

Reconciling and Extending Archival Authorities Data with Wikidata

Example: Archival authorities from *Unlocking the Airwaves*
Unlocking the Airwaves project

[Facet / Filter](#)[Undo / Redo 0 / 0](#)

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started?
[Watch these screencasts](#)

12 rowsShow as: [rows](#) [records](#)Show: [5](#) [10](#) **25** [50](#) [100](#) [500](#) [1000](#) rows[« first](#)[« previous](#)

1

of 1 page [next](#) [last »](#)

<input type="checkbox"/> All	<input checked="" type="checkbox"/> Name	<input type="checkbox"/> Entity Type
1.	National Association of Educational Broadcasters	Corporate Body
2.	WHA (Radio station : Madison, Wis.)	Corporate Body
3.	Mayer, Milton, 1908-1986	Person
4.	Fund for Adult Education (U.S.)	Corporate Body
5.	Gardiner, C. Harvey (Clinton Harvey)	Person
6.	Burrows, E. G. (Edwin Gladding), 1917-	Person
7.	KOAC (Radio/television station : Corvallis, Or.)	Corporate Body
8.	Harvard University. Russian Research Center	Corporate Body
9.	Harley, William G., 1911-1998	Person
10.	Marquis, Chalmers H.	Person
11.	Voice of America (Organization)	Corporate Body
12.	WETA (Radio/television station : Arlington, Va.)	Corporate Body

A representative set of CPF authorities from *Airwaves*, along with entity type. Entities from the *Airwaves* database frequently include parenthetical name additions, which can make reconciliation less effective if they are not trimmed off.

Facet / Filter

Undo / Redo 0 / 0

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started?
[Watch these screencasts](#)

Add column based on column Name

New column name

On error set to blank store error copy value from original column

Expression Language [General Refine Expression Language \(GREL\)](#) No syntax errors

```
value
```

Preview History Starred Help

row	value	value
1.	National Association of Educational Broadcasters	National Association of Educational Broadcasters
2.	WHA (Radio station : Madison, Wis.)	WHA (Radio station : Madison, Wis.)
3.	Mayer, Milton, 1908-1986	Mayer, Milton, 1908-1986
4.	Fund for Adult Education (U.S.)	Fund for Adult Education (U.S.)
5.	Gardiner, C. Harvey (Clinton Harvey)	Gardiner, C. Harvey (Clinton Harvey)
6.	Burrows, E. G. (Edwin Gladding), 1917-	Burrows, E. G. (Edwin Gladding), 1917-

OK Cancel

Create a column that duplicates the initial column, so that the starting names column is preserved for both reference and re-importing data back into the database later.

Facet / Filter

Undo / Redo 1 / 1

Using facets and filters



Use facets and filters to select subsets.

Create a text facet on the entity type column by going to Facet > Text facet. This will allow you to do different transformations on different subsets of the CPF authorities according to their type.

12 rows

Extensions: SNAC ▾ Wikidata ▾

Show as: [rows](#) [records](#) Show: [5](#) [10](#) [25](#) [50](#) [100](#) [500](#) [1000](#) rows

« first

< previous

1 of 1 page

next

last »

All	Name	Recon Name	Entity Type
1.	National Association of Educational Broadcasters	National Association of Educational Broadcasters	Facet
2.	WHA (Radio station : Madison, Wis.)	WHA (Radio station : Madison, Wis.)	Text filter
3.	Mayer, Milton, 1908-1986	Mayer, Milton, 1908-1986	Edit cells
4.	Fund for Adult Education (U.S.)	Fund for Adult Education (U.S.)	Edit column
5.	Gardiner, C. Harvey (Clinton Harvey)	Gardiner, C. Harvey (Clinton Harvey)	Transpose
6.	Burrows, E. G. (Edwin Gladding), 1917-	Burrows, E. G. (Edwin Gladding), 1917-	Sort...
7.	KOAC (Radio/television station : Corvallis, Or.)	KOAC (Radio/television station : Corvallis, Or.)	View
8.	Harvard University. Russian Research Center	Harvard University. Russian Research Center	Reconcile
9.	Harley, William G., 1911-1998	Harley, William G., 1911-1998	Corporate Body
10.	Marquis, Chalmers H.	Marquis, Chalmers H.	Corporate Body
11.	Voice of America (Organization)	Voice of America (Organization)	
12.	WETA (Radio/television station : Arlington, Va.)	WETA (Radio/television station : Arlington, Va.)	

Facet / Filter

Undo / Redo 1 / 1

5 matching rows (12 total)

Extensions: SNAC ▾ Wikidata ▾

Refresh

Reset All Remove All

Show as: [rows](#) [records](#) Show: [5](#) [10](#) **25** [50](#) [100](#) [500](#) [1000](#) rows

« first < previous 1 of 1 page next » last »

<input checked="" type="checkbox"/> All	<input checked="" type="checkbox"/> Name	<input checked="" type="checkbox"/> Recon Name	<input checked="" type="checkbox"/> Entity Type
3.	Mayer, Milton, 1908-1986	Mayer, Milton, 1908-1986	Person
5.	Gardiner, C. Harvey (Clinton Harvey)	Gardiner, C. Harvey (Clinton Harvey)	Person
6.	Burrows, E. G. (Edwin Gladding), 1917-	Burrows, E. G. (Edwin Gladding), 1917-	Person
9.	Harley, William G., 1911-1998	Harley, William G., 1911-1998	Person
10.	Marquis, Chalmers H.	Marquis, Chalmers H.	Person

Select one of the entity types to begin transforming the Recon Name column. Here, we are starting with entity type Person.

Facet / Filter

Undo / Redo 1 / 1

5 matching rows (12 total)

Extensions: SNAC ▾ Wikidata ▾

Refresh

Reset All Remove All

Entity Type change invert reset

2 choices Sort by: name count

Cluster

Corporate Body 7

Person 5

exclude

Facet by choice counts

Show as: [rows](#) [records](#) Show: 5 10 25 50 100 500 1000 rows

« first < previous 1 of 1 page next » last

All **Name**

		3. Mayer, Milton, 1908-1986
		5. Gardiner, C. Harvey (Clinton Harvey)
		6. Burrows, E. G. (Edwin Gladding), 1917
		9. Harley, William G., 1911-1998
		10. Marquis, Chalmers H.

Go to Edit cells > Transform... on the Recon Name column to begin editing the column.

Recon Name	Entity Type
Facet	Person
Text filter	Person
Edit cells	Transform...
Edit column	Common transforms
Transpose	Fill down
Sort...	Blank down
View	Split multi-valued cells...
Reconcile	Join multi-valued cells...
	Cluster and edit...
	Replace

Facet / Filter Undo / Redo 1 / 1

Refresh

Reset All

Entity Type

change in

2 choices Sort by: name count

Corporate Body 7

Person 5

Facet by choice counts

Custom text transform on column Recon Name

Expression

Language General Refine Expression Language (GREL) ▾

```
value.replace(/\s\(.*/,"").replace(/, [0-9]{4}-([0-9]{4})*/, "").match(/(.*) (.*)/).reverse().join(" ").trim()
```

No syntax error.

Preview

History

Starred

Help

row value

value.replace(/\s\(.*/,"").r .

3. Mayer, Milton, 1908-1986

Milton Mayer

5. Gardiner, C. Harvey (Clinton Harvey)

C. Harvey Gardiner

6. Burrows, E. G. (Edwin Gladding), 1917-

E. G. Burrows

9. Harley, William G., 1911-1998

William G. Harley

10. Marquis, Chalmers H.

Chalmers H. Marquis

On error

- keep original
- set to blank
- store error

 Re-transform up to times until no change

Use the GREL expression below to trim off parenthetical name additions and birth and death dates as well as reverse the order of the names to natural order.

```
value.replace(/\s\(.*/,"").replace(/, [0-9]{4}-([0-9]{4})*/, "").match(/(.*) (.*)/).reverse().join(" ").trim()
```

Facet / Filter Undo / Redo 2 / 2

Refresh Reset All Refresh

Entity Type change in

2 choices Sort by: name count

Corporate Body 7

Person 5

Facet by choice counts

Custom text transform on column Recon Name

Expression

Language General Refine Expression Language (GREL) ▾

```
value.replace(/\s\(.*\)/, "")
```

Preview

History

Starred

Help

row value

value.replace(/\s\(.*\)/, "")

1. National Association of Educational Broadcasters

National Association of Educational Bro

2. WHA (Radio station : Madison, Wis.)

WHA

4. Fund for Adult Education (U.S.)

Fund for Adult Education

7. KOAC (Radio/television station : Corvallis, Or.)

KOAC

8. Harvard University. Russian Research Center

Harvard University. Russian Research

11. Voice of America (Organization)

Voice of America

On error

- keep original
- set to blank
- store error

 Re-transform up to 10 times until no change

OK

Cancel

```
value.replace(/\s\(.*\)/, "")
```

Change the text facet to facet on entity type Corporate Body, then repeat use the GREL expression below to transform those cells by trimming off any parenthetical name additions, such as those that appear after radio/TV station names.

Facet / Filter

Undo / Redo 3 / 4

12 rows

Show as: rows records

Show: 5 10 25 50 100 500 1000 rows

Extensions: SNAC ▾ Wikidata ▾

Using facets and filters

If any names in your dataset are not transformed to a trimmed down form with the previous GREL expressions, manually edit them as needed.

		9. Harley, William G., 1911-1998
		10. Marquis, Chalmers H.
		11. Voice of America (Organization)
		12. WETA (Radio/television station : Arlington, Va.)

Then, start reconciling. On the column Recon Name, go to Reconcile > Start reconciling.

Recon Name	Entity Type
Facet	Educational Broadcasters
Text filter	Corporate Body
Edit cells	Person
Edit column	Corporate Body
Transpose	Person
Sort...	Corporate Body
View	Research Center
Reconcile	Corporate Body
Start reconciling	
Facets	Corporate Body
Actions	Corporate Body
Copy reconciliation data...	
Use values as identifiers	
Add entity identifiers column	



Facet / Filter

Undo / Redo

Using facets and filters

Use facets and filters to select parts of your data to act on. Choose and filter methods from the top of each data column.

Not sure how to get started?
[Watch these screencasts](#)

Reconcile column "Recon Name"

Reconcile each cell to an entity of one of these types:

- human
Q5
- scholarly article
Q13442814
- organization
Q43229
- radio station
Q14350
- sports league
Q623109
- foundation
Q157031
- chemical compound
Q11173
- governing body
Q5588651
- version_edition_or_translation

Reconcile against type:

Reconcile against no particular type

Auto-match candidates with high confidence

Maximum number of candidates to return

Add Standard Service...

Start Reconciling...

Cancel

Also use relevant details from other columns:

col
Nam
Enti

In the reconciliation window, change radio button to Reconcile against no particular type (since this dataset includes both humans and corporate bodies). If you are only reconciling the names of people, you can reconcile against type human for better accuracy. Additionally, you can check or uncheck the Auto-match option depending on whether you think Wikidata might correctly identify your entities. Then click Start reconciling...

12 rows

Auto-matched names, like National Association of Educational Broadcasters, will not show the reconciliation options, though you can select Choose new match if the match is incorrect. To inspect a match, hover over it to see the label, short description, and image. Click on the label to open the full Wikidata item page in a new tab. If the auto-match is accurate, no action is required.

v: 5 10 25 50 100 500 1000 rows « first < previous 1 of 1 page next » last »

Recon Name

National Association of Educational Broadcasters
Choose new match

WHA

- World Hockey Association (100)
- World Health Assembly (100)
- World History Association (100)
- WHA (100)
- Whamp station (100)
- World Haiku Association (100)
- west Hockey Association (100)
- Whaml (86)
- WhatsApp (80)
- Eurasian Curlew (75)
- wharf (75)
- whale (75)
- whaler (67)
- Yokohama DeNA BayStars (67)
- whaling (60)
- harbor (60)
- Tipulidae (60)
- Shinhwa (60)
- The Wharton School (60)
- Wha-Koo (60)
- alternate history (55)
- Cafe Wha? (55)
- brown rat (50)
- Whangārei (50)
- Whakatāne (50)
- Create new item

Search for match

Milton Mayer

- Milton Mayer (100)
- May Milton (91)
- May Milton (01)

Entity Type

National Association of Educational Broadcasters (Q17070604)
US organization of broadcasters with the aim of coordinating educational programs

WIKIDATA

		3.	Mayer, Milton, 1908-1986
https://www.wikidata.org/wiki/Q17070604			

Person

[Facet / Filter](#)[Undo / Redo](#) 4 / 4

12 rows

[Refresh](#)[Reset All](#) [Remove All](#)

For entities that don't auto-match, you will be presented with a list of match candidates. Hover over their labels to see some information, and click to open the full item page. Once you have identified the correct match, you can hover over the label and click Match this Cell, or click the single check mark to the left of the label. The double check mark corresponds to the Match All Identical Cells option.

National Association of Educational Broadcasters

WHA

- World Hockey Association (100)
- World Health Assembly (100)
- World History Association (100)
- WHA (10)
- Whampoa (10)
- World Hawa (10)
- West Hockey (10)
- Whaml (6)
- WhatsApp (6)
- Eurasian Cul (6)
- wharf (7)
- whale (7)
- whaler (6)
- Yokohama (6)
- DeNA BayStars (67)
- whaling (60)
- harbor (60)
- Tipulidae (60)
- Shinhwa (60)
- The Wharton School (60)
- Wha-Koo (60)
- alternate history (55)
- Cafe Wha? (55)
- brown rat (50)
- Whangārei (50)
- Whakatāne (50)
- Create new item

Search for match

Milton Mayer

- Milton Mayer (100)
- May Milton (91)
- May Milton (91)

Entity Type

Corporate Body

Corporate Body

edit

Person

Match this Cell [Match All Identical Cells](#) [Cancel](#)

WHA (Q7950064)
Wisconsin Public Radio Ideas Network flagship station in Madison, Wisconsin, United States



Facet / Filter

Undo / Redo 8 / 8

12 rows

Extensions: SNAC ▾ Wikidata ▾

This is an example of a false match candidate. The item "The Russian Research Center at Harvard University" is not the item for the center itself, but rather, an article about the center. However, there is a true match for this item on Wikidata that is not offered here. It does not appear because the name of the center has changed, and Wikidata does not make the connection between the two. Clicking Search for match will only be effective if you already know the exact name. Otherwise, searching Wikidata directly in a new tab is more effective.

	50 100 500 1000 rows	« first < previous	1 of 1 page	next » last »
<input checked="" type="checkbox"/> Recon Name			<input checked="" type="checkbox"/> Entity Type	
Fund for Adult Education			Corporate Body	
Choose new match			Person	
C. Harvey Gardiner			Person	
Choose new match			Corporate Body	
Ed Burrows			Corporate Body	
Choose new match			Corporate Body	
KOAC			Corporate Body	
Choose new match			Person	
Harvard University. Russian Research Center			Person	
<input checked="" type="checkbox"/> The Russian Research Center at Harvard University (92)			Corporate Body	
<input checked="" type="checkbox"/> Create new item			Person	
Search for match			Person	
William G. Harley			Corporate Body	
Choose new match			Voice of America	
Chalmers H. Marquis			<input checked="" type="checkbox"/> Voice of America (100)	
<input checked="" type="checkbox"/> Create new item			<input checked="" type="checkbox"/> Voice of America (100)	
Search for match			<input checked="" type="checkbox"/> Voice of America (100)	
Voice of America			<input checked="" type="checkbox"/> The Voice of America (89)	
<input checked="" type="checkbox"/> Voice of America (100)			<input checked="" type="checkbox"/> The Voice of America (89)	
<input checked="" type="checkbox"/> Voice of America (100)			<input checked="" type="checkbox"/> Voice of America Park (86)	
<input checked="" type="checkbox"/> Voice of America (100)			<input checked="" type="checkbox"/> Slavic Voice of America (82)	
<input checked="" type="checkbox"/> The Voice of America (89)			<input checked="" type="checkbox"/> Voice of America Indonesia (76)	
<input checked="" type="checkbox"/> The Voice of America (89)			<input checked="" type="checkbox"/> Voice of America Jazz Hour (76)	
<input checked="" type="checkbox"/> Voice of America Park (86)			<input checked="" type="checkbox"/> VOA-PNN (60)	
<input checked="" type="checkbox"/> Slavic Voice of America (82)			<input checked="" type="checkbox"/> YOUR VOICE OF AMERICA #5, 6 (NAID 119202) (60)	
<input checked="" type="checkbox"/> Voice of America Indonesia (76)			<input checked="" type="checkbox"/> The voice of America is coming of age (60)	
<input checked="" type="checkbox"/> Voice of America Jazz Hour (76)			<input checked="" type="checkbox"/> Voice of America Bethany Relay Station (59)	
<input checked="" type="checkbox"/> VOA-PNN (60)			<input checked="" type="checkbox"/> VOICE OF AMERICA PRE SPOTS (NAID 122323) (59)	
<input checked="" type="checkbox"/> YOUR VOICE OF AMERICA #5, 6 (NAID 119202) (60)			<input checked="" type="checkbox"/> VOICE OF AMERICA NEWS SPOTS (NAID 124800) (58)	
<input checked="" type="checkbox"/> The voice of America is coming of age (60)			<input checked="" type="checkbox"/> VOICE OF AMERICA ANIMAL FARM (NAID 106218) (57)	
<input checked="" type="checkbox"/> Voice of America Bethany Relay Station (59)				
<input checked="" type="checkbox"/> VOICE OF AMERICA PRE SPOTS (NAID 122323) (59)				
<input checked="" type="checkbox"/> VOICE OF AMERICA NEWS SPOTS (NAID 124800) (58)				
<input checked="" type="checkbox"/> VOICE OF AMERICA ANIMAL FARM (NAID 106218) (57)				



Item Discussion

Read View history

Search Wikidata



Davis Center for Russian and Eurasian Studies, Harvard University (Q45137420)

academic institution in United States of America

edit

Harvard University Russian Research Center

▼ In more languages

Configure

Language

Label

Description

Also known as

United States of America Harvard University Russian Res...

After manual searching, I find the correct item, which does have the alternative label Harvard University Russian Research Center. I can copy the precise label and use it to locate the item in OpenRefine.

Statements

instance of

academic institution

edit

▶ 1 reference

+ add value

Main page
Community portal
Project chat
Create a new Item
Recent changes

Random
Query
Nearby
Help
Donate

Lexicon
Create
Recent
Random

Tools

What links here
Related changes
Special pages
Permanent link
Page information
Concept URI
Cite this page

Facet / Filter

Undo / Redo 8 / 8

12 rows

Extensions: SNAC ▾ Wikidata ▾

Refresh

Reset All Remove All

Show as: rows records Show: 5 10 25 50 100 500 1000 rows

< first < previous 1 of 1 page next > last >

All Name

Recon Name

Entity Type

Paste the exact label into the Search for match box, which should then make the correct item appear for matching.

Search for match

Search for "Harvard University. Russian Research Center"

-
- Match other cells with same content
-
-
- Match this cell only

Studies, Harvard University

Select an item from the list:

Davis Center for Russian and Eurasian Studies Q45137420
academic institution in United States of America

Match

New Item

Don't Reconcile Cell

Cancel

- Voice of America (100)
- Voice of America (100)
- The Voice of America (89)
- The Voice of America (89)
- Voice of America Park (86)
- Slavic Voice of America (82)
- Voice of America Indonesia (76)
- Voice of America Jazz Hour (76)
- VOA-PNN (60)
- YOUR VOICE OF AMERICA #5, 6 (NAID 119202) (60)
- The voice of America is coming of age (60)
- Voice of America Bethany Relay Station (59)
- VOICE OF AMERICA PRE SPOTS (NAID 122323) (59)
- VOICE OF AMERICA NEWS SPOTS (NAID 124800) (58)
- VOICE OF AMERICA ANIMAL FARM (NAID 106218) (57)

Facet / Filter

Undo / Redo 9 / 9

12 rows

Extensions: SNAC ▾ Wikidata ▾

Refresh

Reset All Remove All

 Recon Name: judgment change

2 choices Sort by: name count

matched 9

none 3

Facet

Another example of an entity for which OpenRefine does not automatically find a match. Here again, I know there is a Wikidata item for Chalmers Marquis, so I need to search for it.

Search for match

Search for "Chalmers H. Marquis"

- Match other cells with same content
 Match this cell only

Match

New Item

Don't Reconcile

Chalmers H. Marquis

Select an item from the list:

No suggested matches

Tips on getting better suggestions:

- Enter more or fewer characters
- Add words related to your original search
- Try alternate spellings
- Check your spelling

All	Name	Recon Name	Entity Type
		Choose new match	Corporate Body
	Gardiner, C. Harvey (Clinton Harvey)	C. Harvey Gardiner	Person
	Ed Burrows, E. G. (Edwin Gladding), 1917-	Ed Burrows	Person
		Choose new match	Corporate Body
		Harvard University	Corporate Body
		Choose new match	Corporate Body
		Choose new match	Person
		Choose new match	Corporate Body
		Choose new match	Person
		Choose new match	Corporate Body

Facet / Filter

Undo / Redo 9 / 9

12 rows

Extensions: SNAC ▾ Wikidata ▾

Refresh

Reset All Remove All

Show as: rows records Show: 5 10 25 50 100 500 1000 rows

« first < previous 1 of 1 page next » last »

Recon Name: judgment change
All Name

Recon Name

Entity Type

2 choices Sort by: name_count

Removing the "H." from the name results in the item being found. Sometimes OpenRefine and Wikidata can match entities with the middle initial left in the name, but other times it is necessary to search without the initial.

Search for match

Search for "Chalmers H. Marquis"

-
- Match other cells with same content
-
-
- Match this cell only

American public television and radio advocate

Chalmers Marquis

Select an item from the list:

Chalmers Marquis

American public television and radio advocate

Q97155867

Voice of America (100)
The Voice of America (89)
The Voice of America (89)
Voice of America Park (86)
Slavic Voice of America (82)
Voice of America Worldwide (75)
Voice of America Jazz Hour (76)
VOA-PNN (60)
YOUR VOICE OF AMERICA #5, 6 (NAID 119202) (60)
The voice of America is coming of age (60)
Voice of America Bethany Relay Station (59)
VOICE OF AMERICA PRE SPOTS (NAID 122323) (59)
VOICE OF AMERICA NEWS SPOTS (NAID 124800) (58)
VOICE OF AMERICA ANIMAL FARM (NAID 106218) (57)
VOICE OF AMERICA AIRSHOW NEWS (NAID 121416) (56)
BACKGROUND VOICE OF AMERICA #23 (NAID 121107) (55)

Facet / Filter

Undo / Redo 12 / 12

12 rows

Show as: rows records

Show: 5 10 25 50 100 500 1000 rows

Extensions: SNAC ▾ Wikidata ▾

Using facets and filters

The complete reconciled dataset.
Here, all entities had a match.
However, you can choose not to reconcile any entities that don't have a match, so that they will be skipped when it comes time to extend the dataset with statements from Wikidata.

All Name

	Recon Name	Entity Type
	National Association of Educational Broadcasters Choose new match	Corporate Body
	WHA Choose new match	Corporate Body
	Milton Mayer Choose new match	Person
	Fund for Adult Education Choose new match	Corporate Body
	C. Harvey Gardiner Choose new match	Person
	Ed Burrows Choose new match	Person
	KOAC Choose new match	Corporate Body
	Davis Center for Russian and Eurasian Studies, Harvard University Choose new match	Corporate Body
	William G. Harley Choose new match	Person
	Chalmers Marquis Choose new match	Person
	Voice of America Choose new match	Corporate Body
	WETA Choose new match	Corporate Body

11. Voice of America (Organization)

12. WETA (Radio/television station : Arlington, Va)

Facet / Filter

Undo / Redo 12 / 12

12 rows

Show as: [rows](#) [records](#) Show: 5 10 25 50 100 500 1000 rowsExtensions: [SNAC](#) [Wikidata](#)[All](#) [Name](#)

Using facets and filters

Before extending the dataset, pull in the Wikidata IDs of your reconciled items. This is the unique identifier that can also be used to form the URL to the Wikidata item. Go to Edit column > Add column based on this column...

		9.	Harley, William G., 1911-1998
		10.	Marquis, Chalmers H.
		11.	Voice of America (Organization)
		12.	WETA (Radio/television station : Arlington, Va

Recon Name	Entity Type
Facet	
Text filter	
Edit cells	
Edit column	Split into several columns...
	Join columns...
Sort...	Add column based on this column...
View	Add column by fetching URLs...
Reconcile	Add columns from reconciled values...
	Rename this column
	Remove this column
	Move column to beginning
	Move column to end
	Move column left
	Move column right

[Facet / Filter](#)[Undo / Redo](#) 12 / 12

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started?
[Watch these screencasts](#)

Add column based on column Recon Name

New column name On error set to blank store error copy value from original column**Expression**`cell.recon.match.id`Language [General Refine Expression Language \(GREL\)](#)[Preview](#)[History](#)[Starred](#)[Help](#)**row value****cell.recon.match.id**

1.	National Association of Educational Broadcasters	Q17070604
2.	WHA	Q7950064
3.	Milton Mayer	Q6861322
4.	Fund for Adult Education	Q107621618
5.	C. Harvey Gardiner	Q107621610
6.	E. G. Burrows	Q103208872

[OK](#)[Cancel](#)

Use the GREL expression below to create a new column that has the Wikidata IDs as values.

`cell.recon.match.id`

Facet / Filter

Undo / Redo 13 / 13

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how
Watch these

The completed new column with Wikidata IDs.

12 rowsShow as: **rows** records

Show: 5 10 25 50 100 500 1000 rows

< first < previous 1 of 1 page next > last >

Extensions: SNAC ▾ Wikidata ▾

	Name	Recon Name	Wikidata ID	Entity Type
1.	National Association of Educational Broadcasters	National Association of Educational Broadcasters Choose new match	Q17070604	Corporate Body
2.	WHA (Radio station : Madison, Wis.)	WHA Choose new match	Q7950064	Corporate Body
3.	Mayer, Milton, 1908-1986	Milton Mayer Choose new match	Q6861322	Person
		Fund for Adult Education Choose new match	Q107621618	Corporate Body
		C. Harvey Gardiner Choose new match	Q107621610	Person
		Ed Burrows Choose new match	Q103208872	Person
7.	KOAC (Radio/television station : Corvallis, Or.)	KOAC Choose new match	Q6334739	Corporate Body
8.	Harvard University. Russian Research Center	Davis Center for Russian and Eurasian Studies, Harvard University Choose new match	Q45137420	Corporate Body
9.	Harley, William G., 1911-1998	William G. Harley Choose new match	Q105441878	Person
10.	Marquis, Chalmers H.	Chalmers Marquis Choose new match	Q97155867	Person
11.	Voice of America (Organization)	Voice of America Choose new match	Q228389	Corporate Body
12.	WETA (Radio/television station : Arlington, Va.)	WETA Choose new match	Q7949066	Corporate Body

Facet / Filter

Undo / Redo 13 / 13

12 rows

Show as: rows records Show: 5 10 25 50 100 500 1000 rows

Extensions: SNAC ▾ Wikidata ▾

Using facets and filters

Now you can (optionally) extend your data with selected statements about your entities from Wikidata. Go to Edit column > Add columns from reconciled values...

Recon Name	Wikidata ID	Entity Type
Facet		Corporate Body
Text filter		Corporate Body
Edit cells		Person
Edit column	Split into several columns... 322	Person
	Join columns... 2161	Corporate Body
Transpose		Person
Sort...	Add column based on this column... 2161	Person
View	Add column by fetching URLs... 0887	Person
Reconcile	Add columns from reconciled values... 739	Corporate Body
	Choose new match	Corporate Body
Davis Center for Russian Research Center	7420	Corporate Body
	Choose new match	Person
William G. Harley	4187	Person
	Choose new match	Person
Chalmers Marquis	5867	Corporate Body
	Choose new match	Corporate Body
Voice of America	89	Person
	Choose new match	Corporate Body
WETA	066	Corporate Body
	Choose new match	Corporate Body

Facet / Filter

Undo / Redo 13 / 13

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started?
[Watch these screencasts](#)

Add columns from reconciled column Recon Name

Add Property

Preview

Reset

Len

Select an item from the list:

SPARQL: Len
property path

Len

length P2043
measured dimension of an objectduration P2047
length of time of an event or processtype of lens P9597
type of lens used in a lighthouse, telescope, camera, etc.longest span P2787
length of the longest span of a bridge, a natural archLens ID P7100
identifier for documents in the Lens.org catalog (patents, articles)language of work or name P407
language associated with this creative work (such as books, show

OK Cancel

To pull labels, descriptions, and alt labels, special case-sensitive property code values are needed: "Len", "Aen", and "Den". Type these in the Add Property box, then click on the presented option of SPARQL: [Property code]. These values require special codes because they do not have P-codes like most Wikidata statements.

Facet / Filter

Undo / Redo 13 / 13

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menu at the top of each data column.

Not sure how to get started?
[Watch these screencasts](#)

For standard Wikidata properties, you can use the P-code, such as P31, to find the property you want, and then click on it in the resulting list of options. The preview window will show you a sample of the data from each property you select.

Add columns from reconciled column Recon Name

Add Property

Preview

Reset

P31

Recon

Len

Select an item from the list:

instance of

that class of which this subject is a particular example and member

ve configure
remove configAen
remove configDen
remove config

Educational Broadcasters			coordinating educational programs
WHA	WHA		Wisconsin Public Radio Ideas Network flagship station in Madison, Wisconsin, United States
Milton Mayer	Milton Mayer	Milton Sanford Mayer	American journalist
Fund for Adult Education	Fund for Adult Education		subsidiary foundation of the Ford Foundation
C. Harvey Gardiner	C. Harvey Gardiner		radio broadcaster
Ed Burrows	Ed Burrows	E. G. "Ed" Burrows	radio program director and producer
		Edwin Gladding Burrows	

OK Cancel

Extensions: SNAC ▾ Wikidata ▾

1 of 1 page next last

a ID	Entity Type
1	Corporate Body
2	Corporate Body
3	Person
4	Corporate Body
5	Person
6	Person
7	Corporate Body
8	Corporate Body
9	Person
10	Person
11	Corporate Body
12	Corporate Body
13	Corporate Body
14	Corporate Body

Facet / Filter

Undo / Redo 13 / 13

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menu at the top of each data column.

Not sure how to get started?
[Watch these screencasts](#)

Add columns from reconciled column Recon Name

Add Property

Preview

Reset

date of birth

Recon

Len

Aen

more config

Den

remove config

instance of

remove config

Select an item from the list:

date of birth

P569

date on which the subject was born

EB

US organization of
broadcasters with

organization

date on which the subject was born

If you don't know the property code, you can also just search for property names or keywords that you want to pull, then select a property from the resulting list of options.

Milton Mayer

Fund for Adult
Education

C. Harvey Gardiner

Ed Burrows

Milton Sanford
Mayercoordinating
educational
programsWisconsin Public
Radio Ideas
Network flagship
station in Madison,
Wisconsin, United
States

radio station

radio station

radio station

subsidiary
foundation of the
Ford Foundation

foundation

radio broadcaster

human

radio program
director and
producer

human

OK Cancel

Extensions: SNAC ▾ Wikidata ▾

1 of 1 page next last »

ID	Entity Type
1	Corporate Body
2	Corporate Body
3	Person
4	Corporate Body
5	Person
6	Person
7	Corporate Body
8	Corporate Body
9	Person
10	Person
11	Corporate Body
12	Corporate Body

Facet / Filter

Undo / Redo 13 / 13

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started?
[Watch these screencasts](#)

Add columns from reconciled column Recon Name

Add Property

Preview

Reset

Library of Congress authc

Suggested Properties

Qid

Recon Name	Len	Aen	Den	instance of
National Association of Educational Broadcasters	National Association of Educational Broadcasters	NAEB	US organization of broadcasters with the aim of coordinating educational programs	organization
WHA	WHA		Wisconsin Public Radio Ideas Network flagship station in Madison, Wisconsin, United States	radio station
Milton Mayer	Milton Mayer	Milton Sanford Mayer	American journalist	human
for Adult Education			subsidiary foundation of the Ford Foundation	foundation
Survey Gardiner			radio broadcaster	human
Ed Burrows	Ed Burrows	E. G. "Ed" Burrows	radio program director and	human

Click OK, then wait for OpenRefine to pull in the data from Wikidata.

OK Cancel

Extensions: SNAC ▾ Wikidata ▾

1 of 1 page next last »

a ID	Entity Type
1	Corporate Body
2	Corporate Body
3	Person
4	Corporate Body
5	Person
6	Person
7	Corporate Body
8	Corporate Body
9	Person
10	Person
11	Corporate Body
12	Corporate Body
13	Person
14	Corporate Body
15	Corporate Body



Facet / Filter

Undo / Redo 14 / 14

21 rows

Extensions: SNAC ▾ Wikidata ▾

Show as: [rows](#) [records](#)

Show: 5 10 25 50 100 500 1000 rows

« first

< previous

1 of 1 page

next >

last »

The resulting dataset after extending. The data can be exported and used as is here, but in order to import it into the Airwaves database, there are further processing steps required.

All	Name	Recon Name	Len	Aen	Den	instance of	date of birth	occupation	inception	SNAC ARK ID	Wikidata ARK ID
1.	National Association of Educational Broadcasters	National Association of Educational Broadcasters Choose new match	National Association of Educational Broadcasters	NAEB	US organization of broadcasters with the aim of coordinating educational programs	organization Choose new match			1925-01-01T00:00:00Z	w6rn70pt	n50
2.	WHA (Radio station : Madison, Wis.)	WHA Choose new match	WHA		Wisconsin Public Radio Ideas Network flagship station in Madison, Wisconsin, United States	radio station Choose new match			1920-01-01T00:00:00Z	w6pc89z4	n9-
3.	Mayer, Milton, 1908-1986	Milton Mayer Choose new match	Milton Mayer	Milton Sanford Mayer	American journalist	human Choose new match	1908-08-24T00:00:00Z	journalist Choose new match		w6xw4ndx	n50
4.	Fund for Adult Education (U.S.)	Fund for Adult Education Choose new match	Fund for Adult Education		subsidiary foundation of the Ford Foundation	foundation Choose new match			1951-01-01T00:00:00Z	w6382349	n8-
5.	Gardiner, C. Harvey (Clinton Harvey)	C. Harvey Gardiner Choose new match	C. Harvey Gardiner		radio broadcaster	human Choose new match	1913-01-01T00:00:00Z	broadcaster Choose new match		w6225vc9	n50
6.	Burrows, E. G. (Edwin Gladding), 1917-	Ed Burrows Choose new match	Ed Burrows	E. G. "Ed" Burrows	radio program director and producer	human Choose new match	1917-07-23T00:00:00Z	news presenter Choose new match		w67946bk	n78
7.				Edwin Gladding	announcer Choose new match						

Facet / Filter

Undo / Redo 14 / 27

Using facets and filters

The first step is to wrap values containing a comma in quotation marks so that they are preserved as one string when they are imported into Airtable (the *Airwaves* database platform). Go to Edit cells > Transform...

21 rows

Show as: rows records

Show: 5 10 25 50 100 500 1000 rows

< first < previous 1 of 1 page next > last >

Extensions: SNAC ▾ Wikidata ▾

All Name Recon Name Len

1 National National Association of Educational Broadcasters

new match

WHA

new match

Mayer

new match

Milton Mayer

Milton

Sanford

Mayer

American

journalist

human

Choose new match

Fund for Adult Education (U.S.)

Fund for Adult Education

Choose new match

C. Harvey Gardiner

Choose new match

Ed Burrows

Choose new match

Edwin Gladding

Choose new match

Aen Den instance of date of birth

Facet

Text filter

Edit cells

Edit column

Transpose

Sort...

View

Reconcile

Madison, Wisconsin, United States

cluster and edit...

Replace

human

Choose new match

foundation

Choose new match

radio broadcaster

human

Choose new match

radio program director and producer

human

Choose new match

news presenter

Choose new match

announcer

Choose new match

1925-01-01T00:00:00Z w6rn70pt n5

1920-01-01T00:00:00Z w6pc89z4 n9

1951-01-01T00:00:00Z w6382349 n8

w6225vc9 n5

w67946bk n7

Facet / Filter

Undo / Redo 14 / 27

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started?
[Watch these screencasts](#)

Custom text transform on column Aen

Expression

Language General Refine Expression Language (GREL) ▾

```
if(value.contains(",","\""+value+"\\"",value))
```

Preview

History

Starred

Help

row value

if(value.contains(",","\""+va ...

1. NAEB

NAEB

2. null

null

3. Milton Sanford Mayer

Milton Sanford Mayer

4. null

null

5. null

null

6. E. G. "Ed" Burrows

E. G. "Ed" Burrows

On error

- keep original
- set to blank
- store error

 Re-transform up to times until no change

OK

```
if(value.contains(",","\""+value+"\\"",value))
```

Use the GREL expression below to wrap values containing a comma in quotes.

Facet / Filter

Undo / Redo 15 / 15

21 rows

Extensions: SNAC ▾ Wikidata ▾

Next, join any properties with multiple statements for the same item into a single cell. This will ensure that the number of rows in your OpenRefine project is equal to the number of items you have, instead of having some blank rows to hold the second value for a particular property. Go to Edit cells > Join multi-valued cells... and use the default separator of ", "

5.	Len	Aen	Den	instance of	date of birth	occupation	inception	SNAC ARK ID					
	National Association of Educational Broadcasters	organization	Choose new match						1925-01-01T00:00:00Z	w6rn70pt	n50		
	WHA	Edit cells		Transform...					1920-01-01T00:00:00Z	w6pc89z4	n9-		
	Milton Mayer	Edit column		Common transforms									
		Transpose		Fill down									
		Sort...		Blank down									
		View		Split multi-valued cells...									
		Reconcile		Join multi-valued cells...									
				Cluster and edit...									
				Replace									
4.	Fund for Adult Education (U.S.)	Fund for Adult Education	Choose new match	Fund for Adult Education									
					subsidiary foundation of the Ford Foundation								
5.	Gardiner, C. Harvey (Clinton Harvey)	C. Harvey Gardiner	Choose new match	C. Harvey Gardiner	radio broadcaster	human	1913-01-01T00:00:00Z	broadcaster				w6225vc9	n50
6.	Burrows, E. G. (Edwin Gladding), 1917-	Ed Burrows	Choose new match	Ed Burrows	E. G. "Ed" Burrows	radio program director and producer	human	1917-07-23T00:00:00Z	news presenter			w67946bk	n78
7.				Edwin Gladding Rimmer					announcer				

Facet / Filter

Undo / Redo 16 / 16

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

This results in multiple statements attached to the same item being joined on a comma, instead of being split into separate cells. Repeat the process of wrapping in quotes and joining multi-valued cells for all columns containing commas and/or with multi-valued cells.

21 rows

Extensions: SNAC ▾ Wikidata ▾

Show as: [rows](#) [records](#) Show: 5 10 25 50 100 500 1000 rows

« first < previous 1 of 1 page next » last »

All	Name	Recon Name	Len	Aen	Den	instance of	date of birth	occupation	inception	SNAC ARK ID	...
				Choose new match							
6.	Burrows, E. G. (Edwin Gladding), 1917-	Ed Burrows Choose new match	Ed Burrows	E. G. "Ed" Burrows, Edwin Gladding Burrows	radio program director and producer	human Choose new match	1917-07- 23T00:00:00Z	news presenter Choose new match		w67946bk	n7:
								announcer Choose new match			
								manager Choose new match			
								broadcasting executive Choose new match	edit		
								poet Choose new match			
		KOAC		radio station in Oregon	radio station Choose new match				1922-12- 07T00:00:00Z	w6895kh2	no:
		Davis Center for Russian and Eurasian Studies, Harvard University	Harvard University Russian Research Center	academic institution in United States of America	academic institution Choose new match				1948-01- 01T00:00:00Z	w6fr50kz	n8:
		William G. Harley	William Harley, William Gardner Harley	American broadcasting executive	human Choose new match	1911-10- 09T00:00:00Z	broadcasting executive Choose new match			w6fx87sd	n9:
								consultant Choose new match			
		14.						executive Choose new match			
		15.	Mannus	Chalmers	Chalmers	American	human	1926-11-		w8nn0hr5	

Facet / Filter

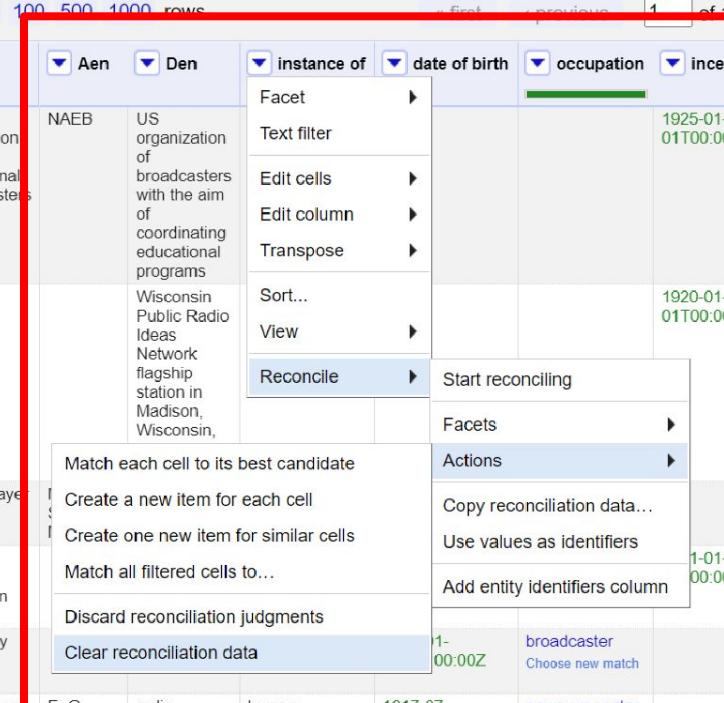
Undo / Redo 18 / 18

18 rows

Show as: [rows](#) [records](#) Show: 5 10 25 50 100 500 1000 rows[All](#) [Name](#) [Recon Name](#) [Len](#)[Aen](#) [Den](#) [instance of](#) [date of birth](#) [occupation](#) [inception](#)[SNAC ARK ID](#)

Using facets and filters

For any columns that you have transformed where the values are reconciled to Wikidata (e.g., are blue), go to Reconcile > Actions > Clear reconciliation data to clear out the reconciliation and make the earlier transformations take effect.





Facet / Filter

Undo / Redo 19 / 19

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started?
[Watch these screencasts](#)

After clearing reconciliation data, the cells should appear properly transformed and joined, as intended.

18 rows

Show as: [rows](#) [records](#)Show: [5](#) [10](#) **25** [50](#) [100](#) [500](#) [1000](#) rowsExtensions: [SNAC](#) [Wikidata](#)

		Name	Recon Name	Len	Aen	Den	instance of	date of birth	occupation	inception	SNAC ARK ID	
		Corvallis, Or.)										
12.	Harvard University. Russian Research Center	Davis Center for Russian and Eurasian Studies, Harvard University	Davis Center for Russian and Eurasian Studies, Harvard	Harvard University Russian Research Center	academic institution	United States of America	academic institution			1948-01-01T00:00:00Z	w6fr50kz	n8(
		Choose new match										
16.												
17.	Voice of America (Organization)	Voice of America Choose new match	Voice of America	VOA, VOA	United States international broadcaster	radio station, international broadcasting, online newspaper, 501(c)(4) organization	human	1942-02-01T00:00:00Z	broadcasting executive Choose new match		w6fx87sd	n9(
18.	WETA (Radio/television station : Arlington, Va.)	WETA Choose new match	WETA		classical music public radio station in Washington, D.C.	radio station	lobbyist Choose new match			1970-01-01T00:00:00Z	w6741pzk	not

Facet / Filter

Undo / Redo 22 / 22

12 records

Show as: [rows](#) [records](#) Show: [5](#) [10](#) [25](#) [50](#) [100](#) [500](#) [1000](#) records

Using facets

Use facets of your data and filter the top of

Not sure how to get started?

[Watch these screencasts](#)

Once you have processed all your extended data as desired, go to Export > Comma-separated value to download a CSV of the data.

2.			WHA (Radio station : Madison, Wls.)	WHA Choose new match	WHA		Wisconsin Public Radio Ideas Network flagship station in Madison, Wisconsin, United States	radio station					
3.			Mayer, Milton, 1908-1986	Milton Mayer Choose new match	Milton Mayer	Milton Sanford Mayer	American journalist	human	1908-08-24T00:00:00Z				
4.			Fund for Adult Education (U.S.)	Fund for Adult Education Choose new match	Fund for Adult Education		subsidiary foundation of the Ford Foundation	foundation					
5.			Gardiner, C. Harvey (Clinton Harvey)	C. Harvey Gardiner Choose new match	C. Harvey Gardiner		radio broadcaster	human	1913-01-01T00:00:00Z	broadcaster			
6.			Burrows, E. G. (Edwin Gladding), 1917-	Ed Burrows Choose new match	Ed Burrows	E. G. "Ed" Burrows, Edwin Gladding Burrows	radio program director and producer	human	1917-07-23T00:00:00Z	news presenter, announcer, manager, broadcasting executive, poet	w67946bk	n78	
7.			KNA&P	KNA&P	KNA&P		radio station	radio station			1922-12-28T00:00:00Z	W8805K9	no

OpenRefine project archive to file

Tab-separated value

Comma-separated value

HTML table

Excel (.xls)

Excel 2007+ (.xlsx)

ODF spreadsheet

Custom tabular exporter...

SQL Exporter...

Templating...

Upload to SNAC

Export to SNAC-Compliant JSON

OpenRefine project archive to Google Drive...

Google Sheets

Wikibase edits...

QuickStatements file

Wikibase schema

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Name	Recon Name	Len	Aen	Den	instance of	date of birth	occupation	inception	SNAC ARK ID	Library of Congress authority ID	Wikidata ID	Entity Type
2	National Association	National Assn	National As	NAEB		US organization	of organization		1925-01-0w6rn70pt	n50063712		Q17070604	Corporate Body
3	WHA (Radio station)	WHA	WHA		Wisconsin Public Ra	radio station			1920-01-0w6pc89z4	n91081816		Q7950064	Corporate Body
4	Mayer, Milton, 1908-	Milton Mayer	Milton May	Milton San	American journalist	human	1908-08-24T00:00:00Z	journalist	w6xw4ndx	n50007303		Q6861322	Person
5	Fund for Adult Education	Fund for Adu	Fund for Educati	subsidiary foundation					1951-01-0w6382349	n83029562		Q107621618	Corporate Body
6	Gardiner, C. Harvey	(C. Harvey Ga	C. Harvey Gardiner		radio broadcaster	human	1913-01-01T00:00:00Z	broadcaster	w6225vc9	n50015237		Q10762161C	Person
7	Ed Burrows, E. G. (Edwi	Ed Burrows	E. G. "Ed"		radio program direc	human	1917-07-23T00:00:00Z	news presenter, anno	w67946bk	n79139003		Q103208872	Person
8	KOAC (Radio/television	KOAC	KOAC		radio station in Ore	radio station			1922-12-0w6895kh2	no2011153873		Q6334739	Corporate Body
9	Harvard University, F	Davis Center	Davis Cente	Harvard U	academic institutio	academic institution			1948-01-0w6fr50kz	n80087594		Q45137420	Corporate Body
10	Harley, William G., 1911-2009	William G. Ha	William Ha	American broadcast	human	1911-10-09T00:00:00Z	broadcasting executiv	w6fx87sd	n92085222			Q105441878	Person
11	Marquis, Chalmers	Chalmers Ma	Chalmers Marquis		American public tele	human	1926-11-12T00:00:00Z	executive, lobbyist	w6qg9hd5			Q97155867	Person
12	Voice of America (Or Voice of Ameri	Voice of Am	VOA, VoA		United States intern	radio station, international broadcast	1942-02-0w6x103xw			n80051822		Q228389	Corporate Body
13	WETA (Radio/television	WETA	WETA		classical music publi	radio station			1970-01-0w6741pzk	no88001746		Q7949066	Corporate Body
14													
15													
16													
17													
18													
19													
20													
21													
22													
23													
24													
25													
26													
27													
28													
29													
30													
	CPF-Authorities-Reconciliation-												

The final CSV, in which each row corresponds to a single Wikidata item. This can be imported into Airtable easily.

Tips and concluding notes:

- This demonstration uses OpenRefine Version 3.5.0, so future versions of OR may diverge from the screencaps used here.

Resources:

- Wikidata's OpenRefine general guide:
<https://www.wikidata.org/wiki/Wikidata:Tools/OpenRefine>
- OpenRefine's Wikidata guides (generalizable to Wikibase):
<https://docs.openrefine.org/manual/wikibase/overview>
- OpenRefine reconciliation guide: <https://docs.openrefine.org/manual/reconciling>
- OpenRefine GREL reference: <https://docs.openrefine.org/manual/grelfunctions>