (P20)	human	no	
Given Name (P14)	string value	yes	this may include middle names and initials
Family Name (P13)	string value	yes	
IR Authority Name Form (P15)	value	yes	in the form of <b>Last Name, First Name</b>
wikidata identifier (P2)	Wikidata QID	yes	this should be just the QID from Wikidata and the link will auto-format
LCNA (P8)	LCNAF number	yes	this is the Library of Congress Name Authority number - the link will auto-format
ORCID (P22)	ORCID identifier	yes	this is the associated ORCID identifier - the link will auto-format
person ID (P7)	value	yes	this is an internal institutional identifier
SSRN author id (P24)	numeric SSRN author ID	yes	author id from SSRN - the link will auto-format
Faculty Profile (P18)	URL	yes	this is the link to a person's faculty profile
Curriculum Vitae (P19)	URL	yes	this is the link to a person's CV
position held (P23)	Item	no	this is the position of the person within the institution

### Query Examples [\[edit\]](#)

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

## Berkeley Law | Library Catalog

Result 1 of 1

Details	
Personal Name	Charles D Weisselberg UC Berkeley School of Law
Link to search	All resources
Record appears in	Authorities

 Edit Bib	 Formats	 Add to Basket
---	--	--

Format		
BibTeX	<a href="#">Q View</a>	<a href="#">Download</a>
MARC	<a href="#">Q View</a>	<a href="#">Download</a>
MARCXML	<a href="#">Q View</a>	<a href="#">Download</a>
DublinCore	<a href="#">Q View</a>	<a href="#">Download</a>
EndNote	<a href="#">Q View</a>	<a href="#">Download</a>
NLM	<a href="#">Q View</a>	<a href="#">Download</a>
RefWorks	<a href="#">Q View</a>	<a href="#">Download</a>
RIS	<a href="#">Q View</a>	<a href="#">Download</a>

```
--<collection>
--<record>
  <controlfield tag="001">1255616</controlfield>
  <controlfield tag="005">20230208154148.0</controlfield>
  <datafield tag="035" ind1=" " ind2=" " >
    <subfield code="a">https://www.wikidata.org/entity/Q63616752</subfield>
  </datafield>
  <datafield tag="100" ind1="1" ind2=" " >
    <subfield code="a">Charles D Weisselberg</subfield>
    <subfield code="u">UC Berkeley School of Law</subfield>
    <subfield code="1">http://id.loc.gov/authorities/names/no2019182936</subfield>
  </datafield>
  <datafield tag="980" ind1=" " ind2=" " >
    <subfield code="a">PERSONAL</subfield>
  </datafield>
  <datafield tag="980" ind1=" " ind2=" " >
    <subfield code="a">AUTHORITY</subfield>
  </datafield>
  <datafield tag="981" ind1=" " ind2=" " >
    <subfield code="a">IR</subfield>
  </datafield>
</record>
</collection>
```

# 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)

## ir-authorities 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
```

# Console output

First lines give an overview/summary of the operations

```
logged in to https://ir-authorities.wikibase.cloud (1e0254467284da953699a05a5ef570f464f768ea+\)  
[2023-09-05T10:44:12-07:00] http://www.wikidata.org/entity/Q63616752  
[2023-09-05T10:44:12-07:00] <https://ir-authorities.wikibase.cloud/entity/Q164>  
[2023-09-05T10:44:12-07:00] pushing ir-auth ORCID (P22) to WD as P496  
[2023-09-05T10:44:14-07:00] lawcat link https://lawcat.berkeley.edu/record/1255616  
[2023-09-05T10:44:14-07:00] updating EXISTING LawCat authority record  
<?xml version="1.0" encoding="UTF-8"?>
```



# 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.

```
[jcera@libdev] ~ 05:56:37 - /home/mlindsey/bin/ir/wikibase:
> for q in Q36 Q68 Q118 Q164; do php syncAuthority.php $q; done
logged in to https://ir-authorities.wikibase.cloud (7e1cc8632d03e76e2edbb6f569bcfe4864f8770d+)
[2023-09-06T05:56:47-07:00] http://www.wikidata.org/entity/Q29961382
[2023-09-06T05:56:47-07:00] <https://ir-authorities.wikibase.cloud/entity/Q36>
[2023-09-06T05:56:48-07:00] lawcat link https://lawcat.berkeley.edu/record/1255670
logged in to https://ir-authorities.wikibase.cloud (d9f3702f9d09662a70736904707dd9fc64f87713+)
[2023-09-06T05:56:52-07:00] http://www.wikidata.org/entity/Q63616476
[2023-09-06T05:56:52-07:00] <https://ir-authorities.wikibase.cloud/entity/Q68>
[2023-09-06T05:56:52-07:00] IMPLEMENT: overwriting WD profile (P973) with ir-auth profile (P18)
[2023-09-06T05:56:54-07:00] lawcat link https://lawcat.berkeley.edu/record/1255649
logged in to https://ir-authorities.wikibase.cloud (e139b5ecbfc2c1b8e3d82963c03e8c864f87718+)
[2023-09-06T05:56:57-07:00] http://www.wikidata.org/entity/Q63616503
[2023-09-06T05:56:57-07:00] <https://ir-authorities.wikibase.cloud/entity/Q118>
[2023-09-06T05:56:59-07:00] lawcat link https://lawcat.berkeley.edu/record/1255633
logged in to https://ir-authorities.wikibase.cloud (a22c46b0e60063b0e519ab8350b1474b64f8771d+)
[2023-09-06T05:57:03-07:00] http://www.wikidata.org/entity/Q63616752
[2023-09-06T05:57:03-07:00] <https://ir-authorities.wikibase.cloud/entity/Q164>
[2023-09-06T05:57:04-07:00] lawcat link https://lawcat.berkeley.edu/record/1255616
[2023-09-06T05:57:04-07:00] updating EXISTING LawCat authority record
[2023-09-06T05:57:04-07:00] curl -X POST \
'https://lawcat.berkeley.edu/api/v1/record?callback_email=jcera@law.berkeley.edu&mode=insertorreplace' -H 'Authorization: Tok
"UTF-8"?'>
<collection xmlns="http://www.loc.gov/MARC21/slim">
<record>
  <controlfield tag="001">1255616</controlfield>
  <datafield tag="035" ind1=" " ind2=" ">
    <subfield code="a">https://www.wikidata.org/entity/Q63616752</subfield>
  </datafield>
  <datafield tag="100" ind1="1" ind2=" ">
    <subfield code="a">Weisselberg, Charles D.</subfield>
    <subfield code="u">UC Berkeley School of Law</subfield>
    <subfield code="1">http://id.loc.gov/authorities/names/no2019182936</subfield>
    <subfield code="0">http://www.wikidata.org/entity/Q63616752</subfield>
    <subfield code="1">https://orcid.org/0009-0001-2365-6075</subfield>
  </datafield>
  <datafield tag="980" ind1=" " ind2=" ">
    <subfield code="a">PERSONAL</subfield>
  </datafield>
  <datafield tag="980" ind1=" " ind2=" ">
    <subfield code="a">AUTHORITY</subfield>
  </datafield>
  <datafield tag="981" ind1=" " ind2=" ">
    <subfield code="a">IR</subfield>
  </datafield>
</record>
</collection>
'
```

% Total	% Received	% Xferd	Average Speed	Time Dload	Time Upload	Time Total	Time Spent	Time Left	Current Speed
100	1147	0	141	100	1006	168	1203	--:--:--	1203
111	1006	0	118	100	1006	85	729	0:00:01	0

# 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!