OpenStreetMap and Wikidata: Awesome Together

State of the Map 2019
Eugene Alvin Villar (OSM: seav)
Edward Betts (OSM: Edward)
CHAPTER 1

What is Wikidata?
Wikidata is a CC0-licensed structured knowledge base under the Wikimedia umbrella that contains over 60 million items.
Wikidata was intended to solve two things:

- The mess of interwiki links in the Wikipedias
- The duplication of data in Wikipedia infoboxes
Now, all interwiki link data is stored in Wikidata (with some exceptions)
...and Wikipedia infoboxes can populate themselves from Wikidata.
Wikidata probably has the largest collection of linked IDs to other third-party databases, data sets, and knowledge bases.
Wikidata has grown outside its original purpose and is fast becoming a central data hub especially for GLAM institutions (galleries, libraries, archives, and museums).
Wikidata data model
OpenStreetMap has nodes, ways, relations, tags, keys, values, roles, etc.
Wikidata has items, statements, properties, values, qualifiers, etc.
Data modeling discussions on the Wikidata:Project chat page are actually quite similar to discussions on OSM’s tagging mailing list.

(Both projects struggle with ontology issues because both projects are examples of folksonomies.)
Wikidata data is stored as a bunch of **triples**. The most basic triples in Wikidata are called **statements**.

This design is heavily based on W3C’s Resource Description Framework [RDF].
Triples, subjects, predicates, and objects are represented as IRIs.

Earth has the IRI: http://www.wikidata.org/entity/Q2

Or Q2 for short.

Wikidata triple

= Earth (Q2)

+ instance of (P31)

+ inner planet (Q3504248)
Triples themselves can be subjects of secondary triples, allowing us to qualify and add sourcing info. These secondary triples are called *qualifiers* and *references*, respectively.
Shown here are how various triples, subjects, predicates, and objects related to Curie are represented in Wikidata, and the common terms Wikidatans use to refer to them depending on context.
Advanced: This is a more complete RDF data model for the main Wikidata space. Useful when you need to fully understand the model or craft advanced queries.

Aside: Wikidata also has lexemes (L-namespace) for lexicographic data, structured data for Wikimedia Commons (SDC) (M-namespace), and shape expressions (E-namespace).
Wikidata Query Service
The Wikidata Query Service (WDQS) is the official API for querying data from Wikidata.

*Think*: Overpass API/Turbo for Wikidata.

https://query.wikidata.org
WDQS uses W3C’s SPARQL as the query language.

https://www.w3.org/TR/sparql11-query/
https://www.wikidata.org/wiki/Wikidata:SPARQL_tutorial
Examples and hands-on time!
Wikidata tools and apps
Monumental is a site that allows you to explore built heritage and learn more about monuments.

https://tools.wmflabs.org/monumental/
Crotos is a search and display engine for visual artworks, based on Wikidata and using Wikimedia Commons files.

https://http://zone47.com/crotos/
Histropedia allows you to generate interactive timelines with events linked to Wikipedia articles.

http://histropedia.com/
Inventaire: keep an inventory of your books!

https://inventaire.io/
More tools are listed on Wikidata:

CHAPTER 2

Linked Open Data
First: How OSM and Wikimedia benefit from each other
OSM has been used to create maps to illustrate Wikipedia articles and populate Wikimedia Commons.
OSM has been used to create maps to illustrate Wikipedia articles and populate Wikimedia Commons.

Bourrichon / CC BY-SA 3.0 .
https://commons.wikimedia.org/wiki/File:Macau_topographic_map-fr.svg
OSM powers the Wikimedia Foundation’s Kartotherian map tile service, which is used by the Kartographer MediaWiki extension and almost all other interactive maps on the Wikimedia projects.
The Wikimedia Foundation released internationalized map tiles for Kartotherian, leveraging OSM’s name:*=* tags.
WikiMiniAtlas, an older MediaWiki plugin still in use in many Wikipedias, is also powered by OSM data, including 3D building data.
The OSM Wiki is powered by MediaWiki, the wiki engine developed by Wikimedia, and this also provides access to Wikimedia Commons images.
The OpenStreetMap Foundation has derived its Local Chapters agreement, Trademark Policy, and Terms of Use from corresponding documents from the Wikimedia Foundation.
Why link OSM and Wikidata?
Tim Berners-Lee, the inventor of the World Wide Web and initiator of the W3C Linked Data project, suggested a 5 star deployment scheme for Linked Open Data.
1. On the Web
2. Machine-readable
3. Open data format
4. In RDF (use IRIs)
5. Linked to other datasets
Wikidata items on places can link to OSM relations using the **OSM relation ID** (P402) property.
OSM objects can link to corresponding Wikipedia articles and Wikidata items using the `wikipedia=*` and `wikidata=*` tags respectively.

There are also secondary tags like `brand:wikidata=*` and `architect:wikidata=*`. 
**Ontology:** Wikidata items about features can link to “equivalent” OSM features using the **OSM tag or key** (P1282) property.
**Ontology:** Tag definitions on the OSM Wiki can link to “equivalent” Wikipedia articles and Wikidata items.
Wikimedia maps can show geometry from OSM.

Pages in category "Heidelberg"

This category contains only the following page.

- Heidelberg

Media in category "Heidelberg"

The following 200 files are in this category out of 366 total.

(previous page) (next page)
Mapbox and MapTiler uses Wikidata to help with the localization of their map products. If there are missing `name:*=*` tags on an OSM object, they pull labels from corresponding Wikidata entries.
For example, Heidelberg has 13 and 5 name:*=* tags in its place=city node and admin boundary relation, while Wikidata has 123 labels for the item.
User tchaddad recently finished a GSoC project to integrate wikidata=* tags into Nominatim. This project is intended to improve the search results by looking at Wikidata for additional relevance data.
The Name Suggestion Index uses Wikidata to provide brand identity and improved tagging in iD and Vespucci.

https://github.com/osmlab/name-suggestion-index
Yurik Astrakhan (OSM: nyuriks) created **Sophox**, a SPARQL endpoint for OSM. This service can use RDF federation to also query linked Wikidata items.

https://wiki.osm.org/Sophox
https://sophox.org/
Addendum: The OSM Wiki has the Wikibase extension installed. This is intended to provide better multilingual machine-readable data on OSM's tags.
CHAPTER 3
OSM Wikidata Matcher
(see second part of the presentation)
Linking OpenStreetMap and Wikidata

A semi-automated, user-assisted editing tool

https://osm.wikidata.link/

Edward Betts
Twitter: @EdwardBetts

Photo by Edward Betts // CC 0
Heidelberg

The software finds Wikidata items with a matching object in OSM

There are 111 matches in Heidelberg
OpenStreetMap and Wikidata match criteria

- entity type AND
- coordinates AND
- name OR street address OR identifier
Match on identifier

- Railway station code
- IATA airport code
- ICAO airport code
- FAA airport code
- ISO 3166-2 country code
- NRHP reference number
- UK Government Statistical Service code
- FIPS 6-4 (US counties)
- FIPS 5-2 (code for US states)
- USGS GNIS ID
- IBNR ID
- National Heritage List for England number
- EDUBase URN
- Admiralty number (lighthouse)
- Website
What about wikipedia tags?
Robot edits

Adding links individually is laborious

OpenStreetMap has strict rules about robot editing

Candidate matches need to be checked for false positives

Machine-assisted editing is a good compromise
Tunnels

Wikidata usually represents tunnels as a single item.

OSM often represents them as two ways, one for each direction of traffic or tunnel bore.

The software should be changed to recognise tunnels and add the Wikidata tag to both ways.

Photo by Edward Betts // CC BY-SA-3.0
Rivers

OSM represents rivers as linear relations

The matcher uses osm2pgsql to load OSM data into Postgis for querying

The only type of relation supported by osm2pgsql is a polygon

The data model it uses has no support for linear relations
Tram stops

Wikidata represents tram stops in both directions as a single item

OSM represents tram stops as a two node, one for each direction

The nodes are combined into a single relation

These relations are unsupported in osm2pgsql, so the matcher fails to identifier the relation
The difference between buildings and institutions

- OSM and Wikidata can have separate items for institutions and their buildings
- But not always
- The matcher has to handle this
Improving the information in Wikidata

- Merging duplicate Wikidata items
- Incomplete next level territorial administrative subdivisions data in Wikidata
OSM edits are grouped in a changeset

Makes it easier to review edits

Photo by Edward Betts // CC 0
The software is written in Python with websockets

- Flask
- SQLAlchemy
- PostgreSQL
- PostGIS
- gevent
- Bootstrap
- jQuery
- Leaflet
Building on the work of others by using APIs

- OSM Overpass
- OSM Nominatim
- OSM map API
- Wikibase API
- Wikipedia API
- Wikidata Query service
Mappers are using the tool

- Uploads by 200 mappers
- 8,800 changesets uploaded to OpenStreetMap
- 235,250 Wikidata links added
Tag count

There are 1,372,000 objects with a ‘wikidata’ tag in OpenStreetMap
Linking OpenStreetMap and Wikidata
A semi-automated, user-assisted editing tool

https://osm.wikidata.link/

Edward Betts
Twitter: @EdwardBetts