

Creating New Wikidata Items with OpenRefine

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

Facet / Filter Undo / Redo 6 / 6

5 rows

Extensions: SNAC▼ Wikidata▼

Show as: rows records Show: 5 10 25 50 rows

« first < previous 1 - 5 next > last »

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](#)

▼ All	▼ Name	▼ Entity Type	▼ SNAC ARK ID
☆	1. Experimental Discussion Project	Corporate Body	w60s7hgw
☆	2. WAFM-TV (Television Station : Birmingham, AL)	Corporate Body	w6qh555q
☆	3. WTOP-TV (Television Station : Washington, D.C.)	Corporate Body	w6mt0z6z
☆	4. Carter, Elmer A. (Elmer Anderson), 1890-1973	Person	w6pr9mh6
☆	5. Breitenfeld, F., Jr.	Person	w6sx7wx

Representative set of *Airwaves* CPF entities that have existing entries on Social Networks and Archival Context (SNAC), but not Wikidata.

Ensure items really lack existing Wikidata items. You will have already tried to reconcile these items, but you may find it useful to search for them on the Wikidata website, particularly if there are multiple possible labels.

Facet / Filter [Undo / Redo](#) 10 / 10

2 matching rows (5 total)

Extensions: [SNAC](#) [Wikidata](#)

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

« first « previous 1 - 2 next » last »

Filter:

0. Create project
1. Create new column SNAC ARK ID based on column Entity Type by filling 0 rows with `grel:null`
2. Edit single cell on row 1, column SNAC ARK ID
3. Edit single cell on row 2, column SNAC ARK ID
4. Edit single cell on row 3, column SNAC ARK ID
5. Edit single cell on row 4, column SNAC ARK ID
6. Edit single cell on row 5, column SNAC ARK ID

All

Name

Wikidata label

Entity Type

SNAC ARK ID

- 4. Carter, Elmer A.
- 5. Breitenfeld, F., Jr.

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 \(GREL\)](#)

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

No syntax error.

[Preview](#) [History](#) [Starred](#) [Help](#)

row	value	value.replace(/, [0-9]{4}-[0-9] ...
4.	Carter, Elmer A. (Elmer Anderson), 1890-1973	Elmer A. (Elmer Anderson) Carter
5.	Breitenfeld, F., Jr.	Jr. Breitenfeld, F.

Reserve names to first-last order, remove birth and death dates, and remove/rearrange other name parts, either using GREL expressions or manually. For the names shown here, the expression below is mostly effective (though still requires manual edits).

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

OpenRefine Airways Wikidata Workflow Documentation [Permalink](#) Open... Export Help

Facet / Filter Undo / Redo 10 / 10 **3 matching rows** (5 total) Extensions: SNAC Wikidata

Refresh Reset All Remove All Show as: **rows** records Show: 5 10 25 50 rows

Entity Type 2 choices Sort by: name **Corporate Body** 3 Person 2 Facet by choice counts

Custom text transform on column Wikidata label

Expression Language General Refine Expression Language (GREL) ▾

`value.replace(/\(.* : .*\)\/, "").trim()` No syntax error.

Preview History Starred Help

row	value	value.replace(/\(.* : .*\)\/, "" ...
1.	Experimental Discussion Project	Experimental Discussion Project
2.	WAFM-TV (Television Station : Birmingham, AL)	WAFM-TV
3.	WTOP-TV (Television Station : Washington, D.C.)	WTOP-TV

Remove corporate body name additions. For radio and television stations, the following GREL expression works:

```
value.replace(/\(.* : .*\)\/, "").trim()
```

Facet / Filter Undo / Redo 11 / 11

5 rows

Extensions: SNAC Wikidata

Show as: rows records Show: 5 10 25 50 rows

« first < previous 1 - 5 next > last »

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.

Start filling a new "instance of" column using a combination of GREL expressions and manual edits. Here, I am filling in "human" as the instance of value for all the items where the entity type is "Person". The GREL expression is:

Add column based on column Entity Type

New column name instance of

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

Expression Language General Refine Expression Language (GREL)

if(value.contains("Person"),"human","")

No syntax error.

Preview History Starred Help

row value

if(value.contains("Person"),"h ...

1. Corporate Body

2. Corporate Body

3. Corporate Body

4. Person

human

5. Person

human

OK

Cancel

`if(value.contains("Person"),"human","")`

Facet / Filter Undo / Redo 12 / 13

5 rows

Extensions: SNAC Wikidata

Show as: rows records Show: 5 10 25 50 rows

« first < previous 1 - 5 next > last »

Using facets and filters



Use facets and filters to select subsets

Another approach is to use custom text facets to narrow your view to only entities with "Television Station" in the name, for example, and then add television station to the instance of column for all the items that are shown in the faceted view. The GREL expression for the facet used here is:

Custom Facet on column Name

Expression Language General Define Expression Language (GREL) ▾

`value.contains("Television Station")` No syntax error.

Preview History Starred Help

row	value	value.contains("Television Sta ...
1.	Experimental Discussion Project	false
2.	WAFM-TV (Television Station : Birmingham, AL)	true
3.	WTOP-TV (Television Station : Washington, D.C.)	true
4.	Carter, Elmer A. (Elmer Anderson), 1890-1973	false
5.	Breitenfeld, F., Jr.	false

OK Cancel

`value.contains("Television Station")`



Facet / Filter

[Undo](#) / [Redo](#) 14 / 14

5 rows

Extensions: [SNAC](#) [Wikidata](#)Show as: [rows](#) [records](#) Show: 5 10 25 50 rows

« first < previous 1 - 5 next > last »

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](#)

All	Name	Wikidata label	Entity type	instance of	SNAC ARK ID
	1. Experimental Discussion Project	Experimental Discussion Project	Corporate body	project	w6d67hgw
	2. WAFM-TV (Television Station : Birmingham, AL)	WAFM-TV	Corporate body	television station	w6d6555q
	3. WTOP-TV (Television Station : Washington, D.C.)	WTOP-TV	Corporate body	television station	w6d60z6z
	4. Carter, Elmer A. (Elmer Anderson), 1890-1973	Elmer A. Carter	Person	human	w6d69mh6
	5. Breitenfeld, F., Jr.	F. Breitenfeld, Jr.	Person	human	w6d67wxs

Instance of property,
completed.

Facet / Filter Undo / Redo 20 / 20

5 rows

Extensions: SNAC▼ Wikidata▼

Show as: rows records Show: 5 10 25 50 rows

« first < previous 1 - 5 next > last »

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](#)

▼ All	▼ Name	▼ Wikidata label	▼ Entity Type	▼ instance of	▼ label description	▼ SNAC ARK ID
☆	1. Experimental Discussion Project	Experimental Discussion Project	Corporate Body	project	project of the Fund for Adult Education	w6s7hgw
☆	2. WAFM-TV (Television Station : Birmingham, AL)	WAFM-TV	Corporate Body	television station	television station in Birmingham, Alabama	w6h555q
☆	3. WTOP-TV (Television Station : Washington, D.C.)	WTOP-TV	Corporate Body	television station	television station in Washington, DC	w6t0z6z
☆	4. Carter, Elmer A. (Elmer Anderson), 1890-1973	Elmer A. Carter	Person	human	commissioner of the New York State Commission Against Discrimination	w6s9mh6
☆	5. Breitenfeld, F., Jr.	F. Breitenfeld, Jr.	Person	human	staff member at the National Association of Educational Broadcasters and Maryland Public Television	w6s7wxs

Create brief descriptions for each of your new items. This is the most manual part of the process, as there is not much way to automate it. If your items have SNAC records or other external authority records, like LCNAF, you may be able to copy or generate a description easily from those sources. For *Airwaves*, we also leaned on the primary sources in the collection to create descriptions for some entities.

Facet / Filter Undo / Redo 20 / 20

5 rows

Extensions: SNAC▼ Wikidata▼

Show as: rows records Show: 5 10 25 50 rows

« first < previous 1 - 5 next > last »

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](#)

Reconcile column "Wikidata label"

» Access Service API

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

Also use relevant details from other columns:

- ☐ item of collection or exhibition
Q18593264
- ☐ human
Q5
- ☐ painting
Q3305213
- ☐ natural watercourse
Q55659167

Column	Include?	As Property
Name	<input type="checkbox"/>	<input type="text"/>
Entity Type	<input type="checkbox"/>	<input type="text"/>
instance of	<input type="checkbox"/>	<input type="text"/>
label description	<input type="checkbox"/>	<input type="text"/>
SNAC ARK ID	<input type="checkbox"/>	<input type="text"/>

- ☐ 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

Reconcile Wikidata labels to Wikidata. Match against no particular type and do not auto-match, as we already know these labels do not have true matching items on Wikidata.

Facet / Filter Undo / Redo 21 / 21

Refresh Reset All Remove All

Wikidata label: judgment change

1 choices Sort by: name count

none 5

Facet by choice counts

Wikidata label: best candidate's score change reset

18.00 — 93.00

☒ Numeric 2 ☐ Non-numeric 0 ☒ Blank 3 ☐ Error 0

5 rows

Show as: rows records Show: 5 10 25 50 rows

Extensions: SNAC Wikidata

« first < previous 1 - 5 next > last »

	All	Name	Wikidata label	Entity Type	instance of	label description	SNAC ARK ID
☆		1. Experimental Discussion Project	Experimental Discussion Project Create new item Search for match	Corporate Body	project	project of the Fund for Adult Education	w60s7hgw
☆		2. WAFM-TV (Television Station : Birmingham, AL)	WAFM-TV Create new item Search for match				
☆		3. WTOP-TV (Television Station : Washington, D.C.)	WTOP-TV FRANK SHAKESPEARE BEING INTERVIEWED BY JOSEPH BENTI WTOP TV (NAID 106276) (18) JOHN CHANCELLOR AND HAROLD WALKER BROADCAST BY WTOP-TV IN THREE SEGMENTS 03/21, 03/22, 03/23/1966 (NAID 106252) (12) Create new item Search for match				
☆		4. Carter, Elmer A. (Elmer Anderson), 1890-1973	Elmer A. Carter Elmer Carter (92) Elmer Anderson Carter (80) The Carters of Elm Street (56) Elm Branch (50) Photograph of Elmer Carter (NAID 78088418) (48) Elm Springs Lake Dam (41) Elmer Spring (38) Dutch elm disease in Illinois (33) Chronic Standing Desk Use and Arterial Stiffness (32) Control of elm diseases in nursery elm plantings (29) Blended learning within an undergraduate exercise physiology laboratory. (28) Progress in the control of elm diseases in nurseries (27) Elm Hollow (25) The 2-hour marathon: what do students think? (25) Pennsylvanians interview new Surgeon General. Interview by John Whyte, Tom McCarter, Elmer Pinzon. (24) A report on tests of fungicides for the control of elm diseases in nurseries (24) An outside-the-box activity to demonstrate how humans and animals turn (24) Host predilection and transmissibility of vesicular stomatitis New Jersey virus strains in domestic cattle (Bos taurus) and swine (Sus scrofa). (22) In Vivo Neural Tissue Engineering: Cylindrical Biocompatible Hydrogels That Create New Neural Tracts in the Adult Mammalian Brain. (20) Combined Measurement of the Higgs Boson Mass in p p Collisions at s = 7 and 8 TeV with the ATLAS and CMS Experiments (18) Mechanical transmission of vesicular stomatitis New Jersey virus by Simulium vittatum (Diptera: Simuliidae) to domestic swine (Sus scrofa). (18) Combinations of single-top-quark production cross-section measurements and J/ψVtbj determinations at s = 7 and 8 TeV with the ATLAS and CMS experiments (17) Combined Measurement of the Higgs Boson Mass in pp Collisions at s = 7 and 8 TeV with the				

Wikidata will likely offer you several match candidates for each label. You can inspect these by hovering over them if you think any could be a true match, but since we have already looked extensively for a match and not found one, we can ignore the match candidates.

Facet / Filter Undo / Redo 21 / 21

Refresh Reset All Remove All

Wikidata label: judgment [change](#)

1 choices Sort by: name count

none 5

Facet by choice counts

Wikidata label: best candidate's score [change](#) [reset](#)

18.00 — 93.00

☒ Numeric ☐ Non-numeric ☒ Blank ☐ Error

2 0 3 0

5 rows

Show as: rows records Show 5 10 25 50 rows

All	Name	Wikidata label	Entry Type	instance of	label description	SNAC ARK ID
☆	1.	Experimental Discussion Project	Corporate Body	project	project of the Fund for	w60s7hgw
☆	2.	WAFM-TV (Television Station : Birmingham, A)	Corporate			
☆	3.	WTOV-TV (Television Station : Washington, D.C.)	Corporate			
☆	4.	Carter, Elmer A. (Elmer Anderson), 1890-1973	Person			

Wikidata label

- Facet
- Text filter
- Edit cells
- Edit column
- Transpose
- Sort...
- View
- Reconcile
- Search for match
- Elmer A. Carter
- ☒ Elmer Carter (92)
- ☒ Elmer Anderson
- ☒ The Carters of E
- ☒ Elm Branch (50)
- ☒ Photograph of E
- ☒ Elm Springs Lak
- ☒ Elmer Spring (34)
- ☒ Dutch elm disease in Illinois (33)
- ☒ Chronic Standing Desk Use and Arterial Stiffness (3)
- ☒ Control of elm diseases in nursery elm plantings (25)
- ☒ Blended learning within an undergraduate exercise physiology laboratory (28)
- ☒ Progress in the control of elm diseases in nurseries (27)
- ☒ Elm Hollow (25)
- ☒ The 2-hour marathon: what do students think? (25)
- ☒ Pennsylvanians interview new Surgeon General. Interview by John Whyte, Tom McCarter, Elmer Pinzon (24)
- ☒ A report on tests of fungicides for the control of elm diseases in nurseries (24)
- ☒ An outside-the-box activity to demonstrate how humans and animals turn (24)
- ☒ Host predilection and transmissibility of vesicular stomatitis New Jersey virus strains in domestic cattle (Bos taurus) and swine (Sus scrofa). (22)
- ☒ In Vivo Neural Tissue Engineering: Cylindrical Biocompatible Hydrogels That Create New Neural Tracts in the Adult Mammalian Brain. (20)
- ☒ Combined Measurement of the Higgs Boson Mass in p p Collisions at s = 7 and 8 TeV with the ATLAS and CMS Experiments (18)
- ☒ Mechanical transmission of vesicular stomatitis New Jersey virus by Simulium vittatum (Diptera: Simuliidae) to domestic swine (Sus scrofa). (18)
- ☒ Combinations of single-top-quark production cross-section measurements and ttVVtb determinations at s = 7 and 8 TeV with the ATLAS and CMS experiments (17)
- ☒ Combined Measurement of the Higgs Boson Mass in pp Collisions at s = 7 and 8 TeV with the

To indicate to OpenRefine that you are creating new items, go to Reconcile > Actions > Create a new item for each cell.

Facet / Filter Undo / Redo 23 / 23 5 rows

Show as: rows records Show: 5 10 25 50 rows

Refresh Reset All Remove All

Instance of: judgment change

2 choices Sort by: name count

The bright green bar and gray "new" next to each item means that new items will be created for all your records.

Next, reconcile any other properties that you are planning to submit to Wikidata whose values refer to Wikidata items (e.g., the value "human" refers to a Wikidata item for "human"). Use the same process as before, except now you should be able to find good matches for all your values. In this example, we only need to reconcile the instance of column.

The reconciliation of any properties with values that refer to Wikidata items is a required step to use those properties in Wikidata uploads.

Wikidata label	Entity Type	instance of	label description	SNAC ARK ID
Experimental Discussion Project new Choose new match	Corporate Body	project <input checked="" type="checkbox"/> Microsoft Project (100) <input checked="" type="checkbox"/> project (100) <input checked="" type="checkbox"/> Project (100) <input checked="" type="checkbox"/> project manager (100) <input checked="" type="checkbox"/> The Project (100) <input checked="" type="checkbox"/> project (100) <input checked="" type="checkbox"/> Wikimedia project page (100) <input checked="" type="checkbox"/> Project (100) <input checked="" type="checkbox"/> Project (100) <input checked="" type="checkbox"/> Project (100) <input checked="" type="checkbox"/> Project (100) <input checked="" type="checkbox"/> Project (100) <input checked="" type="checkbox"/> Proekt (100) <input checked="" type="checkbox"/> projectile (88) <input checked="" type="checkbox"/> Los Alamos National Laboratory (88) <input checked="" type="checkbox"/> bullet (82) <input checked="" type="checkbox"/> map projection (82) <input checked="" type="checkbox"/> GNU Project (78) <input checked="" type="checkbox"/> side project (78) <input checked="" type="checkbox"/> fork (74) <input checked="" type="checkbox"/> Touhou Project (74) <input checked="" type="checkbox"/> Sega 32X (74) <input checked="" type="checkbox"/> Wikipedia:Stub (74) <input checked="" type="checkbox"/> Apollo space program (70) <input checked="" type="checkbox"/> Nintendo DS (70) <input checked="" type="checkbox"/> Create new item Search for match	project of the Fund for Adult Education	w60s7hgw
		television station Choose new match	television station in Birmingham, Alabama	w6qh555q
		television station Choose new match	television station in Washington, DC	w6mt0z6z
		human <input checked="" type="checkbox"/> human (100) <input checked="" type="checkbox"/> Uman (100) <input checked="" type="checkbox"/> Human (100) <input checked="" type="checkbox"/> human (100) <input checked="" type="checkbox"/> Human (100) <input checked="" type="checkbox"/> Human (100) <input checked="" type="checkbox"/> Human from Star Trek (100)	commissioner of the New York State Commission Against Discrimination	w6pr9mh6

« first < previous 1 - 5 next > last »

Extensions: SNAC Wikidata

Map your columns to the Wikidata schema (Wikidata > Edit Wikidata schema) by dragging and dropping them to the appropriate boxes after clicking the "add term" and "add statement" buttons as needed.

The Wikidata schema below specifies how your tabular data will be transformed into Wikidata edits. You can drag and drop the column names below in most input boxes: for each row, edits will be generated with the values in these columns.

Save schema Discard changes

Name Wikidata label Entity Type instance of label description SNAC ARK ID

Wikidata label

Terms

Label

en

Wikidata label

☐ override if present

remove

Description

en

label description

☐ override if present

remove

+ add term

Statements

instance of

instance...

0 references

remove

+ add qualifier

+ add reference

+ add value

SNAC ARK ID

SNAC ARK...

0 references

remove

+ add qualifier

+ add reference

+ add value

+ add statement

+ add item

Labels, descriptions, and aliases are all terms, while properties with P-codes are statements. Search for the property you are you using in the left box, and then drag the column to the right box.

Save the schema, then review the Issues tab. In this example, the issues raised are all due to lack of references, which is okay for the very basic statements being submitted along with the SNAC ARK IDs.

The Wikidata schema below specifies how your tabular data will be transformed into Wikidata edits. You can drag and drop the column names below in most input boxes: for each row, edits will be generated with the values in these columns.

[Save schema](#) [Discard changes](#)

Name Wikidata label Entity Type instance of label description SNAC ARK ID

Wikidata label

remove

Terms

Label

en

Wikidata label

☐ override if present

remove

Description

en

label description

☐ override if present

remove

+ add term

Statements

instance of



instance...

remove

▼ 0 references

+ add qualifier

+ add reference

+ add value

SNAC ARK ID



SNAC ARK...

remove

► 0 references

+ add qualifier

+ add reference

+ add value

+ add statement

+ add item

Go to Wikidata > Upload edits to Wikidata and then sign in (you can use a bot password for this). Provide a short description of your edits as an edit summary to be attached to the edit history of all the newly created items. Then, upload your edits, which will take a few minutes.

WAFM-TV (Q108766150)

television station in Birmingham, Alabama



 edit

▾ In more languages



Configure

Language	Label	Description	Also known as
English	WAFM-TV	television station in Birmingham, Alabama	
Spanish	No label defined	No description defined	
Traditional Chinese	No label defined	No description defined	
Chinese	No label defined	No description defined	

Statements

instance of	 television station	 edit
	▾ 0 references	+ add reference
		+ add value
		+ add statement

Identifiers

SNAC ARK ID	 w6qh555q	 edit
	▾ 0 references	+ add reference

After the upload is complete, you can review some of the new items on Wikidata to ensure the upload appeared correctly. The items will appear in the Contributions list of your Wikidata account.

Facet / Filter Undo / Redo 27 / 27

5 rows Schema Issues 10 Preview

Extensions: SNAC Wikidata

Refresh Reset All Remove All

Show as: rows records Show: 5 10 25 50 rows

« first < previous 1 - 5 next > last »

Instance of: judgment change

All Name Wikidata label Entity Type instance of label description SNAC A

1 choices Sort by: name count

matched 5
Facet by choice counts

Instance of: best candidate's change reset

1. Experi
2. WAFM
3. WTOP
D.C.)

Add column based on column Wikidata label

New column name Wikidata ID

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

Expression Language General Refine Expression Language (GREL)

cell.recon.match.id No syntax error.

Preview History Starred Help

row value

1. Experimental Discussion Project
2. WAFM-TV
3. WTOP-TV
4. Elmer A. Carter
5. F. Breitenfeld, Jr.

cell.recon.match.id
Q108766149
Q108766150
Q108766152
Q108766153
Q108766154

OK Cancel

Create a new column for the IDs of the new Wikidata items. Go to Edit column > Add column based on this column and this GREL expression to pull in the IDs:

`cell.recon.match.id`

From here, you can import the Wikidata IDs and the information you just submitted back into your database, if needed. For *Airwaves*, we downloaded the OR project as CSV and imported it into *Airtable*.

Tips and concluding notes:

- It can be helpful to test this workflow with a small batch of items first.
- There are ways of undoing batches of edits, if you make an error that is best corrected by reverting the entire batch of edits.
- This demonstration uses OpenRefine Version 3.4.1, so future versions of OR may diverge from the screencaps used here.

Resources:

- Wikidata's OpenRefine editing guides:
<https://www.wikidata.org/wiki/Wikidata:Tools/OpenRefine/Editing>
- OpenRefine's Wikidata guides (generalizable to Wikibase):
<https://docs.openrefine.org/manual/wikibase/overview>
- EditGroups: <https://editgroups.toolforge.org/> (to see the progress of your batch edits and undo them, if need be)
- Creating a bot password: <https://www.wikidata.org/wiki/Special:BotPasswords>
- OpenRefine GREL reference: <https://docs.openrefine.org/manual/grelfunctions>