

Enhancing Existing Wikidata Items with OpenRefine

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

Facet / Filter

Undo / Redo 1 / 1

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](#)

99 rowsShow as: **rows** records Show: 5 10 25 50 rows

All	Wikidata ID	Description (Der)	AltLabels (Aen)	date of birth (P5)	date of death (P)	place of birth (P)	place of death (P)	occupation (P10)
26.	Q5000894 Choose new match							Wisconsin
27.	Q6329018 Choose new match							
28.	Q6334739 Choose new match							
29.	Q6336165 Choose new match							
30.	Q6693228 Choose new match	United States educational foundation located in Boston, Massachusetts						
31.	Q6861322 Choose new match							
32.	Q6972391 Choose new match	distributor of radio programs in the United States						
33.	Q7280867 Choose new match							
34.	Q7346886 Choose new match							
35.	Q7858119 Choose new match							
36.	Q7945762 Choose new match	philanthropic foundation						
37.	Q7947006 Choose new match	United States						
38.	Q7947099 Choose new match							

Starting with a CSV of Wikidata IDs and fields with all the enhancements to be submitted according to their corresponding property. This looks a lot like the CSV that results when names are reconciled and data extended from Wikidata, but this is only the data to be submitted.

Go to Reconcile > Use values as identifiers on the Wikidata ID column to match Wikidata IDs to their corresponding items.

Optional: Verify that the statements you intend to submit to Wikidata will not be duplicates by pulling in the data currently on Wikidata for each of the properties you are using. For example, pull in the descriptions on Wikidata for all the items (column Den). Go to Edit column > All columns from reconciled values... and use the property codes for your columns.

	Den	Description (Der)	AltLabels (Aen)	date of birth (P5)	date of death (P)	place of birth (P)	place of death (P)	occupa
	American musician							Wisconsin
	Prairie Public Radio Roots, Rock, and Jazz station in Grand Forks, North Dakota, United States							
	radio station in Oregon							
	public broadcaster in the San Francisco Bay Area							
		United States educational foundation located in Boston, Massachusetts						
	American journalist							
		distributor of radio programs in the United States						
	college radio station of the University of Minnesota Twin Cities							

Use Facet > Customized facets > Facet by blank (null or empty string) on the column pulled in from Wikidata, then repeat on the column to be submitted to Wikidata to determine if any properties for which you are submitting data already have values on Wikidata for that item. Many properties can have multiple statements added, but you should avoid duplicating existing information.

Facet / Filter Undo / Redo 26 / 26

Refresh

Reset All Remove All

 occupation (P106) change invert reset

2 choices Sort by: name count

false 25

exclude

true 76

Facet by choice counts

25 matching records (99 total)

Extensions: SNAC ▾ Wikidata ▾

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

« first < previous 1 - 25 next > last »

of death	occupation (P106)	field of work (P101)	employer (P108)	inception (P571)	located in the ad	country (P17)	member of (P46)	SNAC ARK ID (P6436f)
	author,film producer,broadcast programming director broadcaster	broadcasting	WOR,CBS Radio					w6b6436f
	broadcaster	broadcasting						w6225vc9
	broadcasting executive	broadcasting						w6mw3wdn
	broadcasting executive	broadcasting						w6k37bm0
	broadcasting executive	broadcasting						w6r22hr5
	broadcasting executive,professor broadcaster	broadcasting,education	University of Michigan					w6tf3sk1
	broadcasting executive,professor broadcaster	broadcasting,education	WNYC,City College of New York					w6h72h53
	broadcasting executive	broadcasting						w63b7gb4
	broadcasting executive	broadcasting	University of Alabama				National Association of Educational Broadcasters	
	broadcasting executive	broadcasting,education	University of North Carolina					
	broadcasting executive	broadcasting,education	National Educational Radio				National Association of Educational Broadcasters	
	consultant,professor	broadcasting,education	Michigan State University,Federal Communications Commission,University				National Association of Educational Broadcasters	

In order to submit multiple statements to the same property for the same item, the values must be split into their own rows, rather than comma-separated as shown here.

Facet / Filter Undo / Redo 26 / 26

Refresh

Reset All Remove All

 occupation (P106) change invert reset

2 choices Sort by: name count

false 25

true 76

Facet by choice counts

25 matching records (99 total)

Extensions: SNAC ▾ Wikidata ▾

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

« first < previous 1 - 25 next > last »

of death | occupation (P106) field of work (P101) employer (P108) inception (P571) located in the administrative territorial entity (P276) country (P17) member of (P46) SNAC ARK ID (P649)

- Facet
- Text filter
- Edit cells

 - Transform...

- Edit column

 - Common transforms

- Transpose
- Fill down
- Blank down
- Sort...

 - View
 - Split multi-valued cells...

- Reconcile

 - Join multi-valued cells...

- broadcasting executive,professor
- broadcasting executive,professor
- broadcaster

On the column you want to split, go to Edit cells > Split multi-valued cells on the column you want to split, and then enter your separator exactly as it appears in the column. For these values, the separator is "," (no space).

w63b7gb4

National Association of Educational Broadcasters

National Association of Educational Broadcasters

National Association of Educational Broadcasters

javascript:{}

Facet / Filter Undo / Redo 27 / 27

Refresh

Reset All Remove All

 occupation (P106) change invert reset

2 choices Sort by: name count

false 32

exclude

true 76

Facet by choice counts

25 matching records (99 total)

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

Extensions: SNAC ▾ Wikidata ▾

of death (F) occupation (P106) field of work (P101) employer (P108) Inception (P571) located in the ad country (P17) member of (P46) SNAC ARK ID (P)

author	broadcasting	WOR,CBS Radio				
film producer						
broadcast programming director	broadcasting					
broadcaster	broadcasting					
broadcaster	broadcasting					
broadcasting executive	broadcasting					
broadcasting executive	broadcasting					
broadcasting executive	broadcasting					
broadcasting executive	broadcasting					
professor						
broadcasting executive	broadcasting,					
professor	broadcasting					
broadcaster	broadcasting					
broadcasting executive	broadcasting					w63b7gb4
broadcasting executive	broadcasting	University of Alabama				
broadcasting executive	broadcasting,education	University of North Carolina				National Association of Educational Broadcasters

Now the same column contains multiple rows with values that are still connected to the same Wikidata ID (e.g., the first three light gray rows show three occupation values, all three of which are tied to the Wikidata ID of the first row. Be sure and view your project in "records" mode rather than "rows" mode, so that grouped rows are treated as such and not as individual records.

Facet / Filter Undo / Redo 55 / 55

99 records Schema Issues 1 Preview

Extensions: SNAC ▾ Wikidata ▾

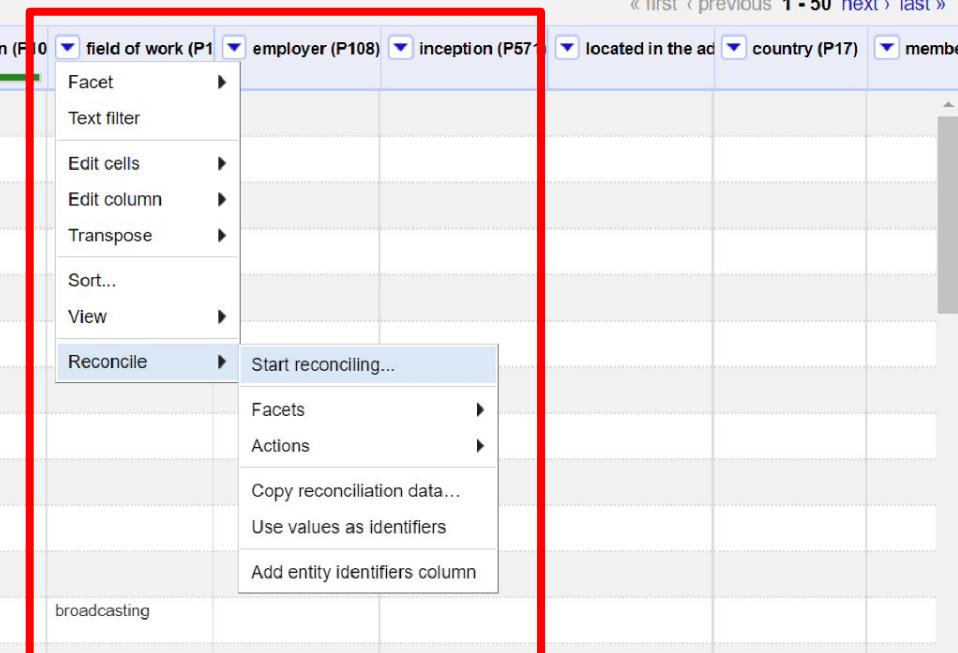
Refresh

Reset All Remove All

 occupation (P106): best candidate's score change reset

100.00 — 101.00

Before data can be submitted to Wikidata, any columns whose values refer to other Wikidata items (e.g., "broadcasting" as the value for property "field of work) must be reconciled with Wikidata. To do this, go to Reconcile > Start reconciling... on the column you want to reconcile. Choose Wikidata (en) as your reconciliation service when prompted.



OpenRefine Wikidata enhancement: Rich Scratch/Rich Enhance

Facet / Filter Undo / Redo Refresh Reset

occupation (P106): best candidate's score
100.00 — 101.00
25 Numeric 0 Non-numeric

Reconcile column "field of work (P101)"

» Access Service API

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

- academic discipline Q11862829
- industry Q268592
- scholarly article Q13442814
- economic activity Q8187769
- academic major Q4671286
- branch of science Q2465832
- packet forwarding Q874048
- art form Q56055944
- specialty

Reconcile against type:

Reconcile against no particular type

Auto-match candidates with high confidence

Maximum number of candidates to return

Add Standard Service...

Also use relevant details from other columns:

Column	Include? As Property
Wikidata ID	<input type="checkbox"/> <input type="text"/>
Description (Den)	<input type="checkbox"/> <input type="text"/>
Alt labels (Aen)	<input type="checkbox"/> <input type="text"/>

Wikidata will try to automatically select a type of entity in your column to reconcile against. In most cases, you want to change this to "Reconcile against no particular type". You can also auto-match high confidence candidates if you would like, or just manually choose the correct candidate from the list of options. For reconciling data you intend to submit, there is not any use in choosing options in the right-hand box on the screen.

Facet / Filter Undo / Redo 56 / 56

34 matching records (99 total)

Schema

Issues 1

Preview

Extensions: SNAC ▾ Wikidata ▾

Refresh

Reset All Remove All

 field of work (P101): judgment

2 choices Sort by: name count

(blank) 73

none

Facet

Each record that contains a value in the column being reconciled will have a list of options appear like this. These are the candidates for reconciling, and the highest confidence candidates will appear at the top, though sometimes the correct match may actually be further down the list. If no candidates appear or none or correct, there is an option to search for a match at the bottom of the candidate list for each record.

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

 field of work (P101) field of work (P101)

broadcasting

- radio broadcasting (100)
- Broadcasting... (100)
- broadcasting (100)
- broadcasting (100)
- Radio Catalana (100)
- Broadcasting (100)
- Broadcasting (100)
- Fox Broadcasting Company (86)
- broadcaster (86)
- Alpha Media (86)
- news broadcasting (83)
- broadcasting licence area (83)
- Broadcasting Board (80)
- broadcasting house (80)
- Broadcasting House (80)
- Broadcasting & Cable (80)
- Broadcasting House (80)
- broadcaster (78)
- public broadcasting (77)
- Turner Broadcasting System (77)
- broadcast network (75)
- Hubbard

 employer (P108) inception (P571) located in the ad country (P

« first < previous 1 - 34 next > last »

To match a value to a Wikidata item, hover over it to see the brief description and photo, if available. You can also click on the label name to open the Wikidata item page in a new tab if it not possible to determine the correct match based on description alone. Once you have found the match, click "Match All Identical Cells" if you have that value in multiple cells and want to match them all at once. The only case where you might choose "Match this Cell" and not all identical cells is if you have two values with the same label that should match to different items (e.g., you have "New York" in two cells, but one refers to the city and one refers to the state).

1 Preview

broadcasting (P10)

field of work (P101) employer (P108) inception (P571) located in the administrative territorial entity (P276) country (P17)

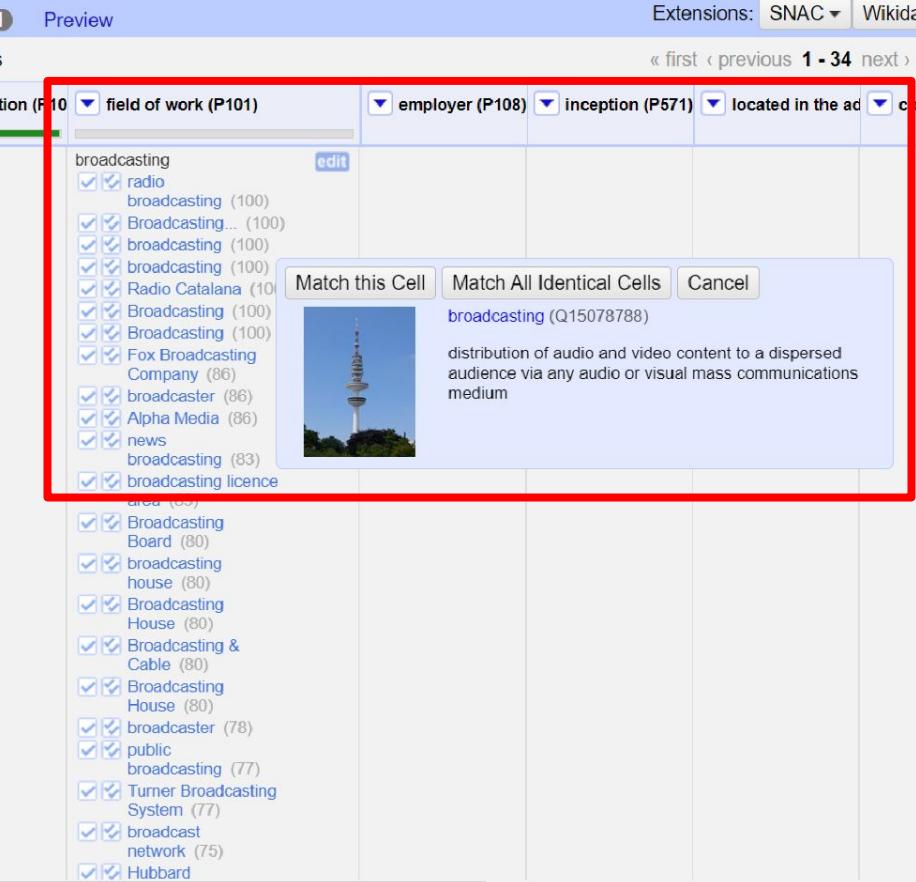
broadcasting

radio broadcasting (100)
 Broadcasting... (100)
 broadcasting (100)
 broadcasting (100)
 Radio Catalonia (100)
 Broadcasting (100)
 Broadcasting (100)
 Fox Broadcasting Company (86)
 broadcaster (86)
 Alpha Media (86)
 news broadcasting (83)
 broadcasting licence
area (83)
 Broadcasting Board (80)
 broadcasting house (80)
 Broadcasting House (80)
 Broadcasting & Cable (80)
 Broadcasting House (80)
 broadcaster (78)
 public broadcasting (77)
 Turner Broadcasting System (77)
 broadcast network (75)
 Hubbard

[edit](#)

Match this Cell Match All Identical Cells Cancel

broadcasting (Q15078788)
distribution of audio and video content to a dispersed audience via any audio or visual mass communications medium



Facet / Filter Undo / Redo 62 / 62

Refresh

Reset All Remove All

 field of work (P101): judgment

2 choices Sort by: name count

(blank) 73

matched 44

exclude

Facet by choice counts

 field of work (P101): best candidate's score

100.00 — 101.00

 Numeric Non-numeric Blank Error

34

0

7

0

34 matching records (99 total) Schema Issues 1 Preview

Extensions: SNAC ▾ Wikidata ▾

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

« first < previous 1 - 34 next > last »

Category (P10) field of work (P1) employer (P108) inception (P571) located in the administrative area (P17) country (P17) member of (P46) SNAC ARK ID (P Library of Congress VIAF)

broadcasting						
Choose new match						
education						
Choose new match						
broadcasting						
Choose new match						
journalism						
Choose new match						
broadcasting						
Choose new match						
education						
Choose new match						
communication studies						
Choose new match						
broadcasting						
Choose new match						
education						
Choose new match						
journalism						
Choose new match						
communication studies						
Choose new match						
broadcasting						
Choose new match						
broadcasting						
Choose new match						

Once all your values in a column are matched, it will look like this and there will be a solid dark green bar below the column name.

Facet / Filter

Undo / Redo 100 / 101

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](#)

Before you start matching columns to Wikidata properties, you must ensure that each ROW has a Wikidata ID in it (not just each record).

99 records Schema * Issues 1 Preview

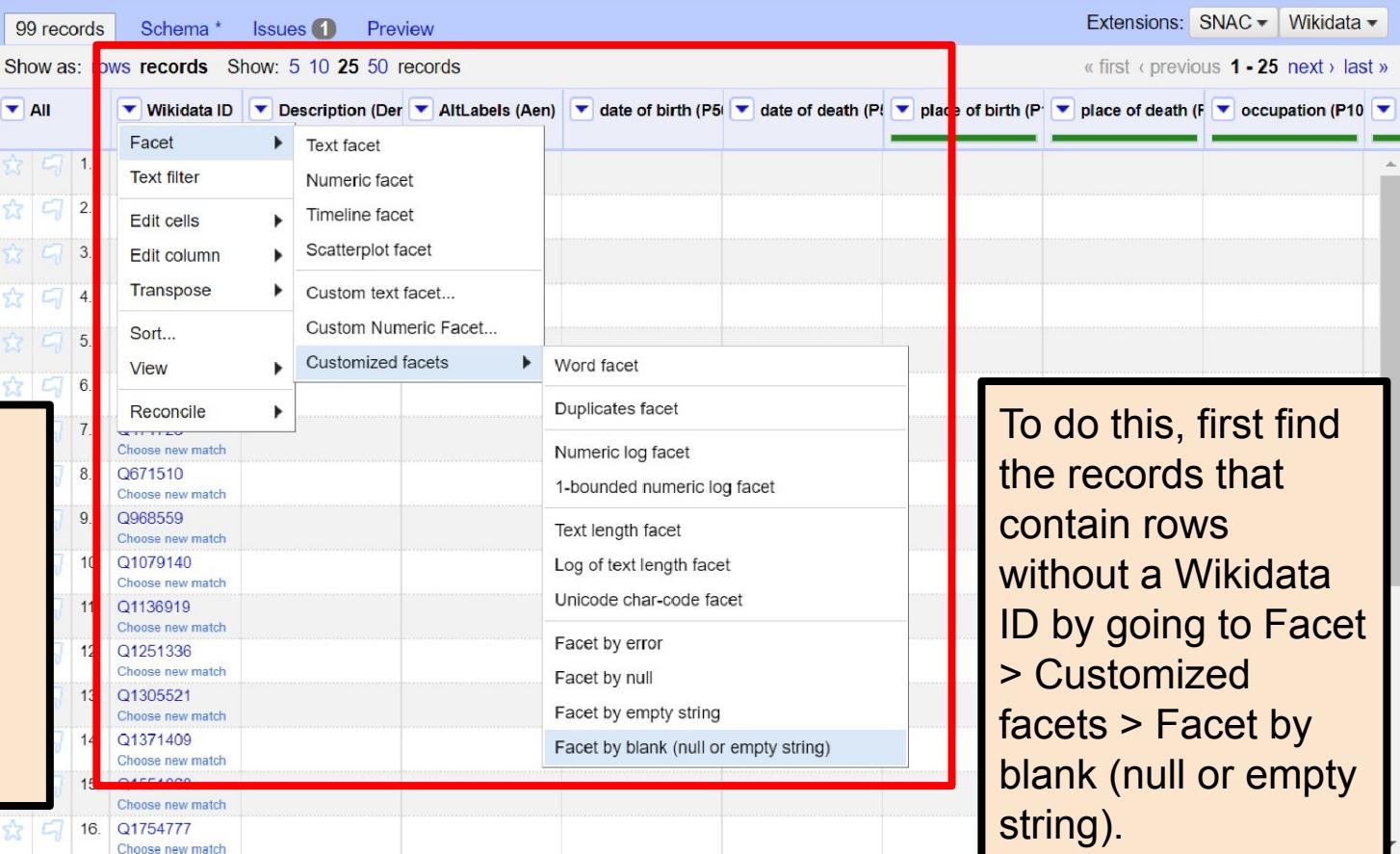
Show as: rows records Show: 5 10 25 50 records

All Wikidata ID Description (Der) AltLabels (Aen) date of birth (P51) date of death (P52) place of birth (P141) place of death (P142) occupation (P101)

Facet Text facet
Text filter Numeric facet
Edit cells Timeline facet
Edit column Scatterplot facet
Transpose Custom text facet...
Sort... Custom Numeric Facet...
View Customized facets Word facet
Reconcile Duplicates facet
Choose new match Numeric log facet
Q671510 1-bounded numeric log facet
Choose new match Text length facet
Q968559 Log of text length facet
Choose new match Unicode char-code facet
Q1079140 Facet by error
Choose new match Facet by null
Q1136919 Facet by empty string
Choose new match Facet by blank (null or empty string)
Q1251336
Q1305521
Q1371409
Q1754777

Extensions: SNAC ▾ Wikidata ▾

« first < previous 1 - 25 next > last »



To do this, first find the records that contain rows without a Wikidata ID by going to Facet > Customized facets > Facet by blank (null or empty string).

Facet / Filter Undo / Redo 100 / 101

Refresh Reset All Remove All

Wikidata ID change invert reset

2 choices Sort by: name count

false 99
true 18

Facet by choice counts exclude

15 matching records (99 total) Schema * Issues 1 Preview Extensions: SNAC ▾ Wikidata ▾

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

All Wikidata ID Description (Der) AltLabels (Aen) date of birth (P51) date of death (P52) place of birth (P14) place of death (P15) occupation (P10)

34. Q7346886

51.

56.

70. Q107621626 Choose new match 1888 1964 broadcasting executive Choose new match university teacher Choose new match

72. Q107621632 Choose new match Seymour Nathaniel Siegel 1978 broadcasting executive Choose new match university teacher Choose new match

79. Q107621648 1925-04-17 2019-04-25 Arkansas broadcasting

Now all the records showing are those which contain multiple rows tied to one record, such as those marked below. The next step is to "fill down" the Wikidata IDs into those rows where the ID column is empty, so that OpenRefine and Wikidata know what IDs each statement is attached to.

Facet / Filter Undo / Redo 100 / 101

Refresh

Reset All Remove All

 Wikidata ID change invert reset

2 choices Sort by: name count

false 99

true 18

exclude

Facet by choice counts

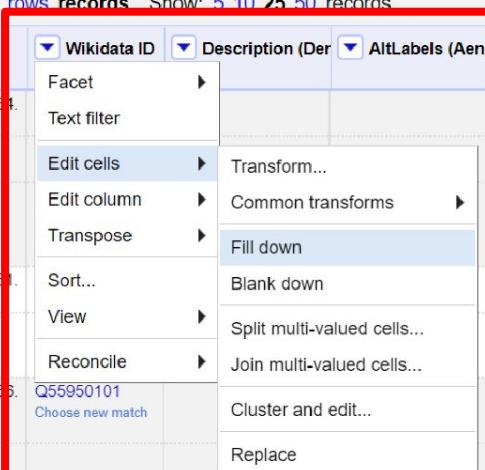
15 matching records (99 total) Schema * Issues 1 Preview

Extensions: SNAC ▾ Wikidata ▾

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

« first < previous 1 - 15 next > last »

All	Wikidata ID	Description (Der)	AltLabels (Aen)	date of birth (P5)	date of death (P)	place of birth (P)	place of death (P)	occupation (P10)



Go to Edit cells > Fill down on the Wikidata ID column while still faceted to only show records that contain multiple rows.

Facet / Filter

Undo / Redo 101 / 101

Refresh

Reset All Remove All

 Wikidata ID

change invert reset

99 choices Sort by: name count

Cluster

Q107621650 3

exclude

Q55950101 3

exclude

Q7346886 3

exclude

Q107621626 2

Q107621632 2

Q107621648 2

Q107621649 2

Q107621652 2

Q107621653 2

Q107621654 2

Q107621655 2

Q107621660 2

9 matching records (117 total)

Schema *

Issues 1

Preview

Extensions: SNAC ▾ Wikidata ▾

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

« first previous 1 - 9 next last »

All	Wikidata ID	Description (Der)	AltLabels (Aen)	date of birth (P50)	date of death (P51)	place of birth (P14)	place of death (P15)	occupation (P10)
34.	Q7346886	Choose new match						writer
35.	Q7346886	Choose new match						film producer
36.	Q7346886	Choose new match						broadcast programming director
59.	Q55950101	Choose new match						education
60.	Q55950101	Choose new match						journalist
61.	Q55950101	Choose new match						communicator
90.	Q107621650	Choose new match						student
91.	Q107621650	Choose new match						researcher
92.	Q107621650	Choose new match						teacher

After filling down, there will be multiple records with the same Wikidata ID, and there should not be any rows that don't have a Wikidata ID now.

Facet / Filter

Undo / Redo 101 / 101

117 records

Schema *

Issues 334

Preview

Extensions: SNAC ▾ Wikidata ▾

Go to Wikidata >
Edit Wikidata
schema to add the
Schema tab to the
blue bar at the top
of the screen. From
here, begin to drag
and drop your
properties to the
schema, using 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

Wikidata ID Description (Den) AltLabels (Aen) date of birth (P569) date of death (P570) place of birth (P19) place of death (P20) occupation (P106) field of work (P101)

employer (P108) inception (P571) located in the administrative territorial entity (P131) country (P17) member of (P463) SNAC ARK ID (P3430)

Library of Congress authority ID (P244) VIAF ID (P214) WorldCat Identities ID (P7859) SNAC IDs for references

Wikidata ID

remove

Terms

Alias

en

AltLabels (Aen)

override if present

remove

Description

en

Description (Den)

override if present

remove

+ add term

Statements

date of birth

date of birth (P5...

remove

+ add qualifier

+ add reference

+ add value

date of death

date of death (P5...

remove

+ add qualifier

+ add reference

+ add value

place of birth

place of birth (P...

remove

Facet / Filter

Undo / Redo 101 / 101

117 records

Schema *

Issues 334

Preview

Extensions: SNAC ▾ Wikidata ▾

For enhancing existing items, you will always need to drag the Wikidata ID column to the top spot in the schema (small red box). All other terms and statements can be mapped in according to your enhancements.

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

Wikidata ID Description (Den) AltLabels (Aen) date of birth (P569) date of death (P570) place of birth (P19) place of death (P20) occupation (P106) field of work (P101)

employer (P108) inception (P571) located in the administrative territorial entity (P131) country (P17) member of (P463) SNAC ARK ID (P3430)

Library of Congress authority ID (P244) VIAF ID (P214) WorldCat Identities ID (P7859) SNAC IDs for references

Wikidata ID

remove

Terms

Alias en AltLabels (Aen)

override if present

remove

Description en Description (Den)

override if present

remove

+ add term

Statements

date of birth

date of birth (P5...

remove

+ add qualifier

+ add reference

+ add value

▶ 0 references

date of death

date of death (P5...

remove

+ add qualifier

+ add reference

+ add value

▶ 0 references

place of birth

place of birth (P...

remove

For adding terms, you'll need to select the appropriate term type from the drop down and enter "en" in the language box, then drag your columns to the term spaces. For adding statements, you'll need to start typing the property name in the left side box, select it from the list, then drag the appropriate column name to the right side box.

117 records Schema * Issues 334 Preview

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

Wikidata ID Description (Den) AltLabels (Aen) date of birth (P569) date of death (P570) place of birth (P19) place of death (P20) occupation (P106) field of work (P101)

employer (P108) inception (P571) located in the administrative territorial entity (P131) country (P17) member of (P463) SNAC ARK ID (P3430)

Library of Congress authority ID (P244) VIAF ID (P214) WorldCat Identities ID (P7859) SNAC IDs for references

Wikidata ID

remove

Terms

Alias en AltLabels (Aen)

 override if present

remove

Description en Description (Den)

 override if present

remove

Statements

date of birth

date of birth (P5...

remove

add qualifier

▶ 0 references

add reference

add value

date of death

date of death (P5...

remove

add qualifier

▶ 0 references

add reference

add value

place of birth

place of birth (P...

remove

[Facet / Filter](#)
[Undo / Redo](#) 101 / 101

117 records

Schema *

Issues 334

Preview

Extensions: SNAC ▾ Wikidata ▾

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](#)

Continue adding statements until you've mapped all the columns you want to use into the schema.

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](#)
[Wikidata ID](#) [Description \(Den\)](#) [AltLabels \(Aen\)](#) [date of birth \(P569\)](#) [date of death \(P570\)](#) [place of birth \(P19\)](#) [place of death \(P20\)](#) [occupation \(P106\)](#) [field of work \(P101\)](#)
[employer \(P108\)](#) [inception \(P571\)](#) [located in the administrative territorial entity \(P131\)](#) [country \(P17\)](#) [member of \(P463\)](#) [SNAC ARK ID \(P3430\)](#)
[Library of Congress authority ID \(P244\)](#) [VIAF ID \(P214\)](#) [WorldCat Identities ID \(P7859\)](#) [SNAC IDs for references](#)
[+ add qualifier](#)
[+ add reference](#)
[+ add value](#)
[remove](#)
[+ add qualifier](#)
[+ add reference](#)
[+ add value](#)
[remove](#)
[+ add qualifier](#)
[+ add reference](#)
[+ add value](#)
[remove](#)
[+ add qualifier](#)
[+ add reference](#)
[+ add value](#)

▶ 0 references

[employer](#)
 employer (P1... [remove](#))

▶ 0 references

[inception](#)
 inception (P5... [remove](#))

▶ 0 references

[located in the admin](#)
 located in the administrative territorial entit... [remove](#))

▶ 0 references

[Facet / Filter](#)[Undo / Redo](#) 101 / 101

117 records

Schema *

Issues 334

Preview

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

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](#)

Statements without references.

Most statements should have references. You can add them easily in the schema.

Description begins with uppercase letter

Description (en) such as [United States educational foundation located in Boston, Massachusetts](#) on [Q6693228](#) ([Q6693228](#)) begins with uppercase letter "U".

Descriptions begin with a lowercase letter except when uppercase would normally be required or expected. See the [manual](#) for more information.

333

At this stage, the Issues tab (to the right of Schema) will indicate that you have a lot of issues, depending on the size of your dataset and number of statements in use. Unless there are other issues you need to address (such as statements that don't conform to formatting requirements), this is because none of the statements have references yet. References aren't required, but are encouraged and help create additional links between Wikidata items and other sources. In this example, we will only be adding references to statements sourced from SNAC.

Facet / Filter

Undo / Redo 121 / 121

117 records

Schema *

Issues 431

Preview

Extensions: SNAC ▾ Wikidata ▾

Click "add reference", then fill in the key parts of the reference: stated in, the identifier (for databases only), and the retrieval date. For our purposes, I took the SNAC IDs of the Wikidata items I am enhancing, then added a column with the value "SNAC" for all the rows for which I wanted to add references, then added a third column with the retrieval date for the information. Instead of manually filling "SNAC" as the stated in value, I can then drag the stated in column to fill in that reference statement value.

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

Save schema

Discard changes

[labels \(Aen\)](#) [date of birth \(P569\)](#) [date of death \(P570\)](#) [place of birth \(P19\)](#) [place of death \(P20\)](#) [occupation \(P106\)](#) [field of work \(P101\)](#)[located in the administrative territorial entity \(P131\)](#) [country \(P17\)](#) [member of \(P463\)](#) [SNAC ARK ID \(P3430\)](#)[4\)](#) [VIAF ID \(P214\)](#) [WorldCat Identities ID \(P7859\)](#) [SNAC IDs for references](#) [Stated in](#) [Retrieved](#)

+ add reference

+ add value

country (P...

remove

+ add qualifier

▼ 1 references

copy

remove

remove

remove

remove

remove

remove

remove

remove

remove

+ add

+ add reference

+ add value

member of (P4...

remove

▼ 0 references

+ add qualifier

+ add reference

+ add value

remove

Facet / Filter

Undo / Redo 121 / 121

117 records

Schema *

Issues 431 Preview

Extensions: SNAC ▾ Wikidata ▾

If you want to use the exact same reference for any and all statements you make on an item, click the copy button on a completed reference, then click the "paste reference" button on any statements for which you still need to add an reference.

The references will only be added to items that have data in any of the three columns that form the references, so for any rows that shouldn't have references, leave the stated in, SNAC ID, and retrieved columns blank.

how your tabular data will be transformed into Wikidata edits. You can drag and drop the boxes: for each row, edits will be generated with the values in these columns.

Save schema Discard changes

els (Aen) date of birth (P569) date of death (P570) place of birth (P19) place of death (P20) occupation (P106) field of work (P101)

located in the administrative territorial entity (P131) country (P17) member of (P463) SNAC ARK ID (P3430)

VIAF ID (P214) WorldCat Identities ID (P7859) SNAC IDs for references Stated in Retrieved

▼ 1 References

copied	remove
stated in	remove
SNAC ARK ID	remove
retrieved	remove

+ add
+ add reference
+ add value

member of (P4... remove

▼ 0 references

+ add qualifier
paste reference
+ add reference
+ add value

SNAC ARK ID (P34... remove

▼ 0 references

+ add qualifier
paste reference
+ add reference
+ add value

The references will only be added to items that have data in any of the three columns that form the references, so for any rows that shouldn't have references, leave the stated in, SNAC ID, and retrieved columns blank. In general, OpenRefine will only do actions on items if you have filled in that column on that item's record. OR will not create completely blank statements or references.

A completed pasted reference.

444 Preview

Extensions: SNAC ▾ Wikidata ▾

how your tabular data will be transformed into Wikidata edits. You can drag and drop the boxes: for each row, edits will be generated with the values in these columns.

Save schema Discard changes

els (Aen) date of birth (P569) date of death (P570) place of birth (P19) place of death (P20) occupation (P106) field of work (P101)
located in the administrative territorial entity (P131) country (P17) member of (P463) SNAC ARK ID (P3430)
VIAF ID (P214) WorldCat Identities ID (P7859) SNAC IDs for references Stated in Retrieved

▼ 1 references

copy	remove
stated in	Stated in remove
SNAC ARK ID	SNAC IDs for refe... remove
retrieved	Retrieved remove

+ add
+ add reference
+ add value

member of (P4... remove)

▼ 1 references

copy	remove
stated in	Stated in remove
SNAC ARK ID	SNAC IDs for refe... remove
retrieved	Retrieved remove

+ add
+ add reference

Facet / Filter

Undo / Redo 122 / 122

117 records

Schema

Issues 444

Preview

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

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](#)

Statements without references.

Most statements should have references. You can add them easily in the schema.

**Statements without references.**

Most statements should have references. You can add them easily in the schema.

282

Some references were ignored.

None of their statements could be parsed, so no reference was added.

161

**Description begins with uppercase letter**

Description (en) such as [United States educational foundation located in Boston, Massachusetts](#) on [Q6693228 \(Q6693228\)](#) begins with uppercase letter "U". Descriptions begin with a lowercase letter except when uppercase would normally be required or expected. See the [manual](#) for more information.

1

The number of issues after adding references may actually go UP, as it does here, but that is because OpenRefine will warn you every time it is ignoring a blank reference. You should have fewer statements without references now though, which was the goal.

Facet / Filter Undo / Redo 122 / 122

Refresh

Reset All Remove All

 occupation (P106) change invert reset

2 choices Sort by: name count

false 32 exclude

true 85

Facet by choice counts

Look at the Preview tab to see how your statements will appear on some example items. You can also facet your dataset to inspect how specific items will preview, if you are concerned about certain items, statements, or references.

32 matching records (117 total) Schema Issues 261 Preview

Extensions: SNAC ▾ Wikidata ▾

This tab shows the first edits (out of 25) that will be made once you upload the changes to Wikidata. You can use facets to inspect the edits on particular items.

Q7346886 (Q7346886)

employer (P108)

WOR (Q3564763)

▼ 1 references

SNAC ARK ID (P3430) w6w68rzn

stated in (P248) SNAC (Q29861311)

retrieved (P813) 14 October 2021

CBS Radio (Q1022995)

► 1 references

occupation (P106)

writer (Q36180)

► 1 references

film producer (Q3282637)

► 1 references

broadcast programming director (Q52085732)

► 1 references

field of work (P101)

broadcasting (Q15078788)

► 1 references

Go to Wikidata > Upload edits to Wikidata, and then sign into Wikidata. You can use a bot password for this. After logging in, you will be asked to provide an edit summary for your edits before submitting them. This should briefly describe the changes you made. For example: "add biographical or historical statements primarily using data from SNAC; add external authority identifiers".

Not sure how to get started?

[Watch these screencasts](#)

Wikidata account

Logging in to [Wikidata](#) lets you to upload edits directly from OpenRefine.

Username: remove

Password: remove

Remember credentials (stored unencrypted in OpenRefine's preferences)

[Close](#) [Log in](#)

▶ 1 references

date of death remove
+ add reference
+ add value

▶ 1 references

place of birth remove
+ add reference
+ add value

place of birth remove
+ add reference
+ add value

Facet / Filter

Undo / Redo 122 / 122

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](#)

Perform Wikibase edits
16% complete Cancel

117 records

Schema

Issues 41

Preview

Extensions: SNAC ▾ Wikidata ▾

The edits will take some time to process, depending on how large your dataset is. You can check how the edits are doing by looking at your contributions page on your Wikidata account, or by looking at the EditGroups tool.

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](#)

Wikidata ID Description (Den) AltLabels (Aen) date of birth (P569) date of death (P570) place of birth (P19) place of death (P20) occupation (P106) field of work (P101)

employer (P108) inception (P571) located in the administrative territorial entity (P131) country (P17) member of (P463) SNAC ARK ID (P3430)

Library of Congress authority ID (P244) VIAF ID (P214) WorldCat Identities ID (P7859) SNAC IDs for references Stated in Retrieved

Wikidata ID

remove

Terms

Alias

en

AltLabels (Aen)

override if present

remove

Description

en

Description (Den)

override if present

remove

[+ add term](#)

remove

[+ add qualifier](#)

[+ add reference](#)

[+ add value](#)

remove

[+ add qualifier](#)

[+ add reference](#)

[+ add value](#)

remove

► 1 references

place of birth

place of birth (P...

(latest | earliest) View (newer 50 | older 50) (20 | 50 | 100 | 250 | 500)

- 16:28, 14 October 2021 (diff | hist) . . (+1,363) . . W1XAL (Q107621675) (Updated Item: add biographical or historical statements primarily using data from SNAC; add external authority identifiers (details)) (current) (Tag: OpenRefine [3.4])
- 16:28, 14 October 2021 (diff | hist) . . (+1,361) . . Cohn and Marks (Q107621767) (Updated Item: add biographical or historical statements primarily using data from SNAC; add external authority identifiers (details)) (current) (Tag: OpenRefine [3.4])
- 16:28, 14 October 2021 (diff | hist) . . (+3,602) . . Cecil S. Bidlack (Q107621674) (Updated Item: add biographical or historical statements primarily using data from SNAC; add external authority identifiers (details)) (current) (Tag: OpenRefine [3.4])
- 16:28, 14 October 2021 (diff | hist) . . (+1,369) . . W9XHZ (Q107621762) (Updated Item: add biographical or historical statements primarily using data from SNAC; add external authority identifiers (details)) (current) (Tag: OpenRefine [3.4])
- 16:28, 14 October 2021 (diff | hist) . . (+852) . . WTOM (Q107621760) (Updated Item: add biographical or historical statements primarily using data from SNAC; add external authority identifiers (details)) (current) (Tag: OpenRefine [3.4])
- 16:28, 14 October 2021 (diff | hist) . . (+1,816) . . Robert J. Coleman (Q107621673) (Changed label, description and/or aliases in en, and other parts: add biographical or historical statements primarily using data from SNAC; add external authority identifiers (details)) (current) (Tag: OpenRefine [3.4])
- 16:28, 14 October 2021 (diff | hist) . . (+1,363) . . WRUL (Q107621756) (Updated Item: add biographical or historical statements primarily using data from SNAC; add external authority identifiers (details)) (current) (Tag: OpenRefine [3.4])
- 16:28, 14 October 2021 (diff | hist) . . (+3,487) . . John W. Dunn (Q107621672) (Updated Item: add biographical or historical statements primarily using data from SNAC; add external authority identifiers (details)) (current) (Tag: OpenRefine [3.4])

When your edits are complete, you can look on your contributions page to review them. Click on the label name to go to the Wikidata item page and see your edits. It is often useful to spot check a few different types of entities, as well as some that have references to make sure the edits were successful, especially when you are using OpenRefine to make edits for the first few times.

[external authority identifiers \(details\)](#) (current) (Tag: OpenRefine [3.4])

- 16:27, 14 October 2021 (diff | hist) . . (+3,997) . . T. M. Baird (Q107621655) (Changed label, description and/or aliases in en, and other parts: add biographical or

AC; add

phical or

phical or

IAC; add

Tips and concluding notes:

- It can be helpful to test this workflow with a small batch of items first, ideally a set where you can test adding references, and adding multiple statements of the same property type to a single item.
- 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>