

Workshop: Reconciling to Wikidata with OpenRefine

Wikidata

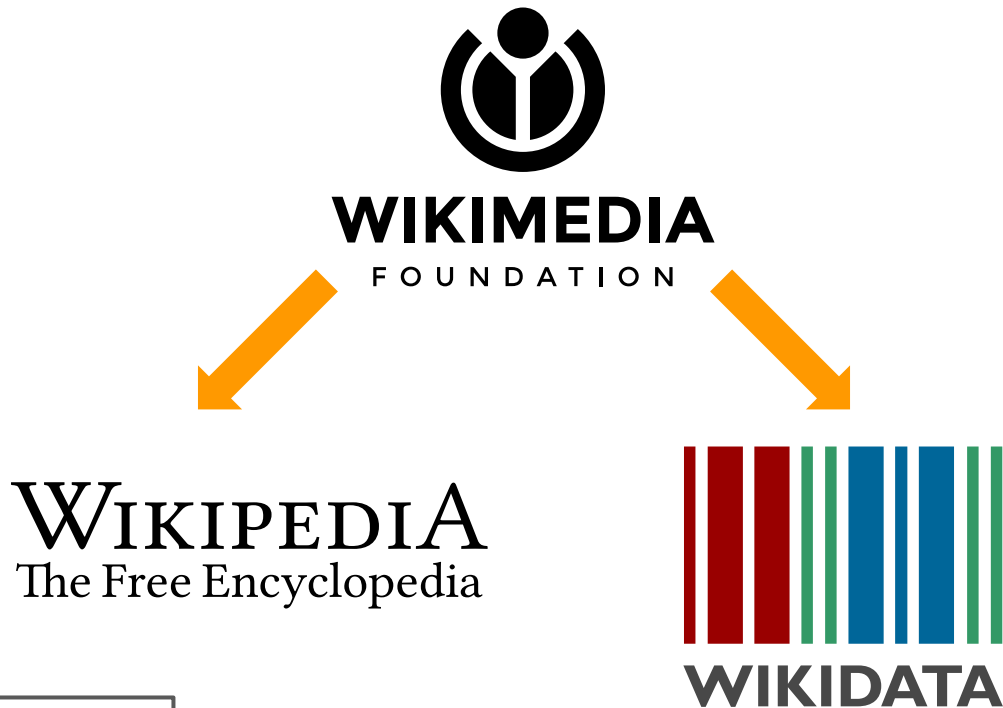
- Free
- Collaborative
- Structured data
- Items can describe people, places, organizations, things, publications, ideas, and more

Barack Obama (Q76)

president of the United States from 2009 to 2017

free will (Q9476)

ability of agents to make choices free from certain kinds of external constraints



Wikidata interfaces

Manual editing interface:
human-readable



- Main page
- Community portal
- Project chat
- Create a new Item
- Recent changes
- Random Item
- Query Service
- Nearby
- Help
- Donate

- Lexicographical data
- Create a new Lexeme
- Recent changes
- Random Lexeme

- Tools
- What links here
- Related changes
- Special pages

Item **Discussion** Read View history Search

Graydon Ausmus (Q107621647)

radio executive from Alabama [edit](#)

[In more languages](#)

Statements

instance of	human	edit
	0 references	+ add reference
		+ add value

date of birth	13 February 1911 <i>Gregorian</i>	edit
	0 references	+ add reference
		+ add value

place of birth	Texas	edit
----------------	-------	----------------------

OpenRefine interface:
machine-readable/transformable



OpenRefine CPF Authorities Reconciliation Example OR [Permalink](#) [Open...](#) [Export](#) [Help](#)

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 »

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	Recon Name	Len	Aen	Den	Instance of	date of birth	occupation	Inception	SNAC ARK ID
1	National Association of Educational Broadcasters	National Association of Educational Broadcasters	National Association of Educational Broadcasters	NAEB	US organization of broadcasters with the aim of coordinating educational programs	organization	1925-01-01T00:00:00Z			w6m70pt
2	WHA (Radio station in Madison, Wis.)	WHA	WHA		Wisconsin Public Radio Ideas Network flagship station in Madison, Wisconsin, United States	radio station	1920-01-01T00:00:00Z			w6pc86z4
3	Meyer, Milton, 1906-1986	Milton Mayer	Milton Mayer	Milton Sanford Meyer	American journalist	human	1906-06-24T00:00:00Z	journalist		w6sew4ndx
4	Fund for Adult Education (U.S.)	Fund for Adult Education	Fund for Adult Education		subsidiary foundation of the Ford Foundation	foundation	1951-01-01T00:00:00Z			w6382349
5	Gardiner, C. Harvey (Clinton Harvey)	C. Harvey Gardiner	C. Harvey Gardiner		radio broadcaster	human	1913-01-01T00:00:00Z	broadcaster		w6225vc9
6	Burrows, E. G. (Edwin Gladding), 1917-	Ed Burrows	Ed Burrows	E. G. "Ed" Burrows	radio program director and producer	human	1917-07-23T00:00:00Z	news presenter		w67946bk
7				Edwin Gladding Burrows		announcer				

Wikidata structure

item	property	value
<i>ID begins with 'Q'</i>	<i>ID begins with 'P'</i>	<i>ID begins with 'Q' -OR- is date, number, URL, or other non-item data type</i>
Marshall McLuhan (Q193871)	employer (P108)	University of Toronto (Q180865)
Marshall McLuhan (Q193871)	date of birth (P569)	21 July 1911 (no item ID)
Marshall McLuhan (Q193871)	Library of Congress authority ID (P244)	n80126329 (no item ID, links to: https://id.loc.gov/authorities/n80126329)

1 row = 1 statement on Wikidata

OpenRefine

- Designed for structured data
- Open-source
- Train station, NOT a database
- 3 major functions:
 1. Exploring data
 2. Transforming data
 3. Reconciling, or matching, data
- Basic building blocks: **General Refine Expression Language**
 - Rearranging strings
 - Concatenating strings
 - Extracting parts of strings
 - Math functions

OpenRefine extensions

- Wikidata extension:
 - Comes pre-installed
 - Connects your instance of OpenRefine to Wikidata
 - Major functions:
 - Reconcile your data to Wikidata and get back additional data based on matches
 - Submit new and edited Wikidata statements
- Other extensions & reconciliation services:
 - Social Networks and Archival Context extension
 - Virtual International Authority File
 - Advanced tools for reconciliation and extension

Reconciliation

- Exists outside of OpenRefine or Wikidata
- Match entities between two datasets
- Allows you to get unique IDs that refer to external databases
- Can "extend" data by pulling other columns or statements from external DBs
- Core function of OpenRefine

