Wikidata as the world's global gazetteer

Alex Lum (林信帝)

User:Canley

Wikimedia Australia

WDC

WIKI DATA CON

What is a gazetteer?

- A geographical index or dictionary of place names

- Often used in conjunction with maps to provide more information

The goal

Why Wikidata?

- There's nothing like Wikidata for this
  - There is Geonames, but will discuss problems with this later
- Linked data
- Multi-lingual
- And of course, open data

WDC
WIKI DATA CON

Detailed information for human settlements

- Every locality/human settlement has (where possible) a consistent set of information:
  - Instance of (P31) – as detailed as possible
  - Country (P17)
  - Gazetted name (Label)
  - Administrative territorial entity (P131)
  - Postcode/postal code/zip code (P281)
  - Coordinates (P625)

Detailed information for human settlements

- Additional information:
  - Population (P1082)
  - Area (P2046)
  - Image (P18)
  - Electoral district (P7938)
  - Any identifiers available

And everything else…

- Rivers and streams (with tributaries and watercourse mouth)
- Bodies of water
- Hills and mountains (with elevation)
- Bays, geological formations
- Protected areas
- Heritage sites

OpenStreetMap

- Wikidata can link to OSM via related properties
- OSM can link to Wikidata Q-IDs
- This allows mapping of polygons (for example, in Wikipedia infoboxes) and even federated SPARQL search

WDC
WIKI DATA CON

OpenStreetMap

- Let's get the import question out of the way…
  - Wikidata licence is CC0 (public domain)
  - OSM is ODbL (with attribution and share-alike provisions)
- OSM to Wikidata – Wikidata would not comply with ODbL
- Wikidata to OSM – technically OK, but…
  - Provenance issues with data in Wikidata (and Wikipedia)
  - OSM community consensus is not to import from those sources
- However, with linking, importing is not necessary…

The difficulties

WDC
WIKI DATA CON

It's difficult!

- Comprehensive data may not be available
  - Or may not be open licenced
- Reconciliation with existing items (avoid duplicates)
- Speaking of duplicates, identify where these already exist in Wikidata and merge

Duplicating Everywhere All At Once

- My presentation at Wikimania 2023 in Singapore
  - https://fb.watch/nJIa1vTStY/?mibextid=2Rb1fB
- Trying to identify duplicate place items created from Geonames by bots in Cebuano Wikipedia, then transferred by other bots to Wikidata
- Using machine learning clustering algorithm to identify duplicates
- Method to merge items using QuickStatements and the WM API

Other issues

- Names and other facts can be controversial
  - Disputed by communities, by the government or by other countries
- Official names, common names, bilingual names
- Linking to Wikipedia
  - Sometimes two different entities are in one article
  - They should probably be separate on Wikidata but then one doesn't link
  - Can now link to a redirect (since October 2022)

Some cool projects and tools

Australian Census Tool

- Lua module by Maia Williams which retrieves latest population value from Wikidata for Australian localities
- https://wikimedia.org.au/wiki/Populating_Wikipedia:_New_tool _integrating_Australian_Census_data

WDC
WIKI DATA CON

New Zealand cemeteries

- Small team of Wikidata enthusiasts compiled and mapped every cemetery in New Zealand in three days

- Map query: https://w.wiki/7qpE

Some recommendations

Recommendations

- Work with a team of interested Wikimedians

- Use a project page to track progress and discuss issues

- Set up a Mix 'n' Match catalog if you have an official identifier:

  - This helps with reconciling existing items

Thanks for your attention!

Get in touch:

Alex Lum (林信帝)

User:Canley

metacoretechs
@fosstodon.org

Credits

Information about the licence of your slides, the pictures used in your slides, links to additional resources, etc.

WDC
WIKI DATA CON