Wikidata + OpenStreetMap

വിക്കിഡാറ്റ

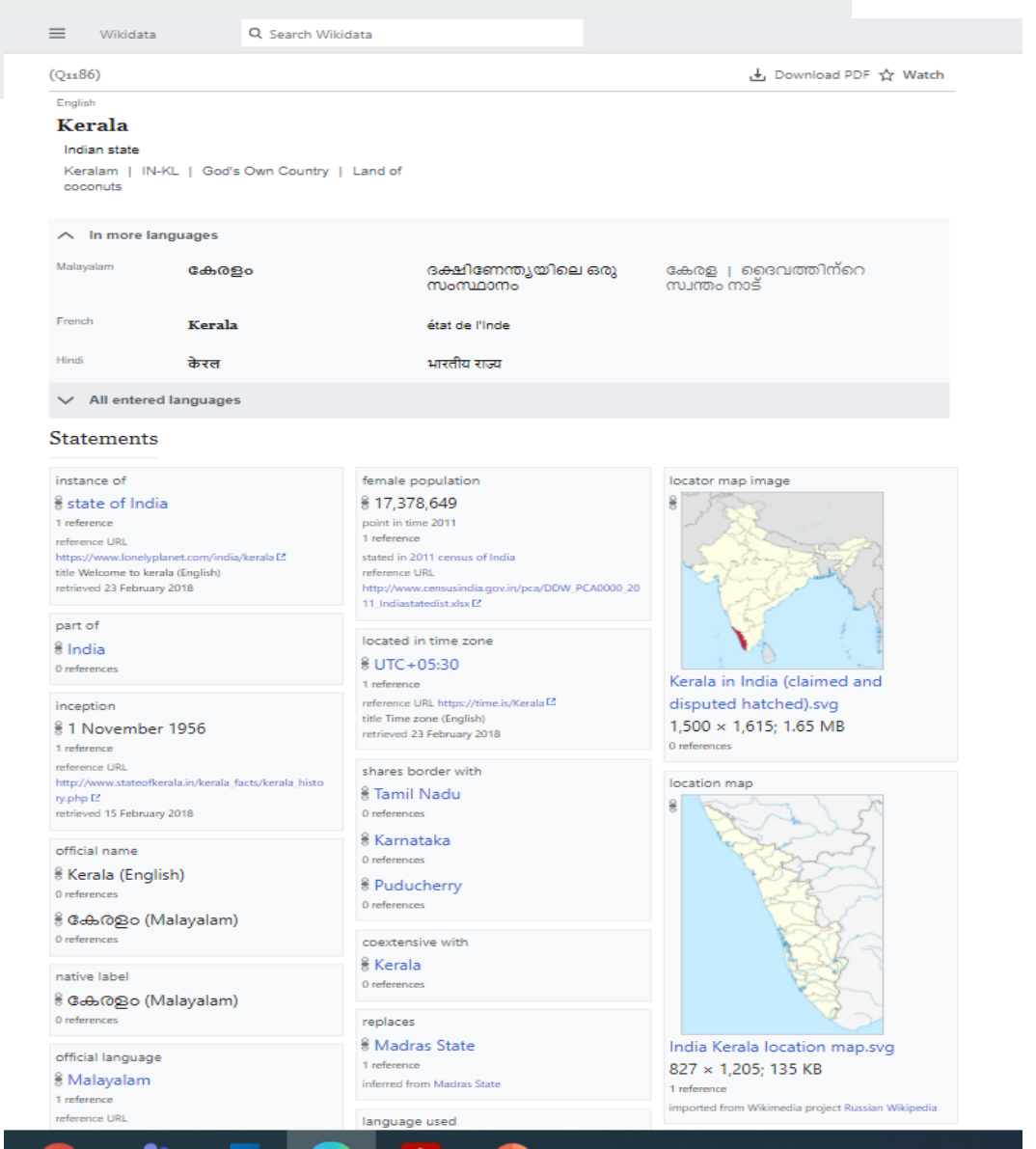
Naveen Francis
Digital Humanitarian
@naveenpf

CHAPTER 1

What is Wikidata?

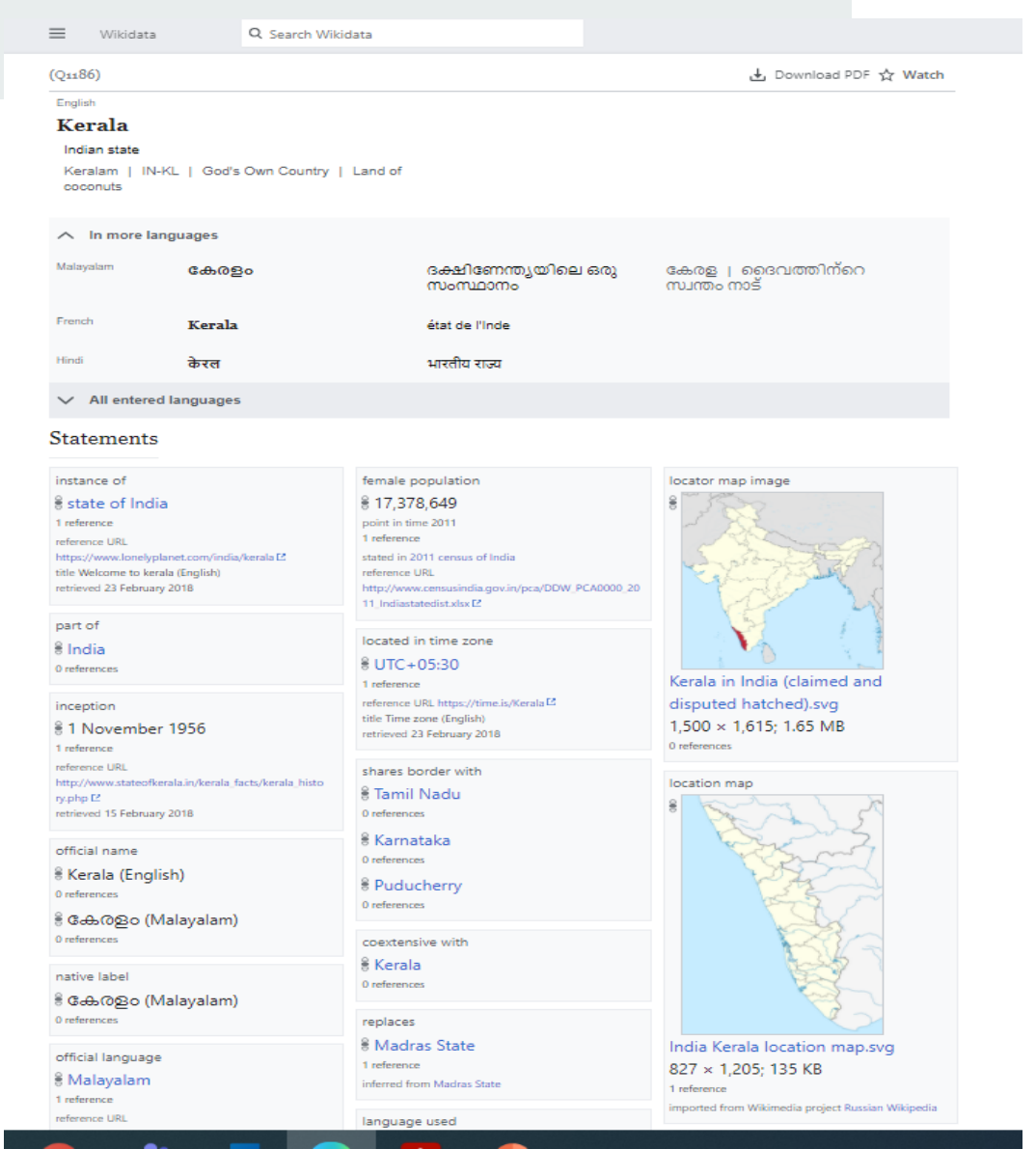
Wikidata is a CC0-licensed structured knowledge base under the Wikimedia umbrella that contains over 60 million items.

WIKIDATA



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.

del riu Neckar, que divideix la ciutat en dues meitats i transcorre d'est a oest per anar a desembocar al Rin uns 22 km més endavant, passada la ciutat veïna de Mannheim.

Al final d'aquesta vall hi trobem els turons que envolten el barri antic: el Königstuhl (568 m), el més alt, i el Gaisberg (375 m).

Barris

Heidelberg té 15 barris i 47 districtes municipals diferents. Alguns d'ells eren antics pobles veïns que, fins a l'any 1975, va anar absorbint la ciutat.

El barri antic

També anomenada "ciutat vella" per analogia amb el terme alemany corresponent, **Altstadt**. És el nucli històric de la ciutat i és on es troben la major part de les zones d'interès turístic. Està situat a la banda sud del Neckar, en l'estreta franja plana que hi ha entre la llera del riu i el turó del Königstuhl.

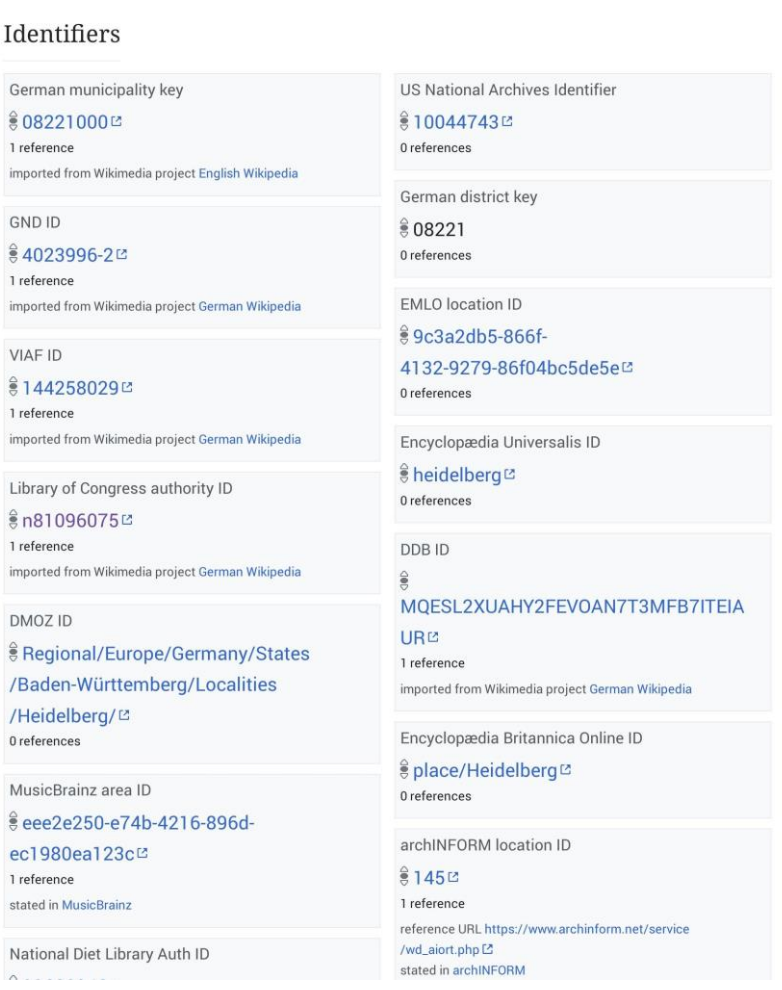
En aquest barri hi trobem el castell de Heidelberg (*Heidelberger Schloß*), el pont vell (*Alte Brücke*), les esglésies com la Heiliggeistkirche i la Jesuitenkirche, l'antiga presó dels estudiants (*Studentenkarzer*) o les places com ara la Karlsplatz, el Kornmarkt, la Martkplatz o la Universitätsplatz.

És al llarg de tot el barri antic que, durant les quatre setmanes d'advent, s'hi instal·la el mercat de Nadal (*Wehinachtsmarkt*).

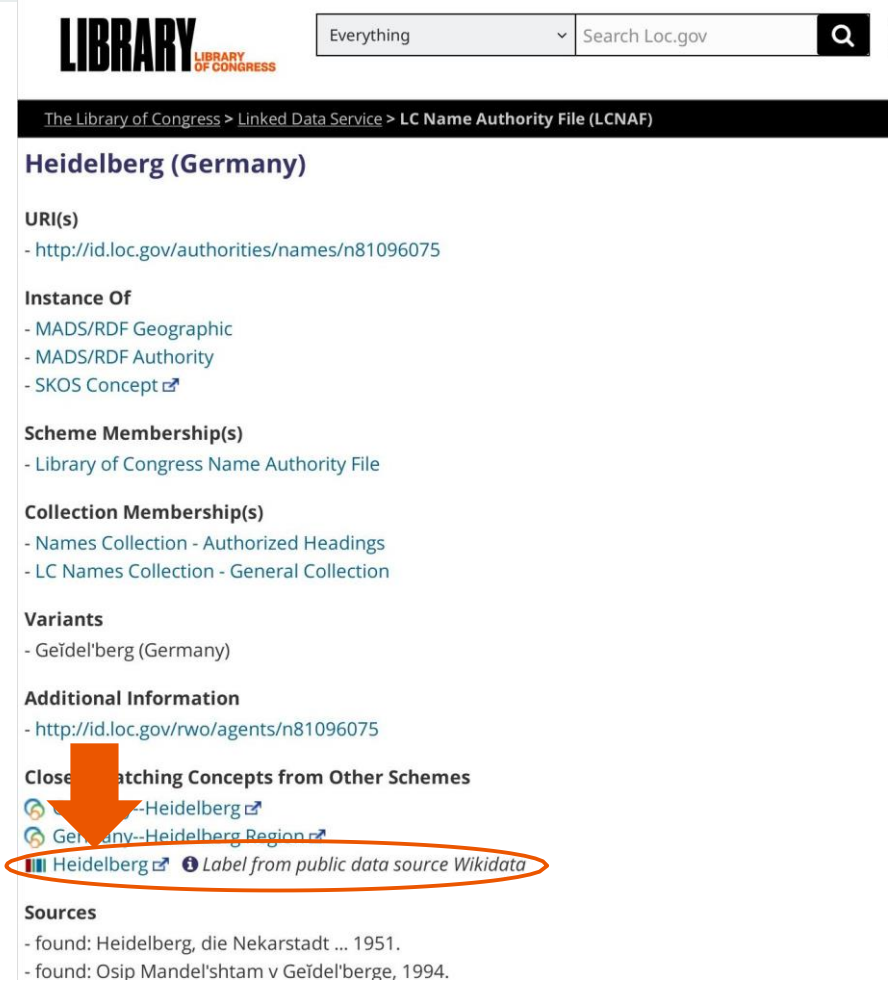
A l'extrem oest de l'Altstadt hi ha la plaça Bismarckplatz, que és el punt central del transport públic local i marca el final del barri.

Localització	
49° 24' 44" N, 8° 42' 36" E	
Estat	Alemanya
Estat federat	Baden-Württemberg
Regió administrativa	Karlsruhe
Capital de	Electorat del Palatinat
Població	
Total	160.601 (2017)
• Densitat	1.474,89 hab/km²
Geografia	
Part de	Àrea metropolitana Rin-Neckar, Karlsruhe i Q1496439
Superfície	108,89 km²
Banyat per	Neckar
Altitud	114 m
Limita amb	Mannheim
Història i celebracions	
Esdeveniment clau	Siege of Heidelberg

Wikidata probably has the largest collection of linked IDs to other third-party databases, data sets, and knowledge bases.

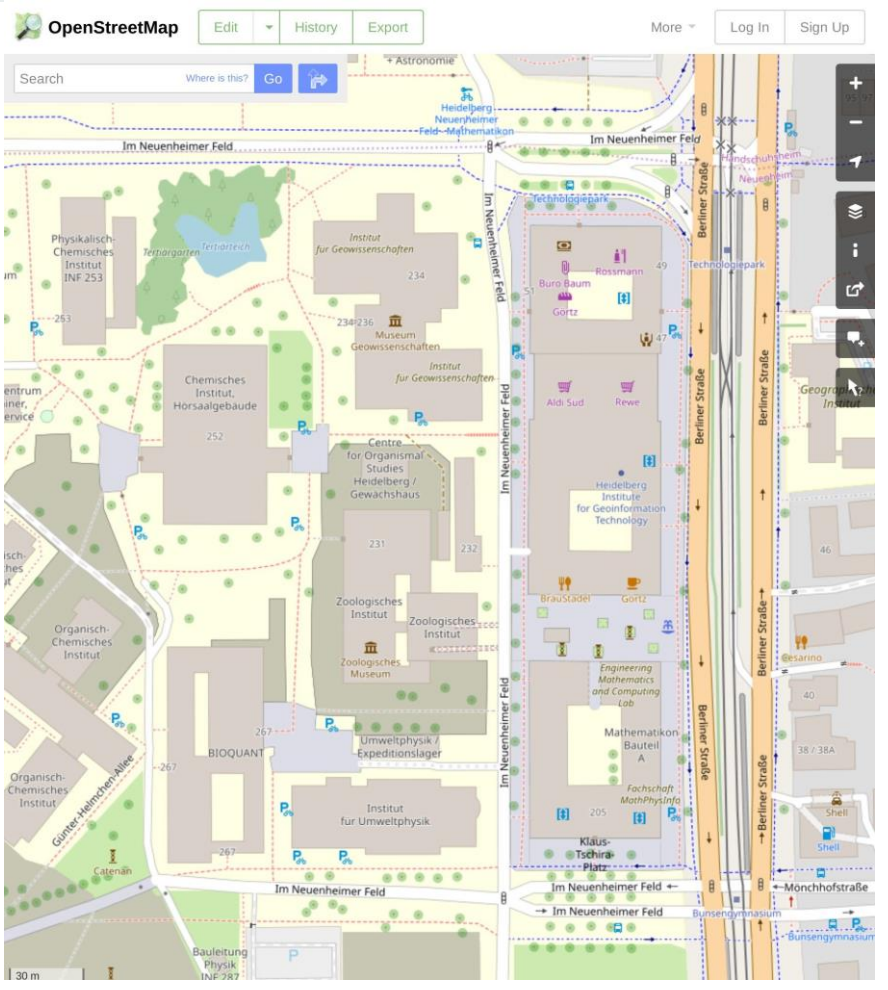


Wikidata has grown outside its original purpose and is fast becoming a central data hub especially for GLAM institutions (galleries, libraries, archives, and museums).



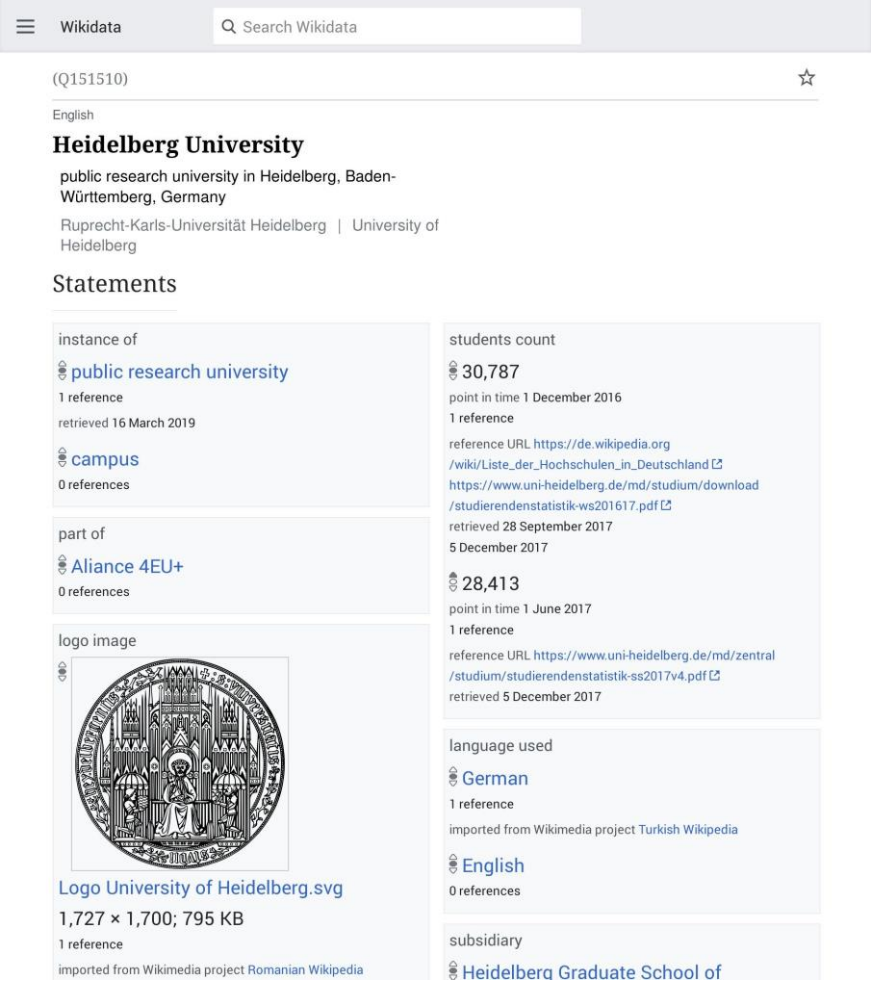
Wikidata data model

OpenStreetMap has nodes, ways, relations, tags, keys, values, roles, etc.



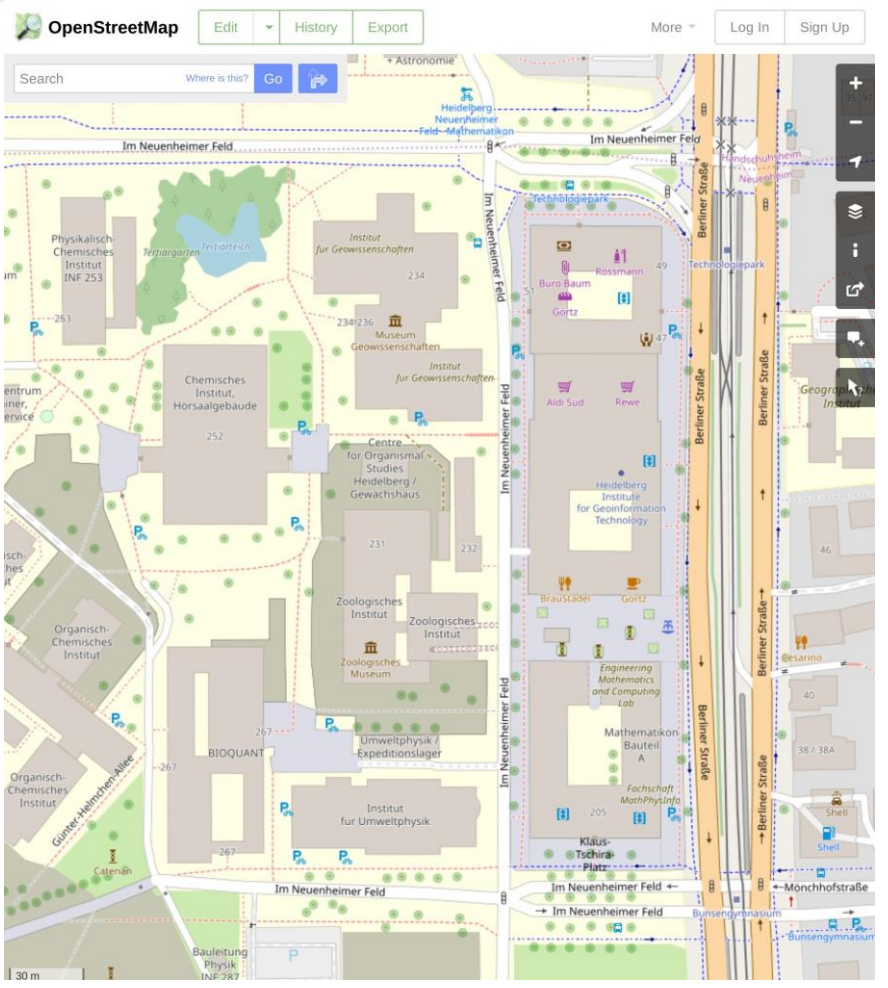
© OpenStreetMap contributors, ODbL 1.0

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.

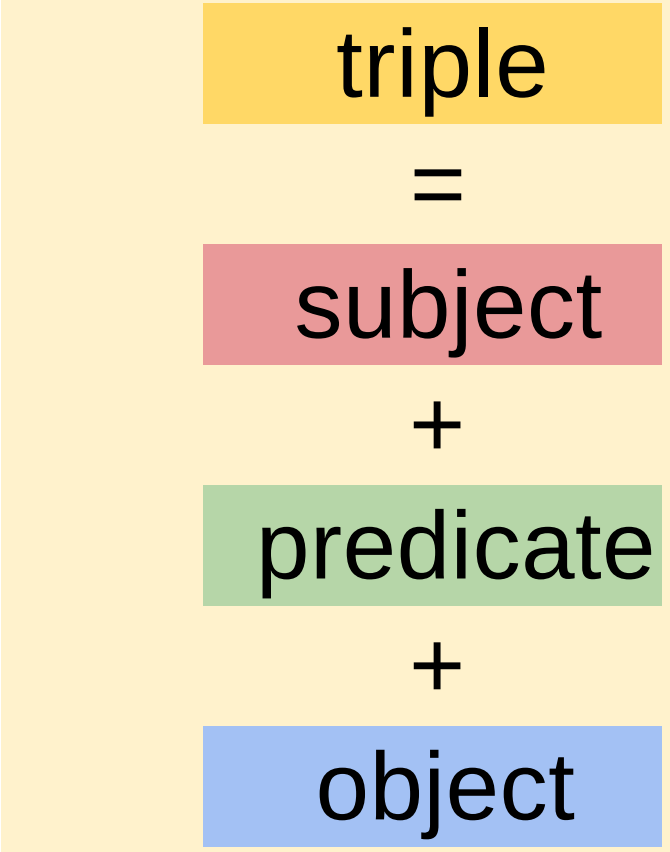
(Both projects struggle with ontology issues because both projects are examples of folksonomies.)



© OpenStreetMap contributors, ODbL 1.0

Wikidata data is stored as a bunch of triples. The most basic triples in Wikidata are called *statements*.

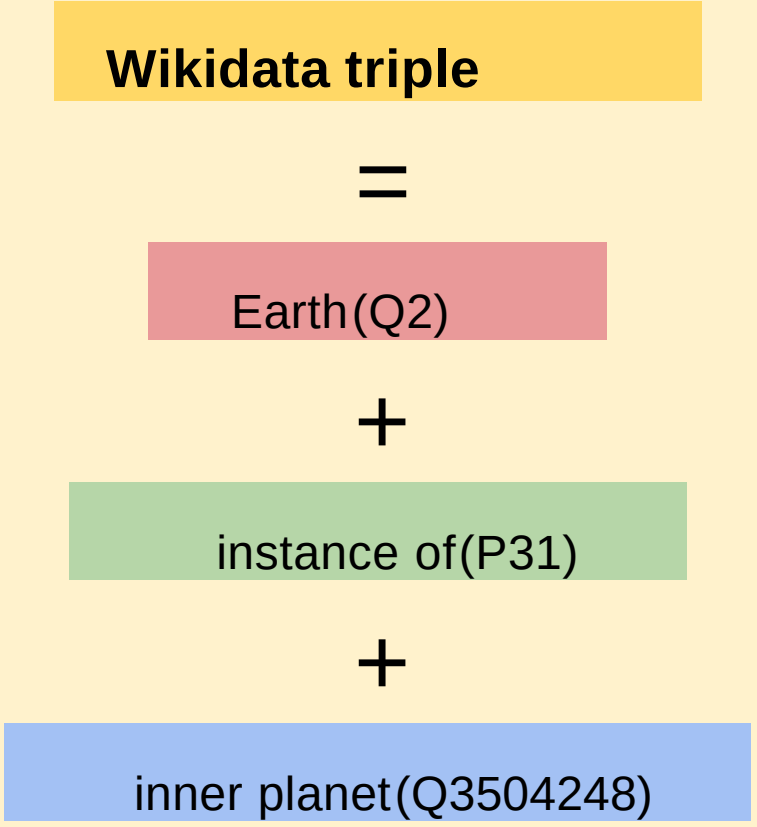
This design is heavily based on W3C's Resource Description Framework [RDF].



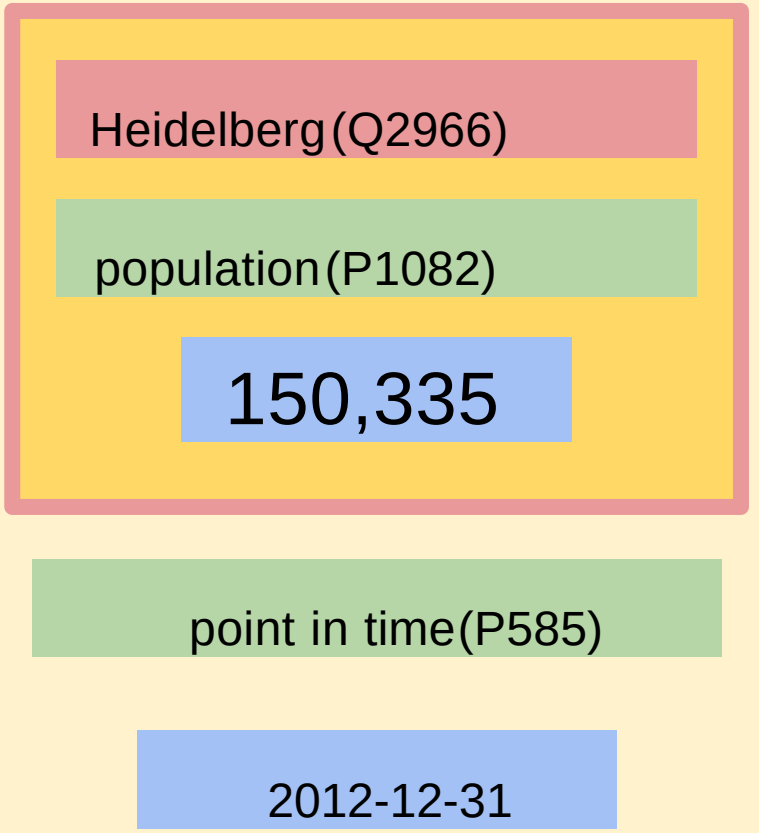
Triples, subjects, predicates, and objects are represented as IRIs.

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

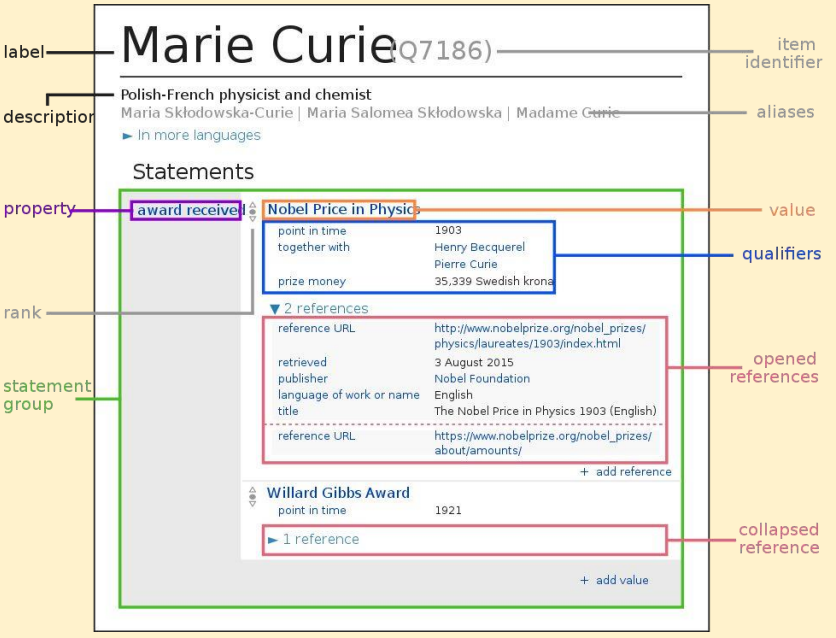
Or Q2 for short.



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.



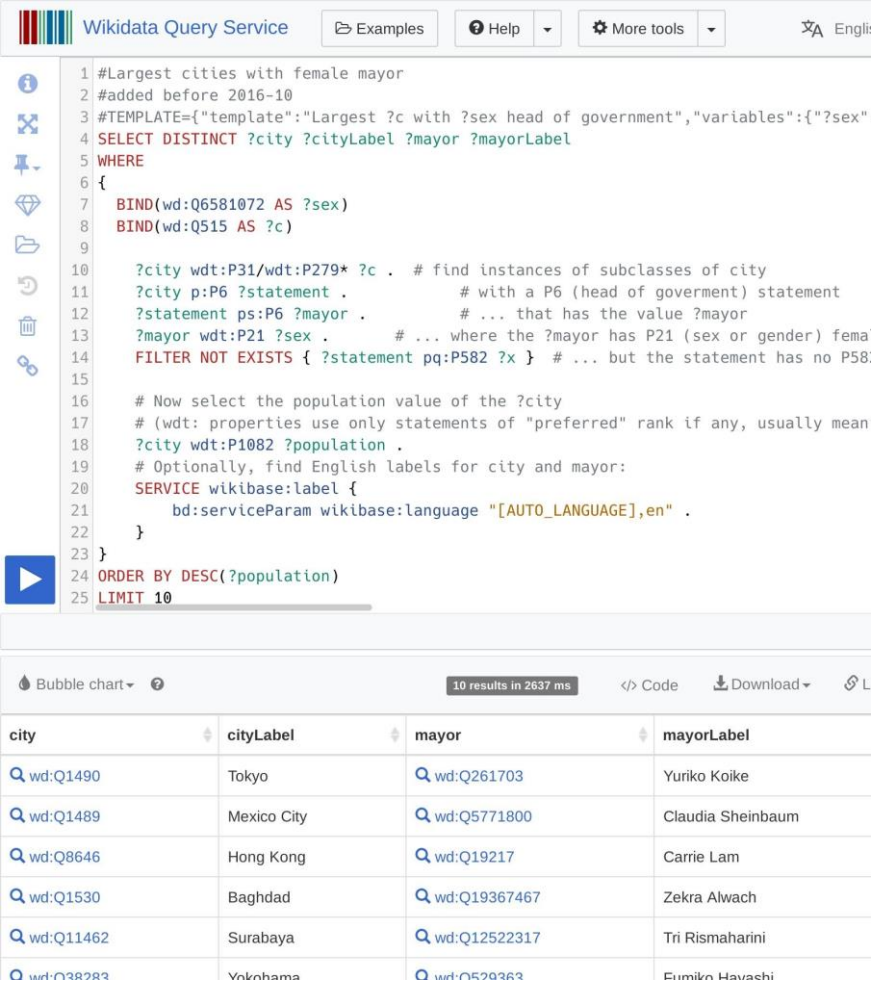
Aside: Wikidata also has lexemes (L-namespace) for lexicographic data, structured data for Wikimedia Commons (SDC) (M-namespace), and shape expressions (E-namespace).

Wikidata Query Service

The Wikidata Query Service (WDQS) is the official API for querying data from Wikidata.

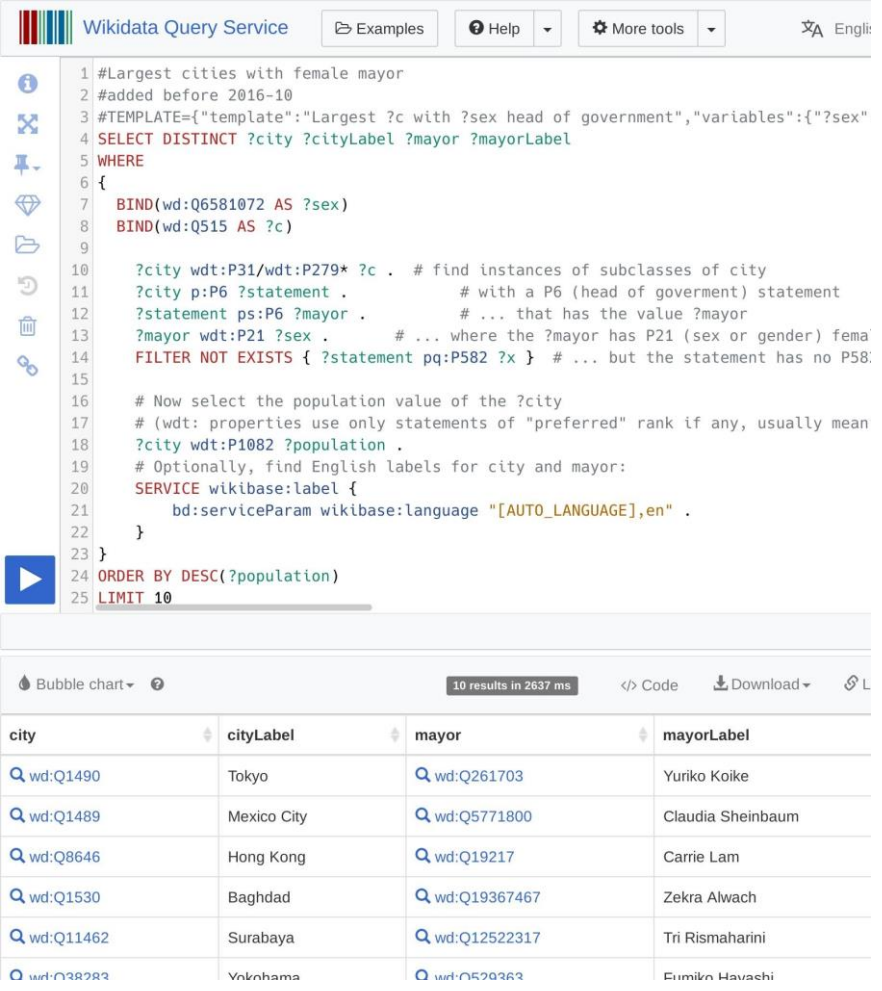
Think: Overpass API/Turbo for Wikidata.

https://query.wikidata.org



WDQS uses W3C’s SPARQL as the query language.

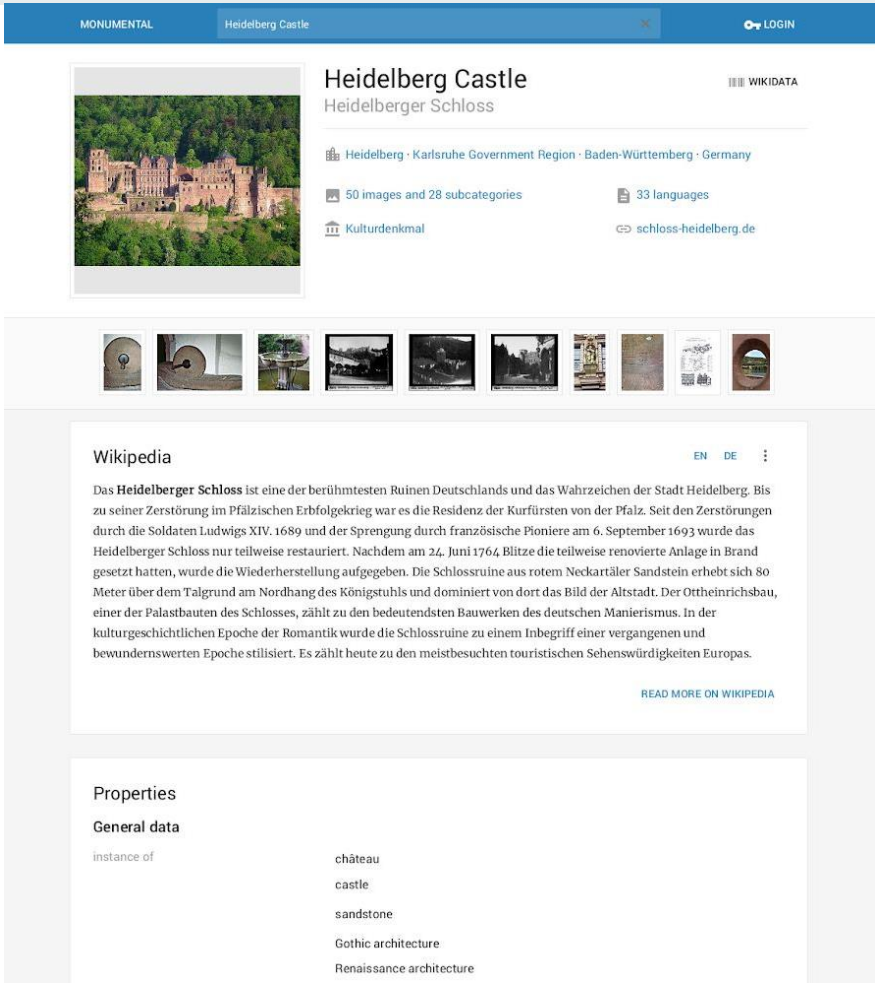
https://www.w3.org/TR/sparql11-query/
https://www.wikidata.org/wiki/Wikidata:SPARQL_tutorial



Wikidata tools and apps

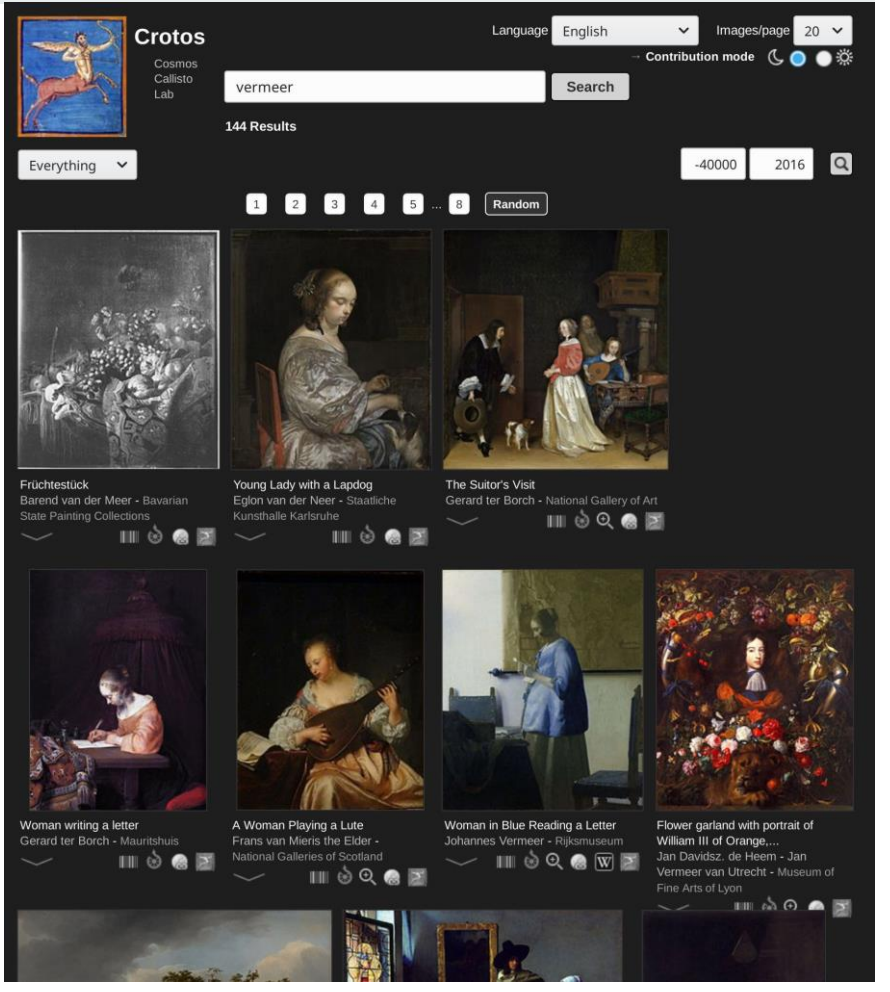
Monumental is a site that allows you to explore built heritage and learn more about monuments.

https://monumental.toolforge.org/



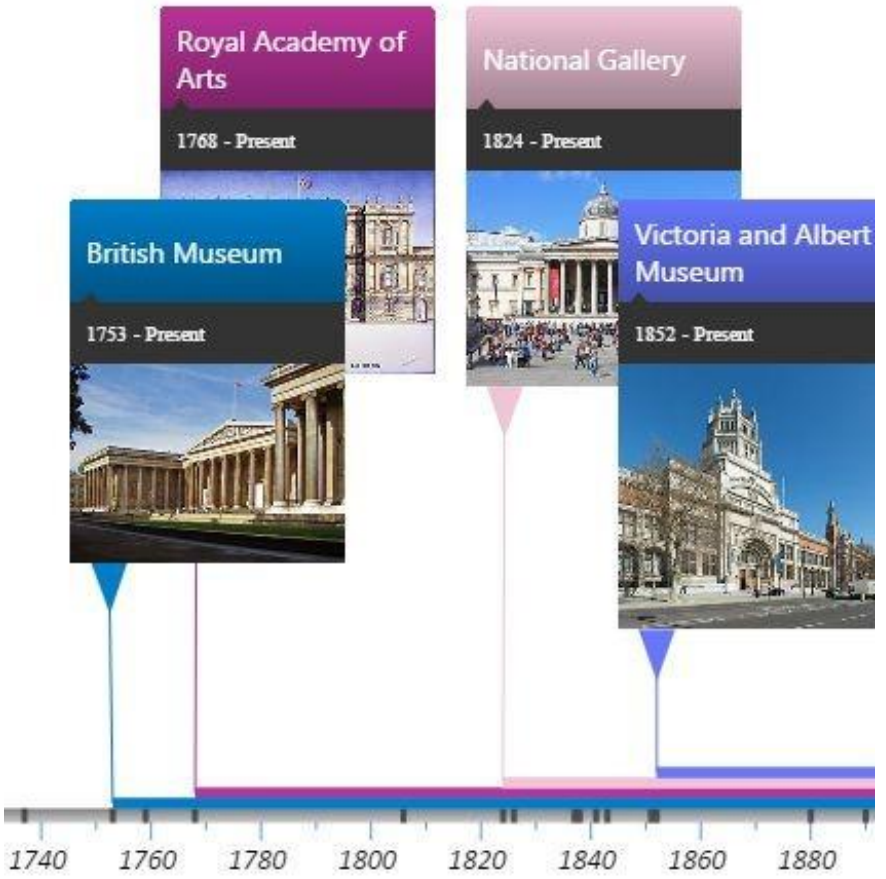
Crotos is a search and display engine for visual artworks, based on Wikidata and using Wikimedia Commons files.

http://zone47.com/crotos/



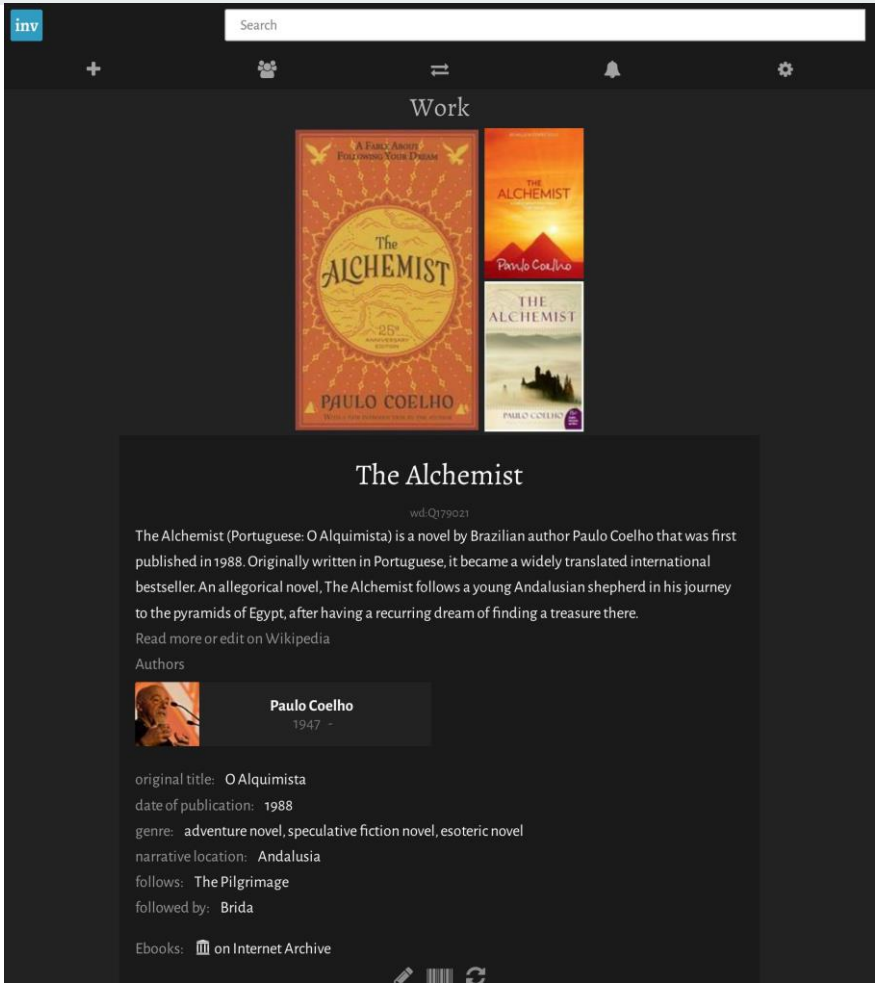
Histropedia allows you to generate interactive timelines with events linked to Wikipedia articles.

http://histropedia.com/



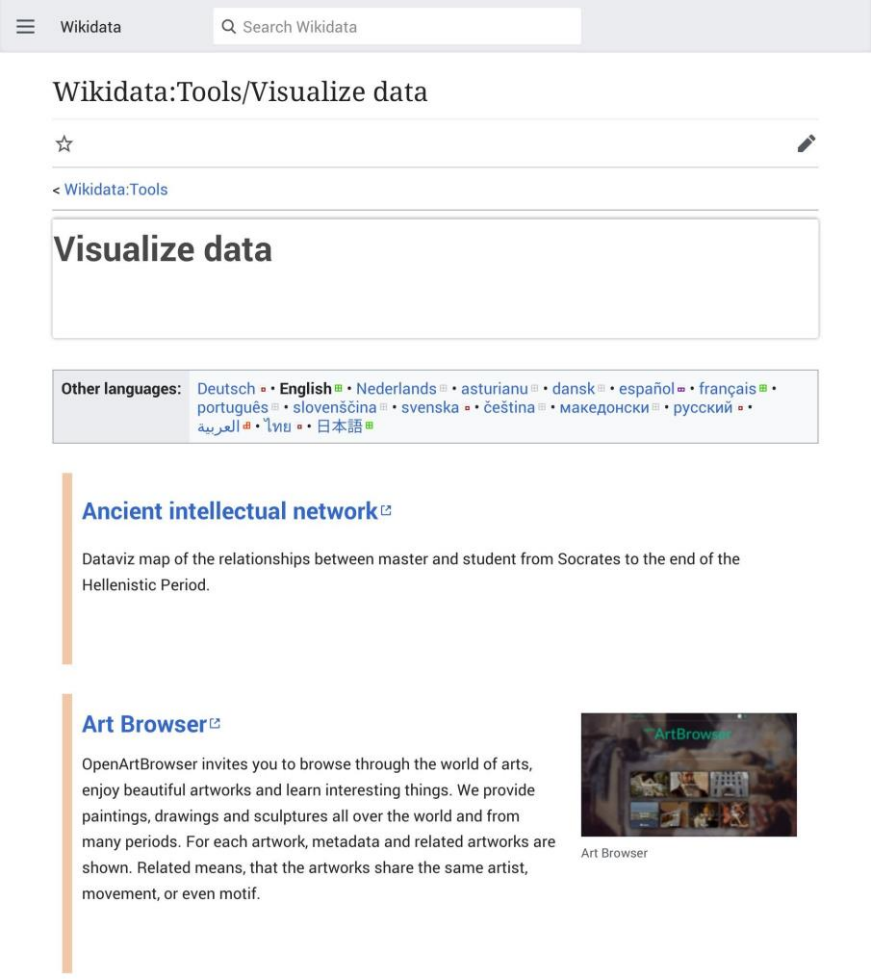
Inventaire: keep an inventory of your books!

https://inventaire.io/



More tools are listed on Wikidata:

https://www.wikidata.org/wiki/Wikidata:Tools/Visualize_data

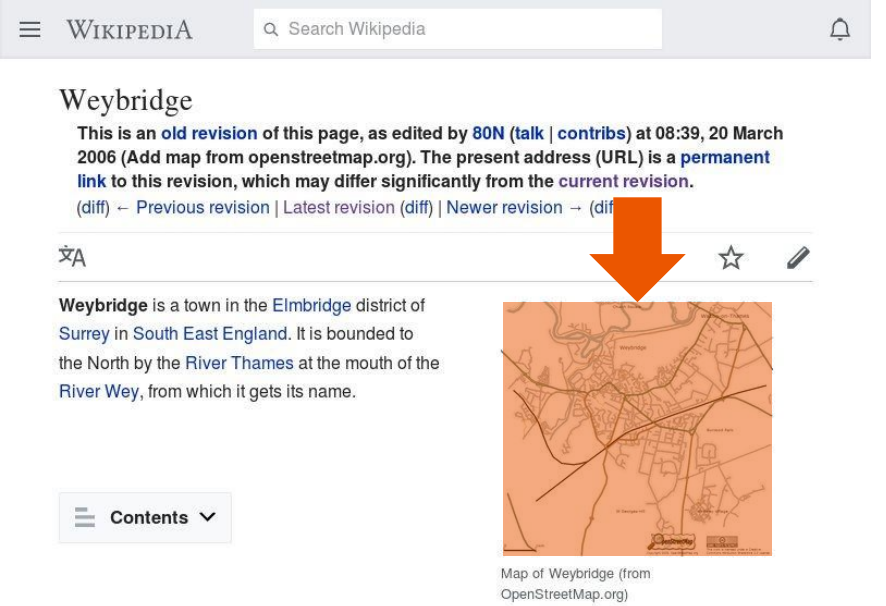


CHAPTER 2

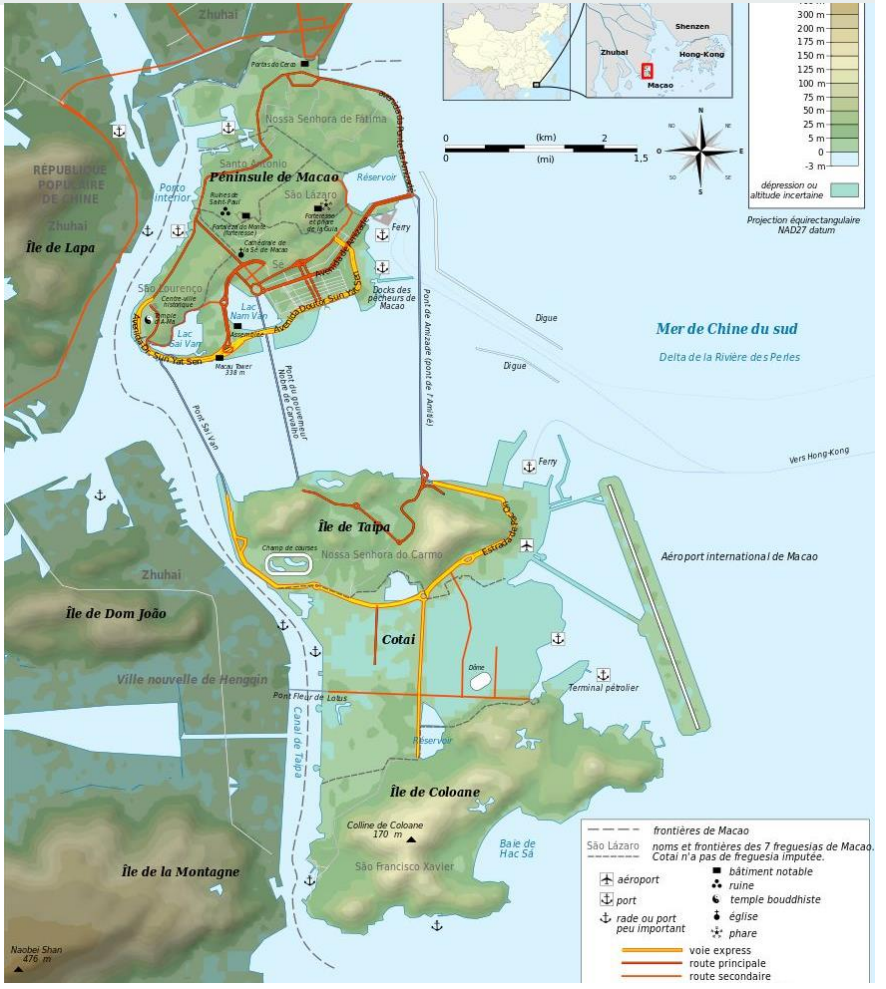
Linked Open Data

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.

Centro Storico is just as full of historic monuments as it is of designer boutiques and showrooms, restaurants and caffes - you will find almost all one comes to find in Milan for within not more than a quarter's walk.

The defunct old Milan district of "*Centro Storico*" included a broader area; the area covered in this guide corresponds to the *Quartiere 101 - Duomo* of the present-day *Zona 1*.

Get in

To get to the Centro Storico usually means getting to its very central point - the Piazza Duomo. Two metro lines cross ride underneath it, and many trams stop right next to the Piazza. The only train station within the Centro Storico is Milano Cadorna, on its western outskirts.

Map of Milan/Centro Storico

From Milano Malpensa Airport

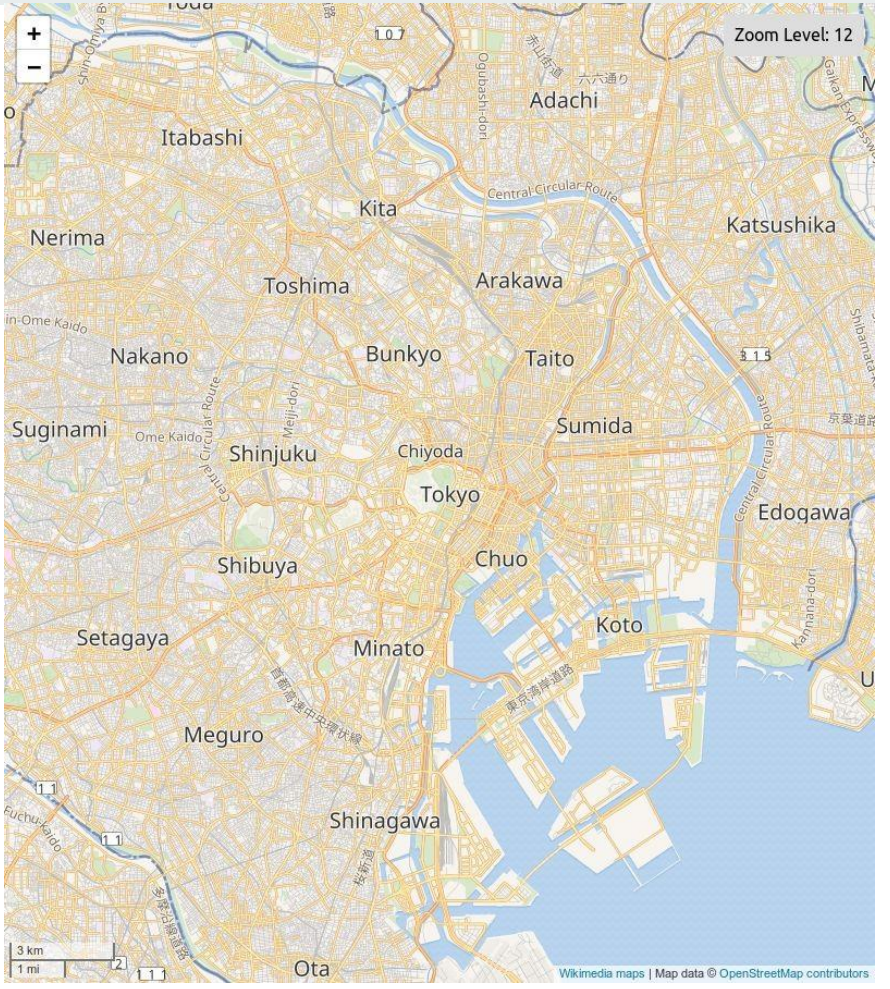
Take the Malpensa Express to Milano Cadorna station and from there simply enjoy the Centro Storico by walking eastwards, or take the metro for the 1 **Duomo** M1 M3. Trams 1 and 27 will also take you into the heart of the Centro Storico, over a route with some memorable views.

From Linate Airport

The bus lines 73 and X73 connect the Linate Airport with Piazza San Babila in the east of the Centro Storico, which has a station on the M1 network.

The Duomo metro station on M1 and

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.

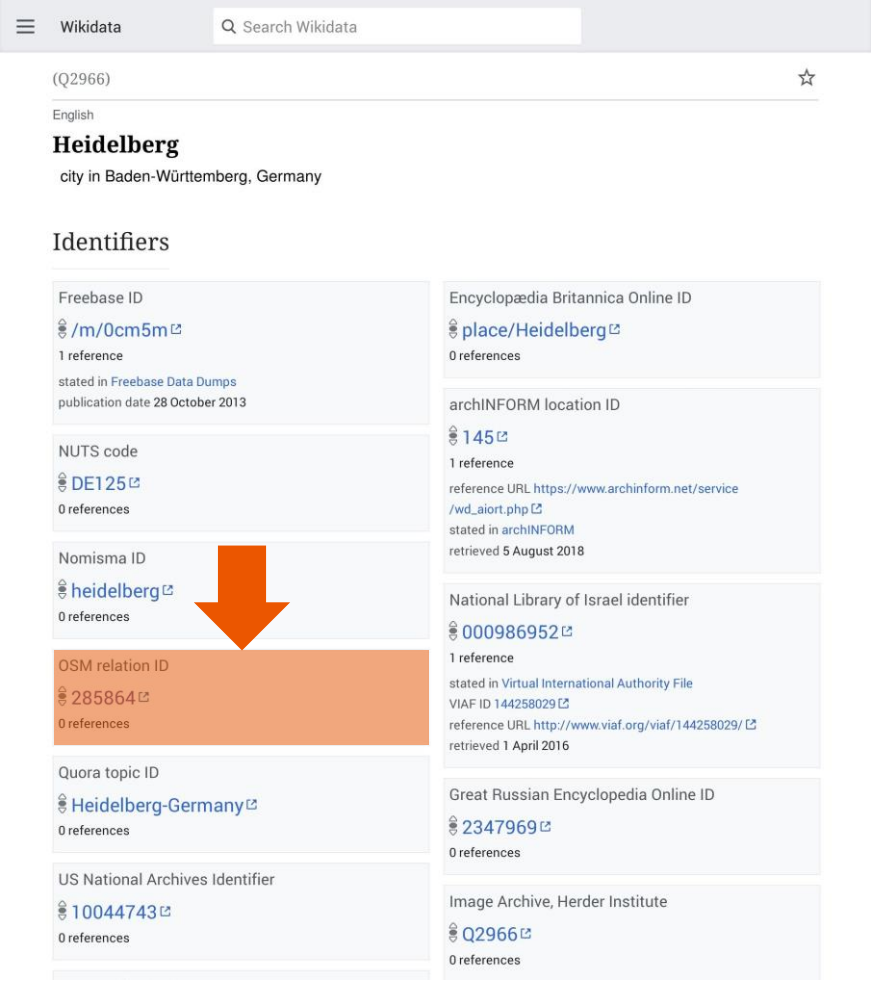
Why link OSM and Wikidata?

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

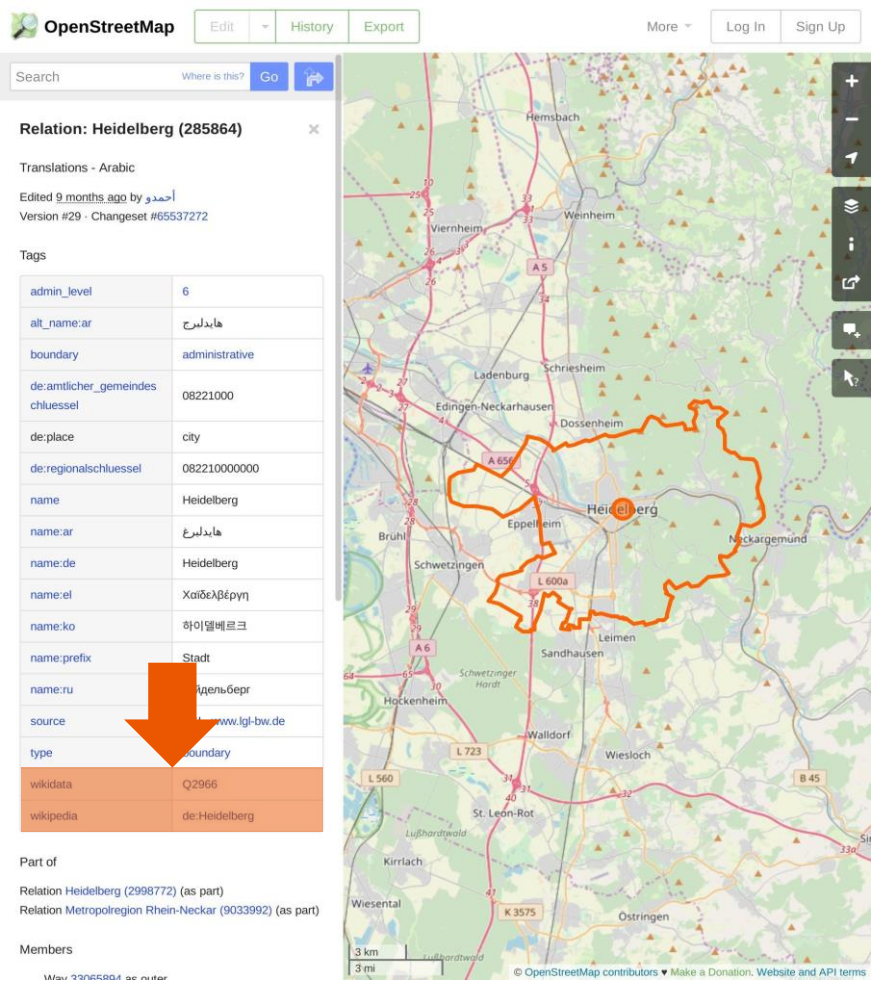
Linked Open data (LOD)

Wikidata items on places can link to OSM relations using the OSM relation ID (P402) property.

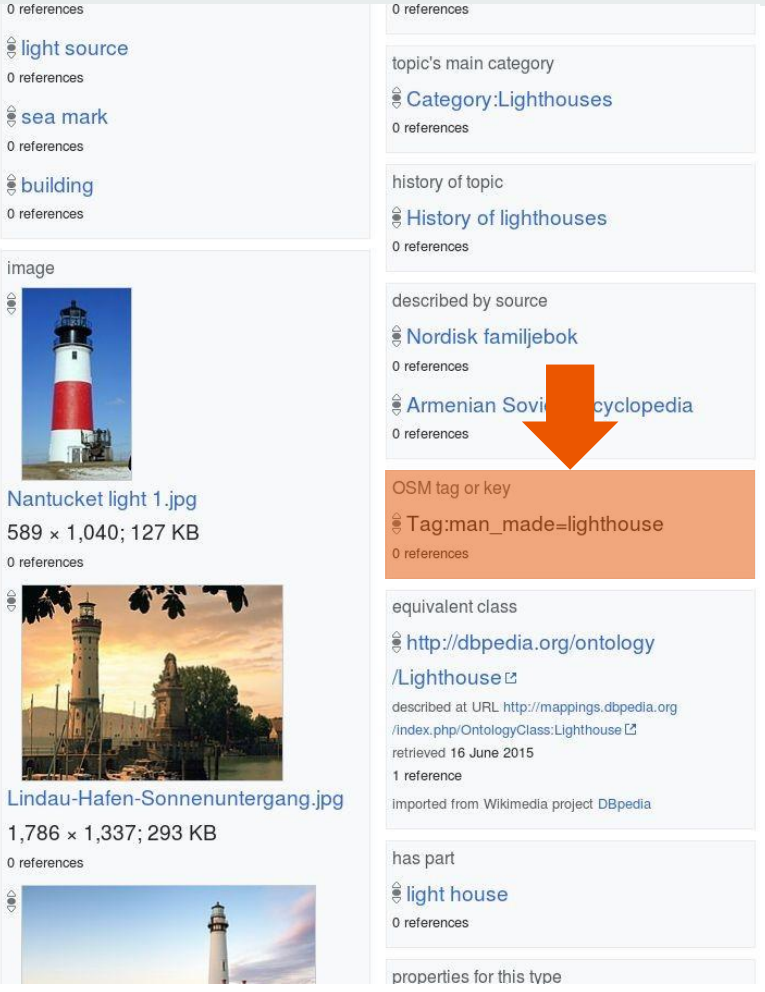


OSM objects can link to corresponding Wikipedia articles and Wikidata items using the wikipedia=* and wikidata=* tags respectively.

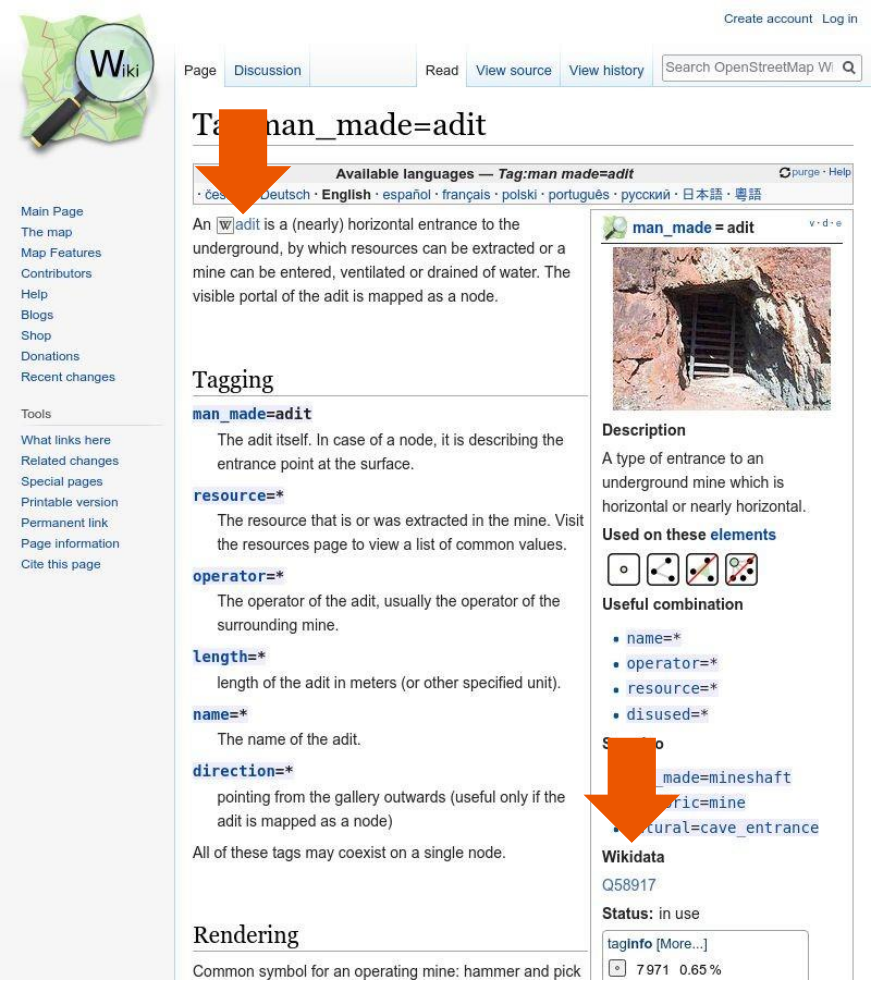
There are also secondary tags like brand:wikidata= and architect:wikidata=*.*



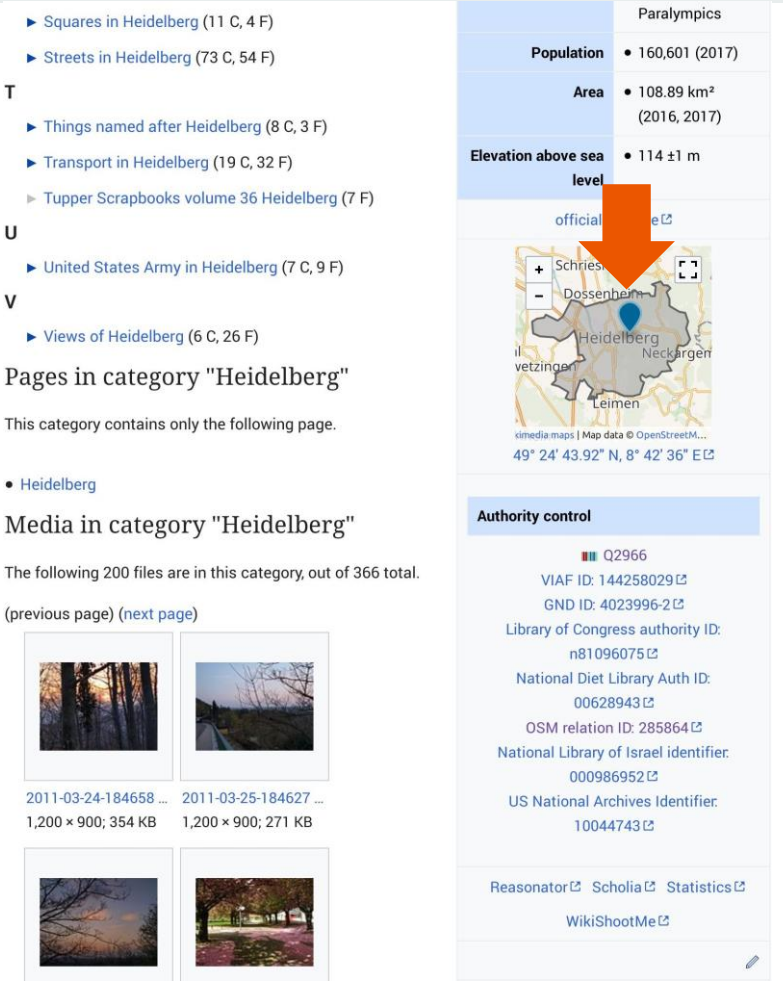
Ontology: Wikidata items about features can link to “equivalent” OSM features using the OSM tag or key (P1282) property.



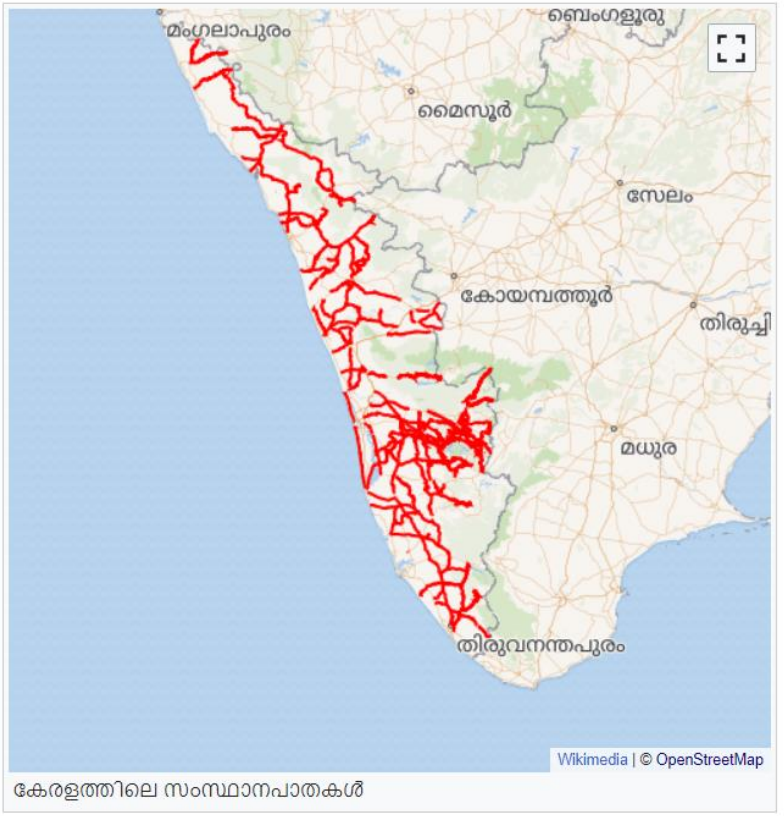
Ontology: Tag definitions on the OSM Wiki can link to “equivalent” Wikipedia articles and Wikidata items.



Wikimedia maps can show geometry from OSM.

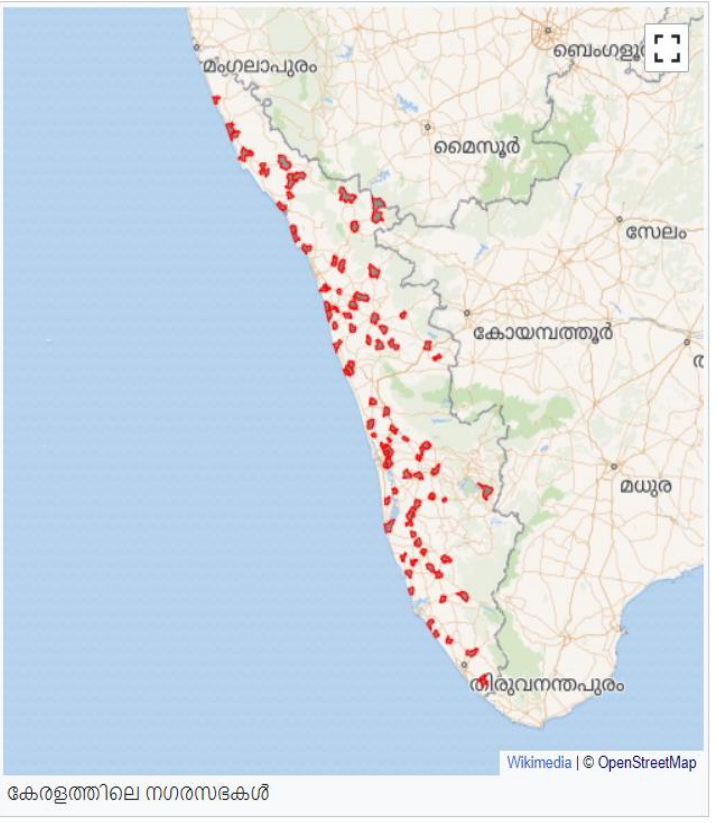


കേരളത്തിലെ സംസ്ഥാനപാതകൾ



കേരളത്തിലെ സംസ്ഥാനപാതകൾ

കേരളത്തിലെ നഗരസഭകൾ



കേരളത്തിലെ നഗരസഭകൾ

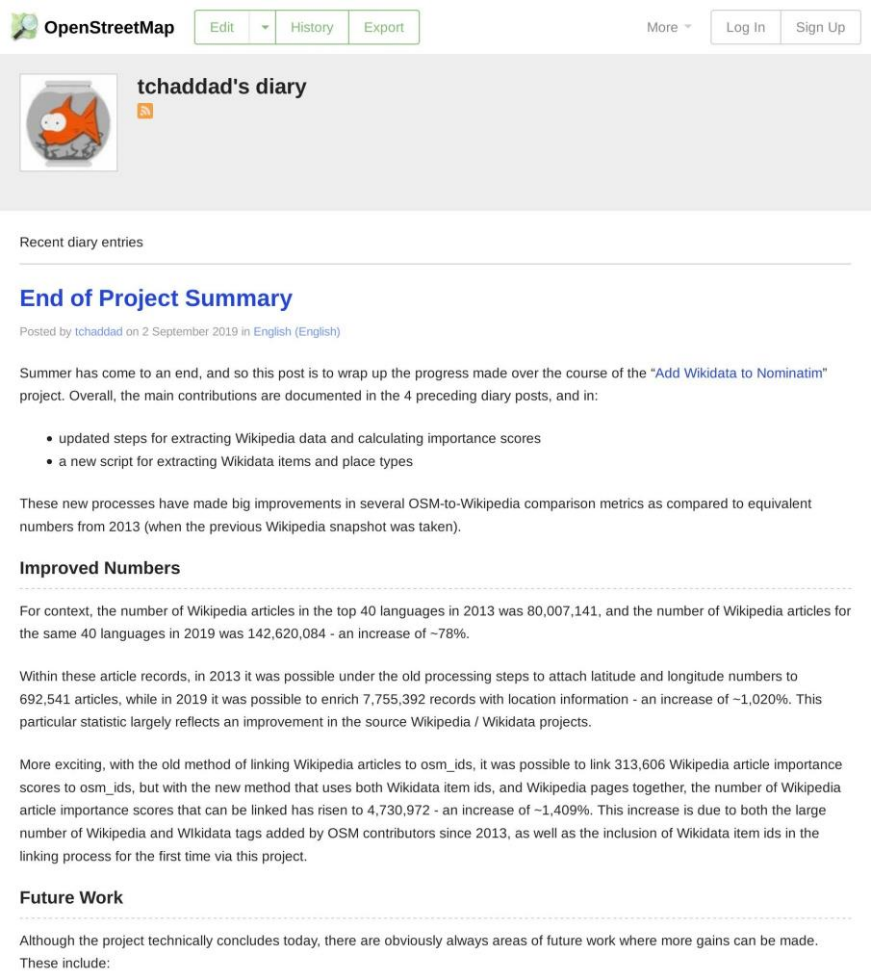
https://ml.wikipedia.org/wiki/List_of_state_highways_in_Kerala
https://ml.wikipedia.org/wiki/Municipalities_of_Kerala

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.



OpenStreetMap Edit History Export More Log In Sign Up

tchaddad's diary

Recent diary entries

End of Project Summary

Posted by tchaddad on 2 September 2019 in English (English)

Summer has come to an end, and so this post is to wrap up the progress made over the course of the "Add Wikidata to Nominatim" project. Overall, the main contributions are documented in the 4 preceding diary posts, and in:

- updated steps for extracting Wikipedia data and calculating importance scores
- a new script for extracting Wikidata items and place types

These new processes have made big improvements in several OSM-to-Wikipedia comparison metrics as compared to equivalent numbers from 2013 (when the previous Wikipedia snapshot was taken).

Improved Numbers

For context, the number of Wikipedia articles in the top 40 languages in 2013 was 80,007,141, and the number of Wikipedia articles for the same 40 languages in 2019 was 142,620,084 - an increase of ~78%.

Within these article records, in 2013 it was possible under the old processing steps to attach latitude and longitude numbers to 692,541 articles, while in 2019 it was possible to enrich 7,755,392 records with location information - an increase of ~1,020%. This particular statistic largely reflects an improvement in the source Wikipedia / Wikidata projects.

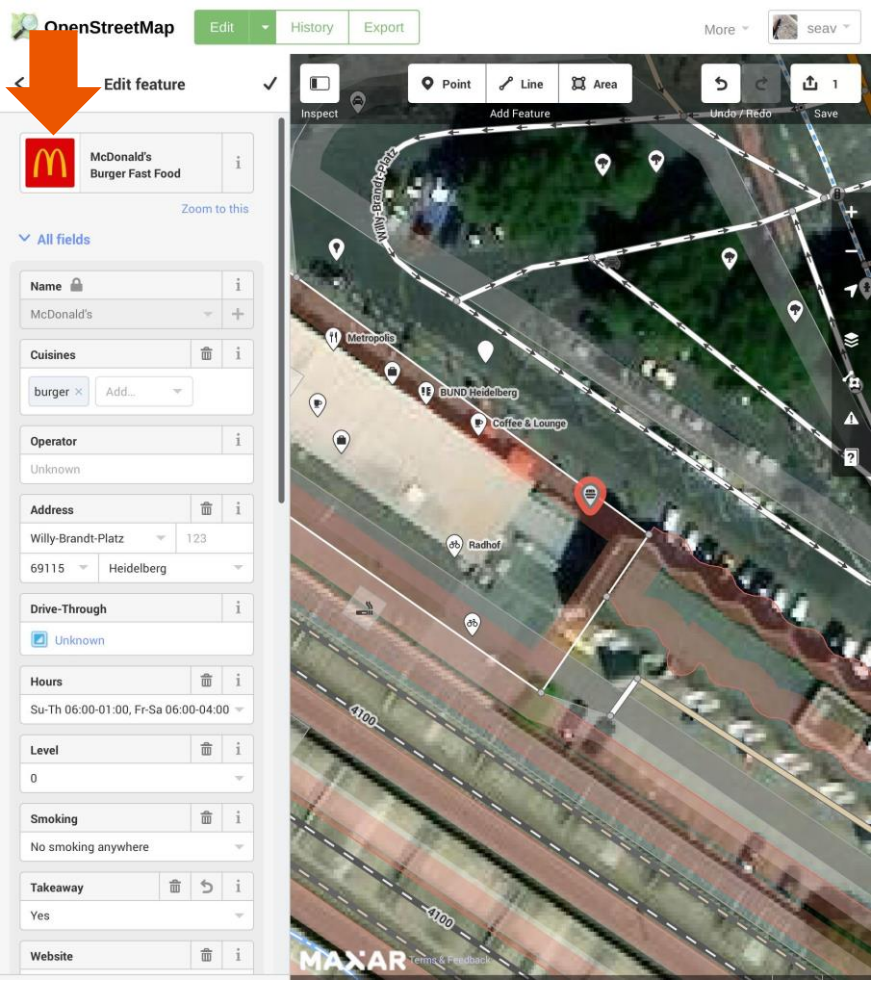
More exciting, with the old method of linking Wikipedia articles to osm_ids, it was possible to link 313,606 Wikipedia article importance scores to osm_ids, but with the new method that uses both Wikidata item ids, and Wikipedia pages together, the number of Wikipedia article importance scores that can be linked has risen to 4,730,972 - an increase of ~1,409%. This increase is due to both the large number of Wikipedia and WIkidata tags added by OSM contributors since 2013, as well as the inclusion of Wikidata item ids in the linking process for the first time via this project.

Future Work

Although the project technically concludes today, there are obviously always areas of future work where more gains can be made. These include:

The Name Suggestion Index uses Wikidata to provide brand identity and improved tagging in iD and Vespucci.

https://github.com/osmlab/name-suggestion-index

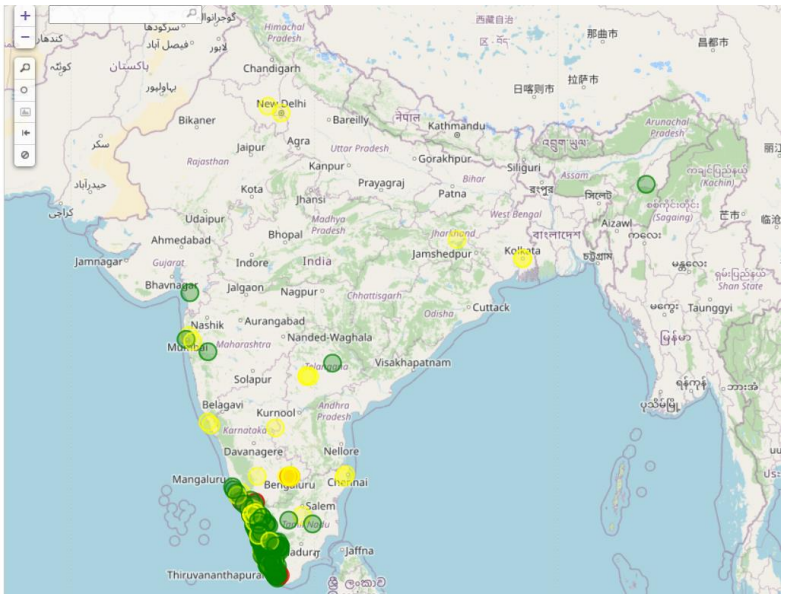


Brand - Name Suggestion Index

# Federal Bank federalbank-34f411 {"include":["in"]} Search Overpass Turbo Search Google Search Wikipedia	344	"amenity": "bank" "brand": "Federal Bank" "brand:wikidata": "Q2044983" "brand:wikipedia": "en:Federal Bank" "name": "Federal Bank"	**Federal Bank** "large commercial bank located in India" Q2044983 https://www.federalbank.co.in/			

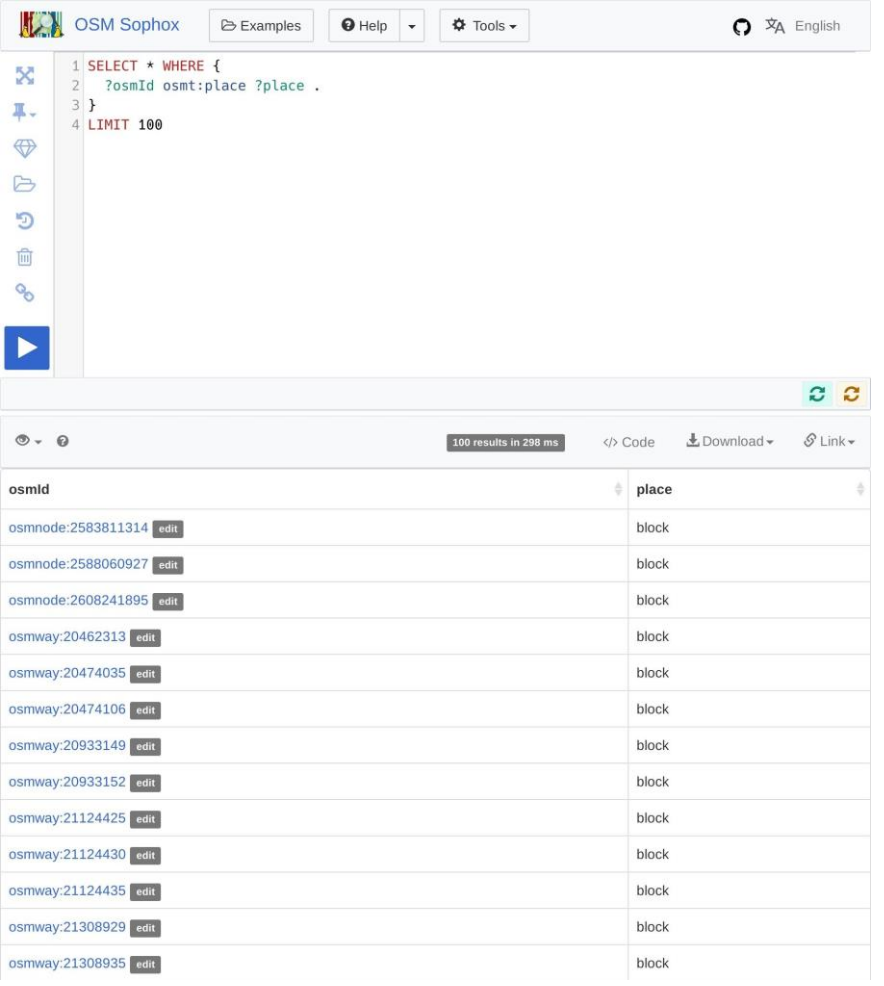
```
[out:json][timeout:100];
(nwr["name"="Federal Bank"];);
out body;
>;
out skel qt;

{{style:
node[name=Federal Bank],
way[name=Federal Bank],
relation[name=Federal Bank]
{ color:red; fill-color:red; }
node[amenity=bank][name=Federal Bank],
way[amenity=bank][name=Federal Bank],
relation[amenity=bank][name=Federal Bank]
{ color:yellow; fill-color:yellow; }
node[amenity=bank][name=Federal Bank][brand=Federal Bank][brand:wikidata=Q2044983],
way[amenity=bank][name=Federal Bank][brand=Federal Bank][brand:wikidata=Q2044983],
relation[amenity=bank][name=Federal Bank][brand=Federal Bank][brand:wikidata=Q2044983]
{ color:green; fill-color:green; }
}}
```



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/



Federated Queries

Nature [edit | edit source]

Threatened species of trees [edit | edit source]

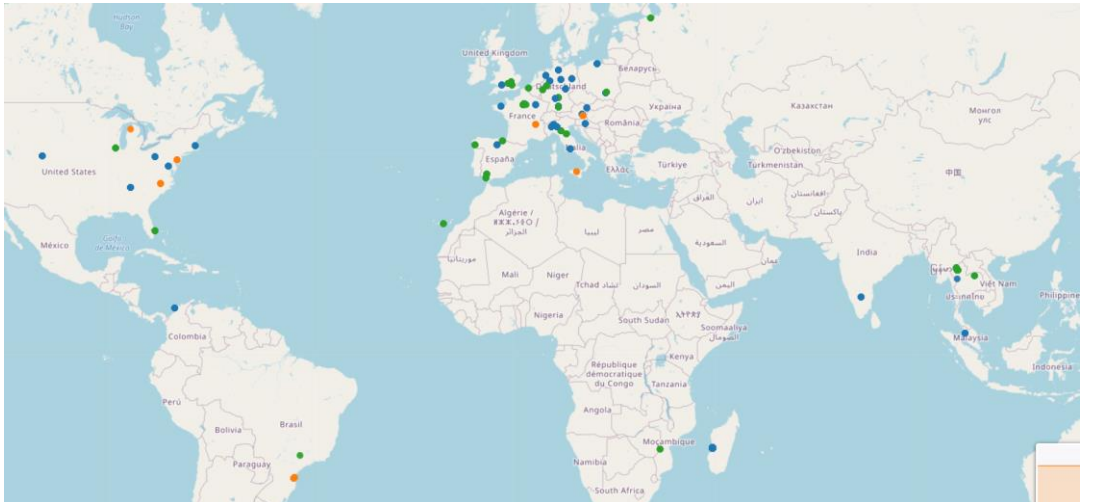
Displays an interactive map of trees whose species is categorized as threatened on the IUCN Red List.

```
#defaultView:Map
SELECT ?tree ?taxon ?taxonLabel ?layer ?coordinates WHERE {
  # Query OpenStreetMap for trees and their taxa
  ?tree osmt:natural "tree";
        (osmt:genus:wikidata|osmt:species:wikidata|osmt:taxon:wikidata) ?taxon;
        osmm:loc ?coordinates.

  # Query Wikidata for
  SERVICE <https://query.wikidata.org/sparql> {
    # Taxa categorized as threatened on the IUCN Red List
    VALUES (?status ?layer) {
      (wd:Q278113 "Vulnerable (VU)")
      (wd:Q11394 "Endangered (EN)")
      (wd:Q219127 "Critically endangered (CR)")
    }
    ?taxon wdt:P141 ?status.

    # Get the English common name
    OPTIONAL {
      ?taxon wdt:P1843 ?taxonLabel.
      FILTER(LANG(?taxonLabel) = "en")
    }
  }
}
```

Run it (edit query)



Node: Guaiacum officinale (3476635853)

Version #5

fixing all trees -- with species:wikidata - found them with http://tinyurl.com/y72v4yv3 - using https://wiki.openstreetmap.org/wiki/Wikidata%2BOSM

Edited about 3 years ago by nyuriks · Changeset #51982127
Location: 12.9533522, 77.5848743

Tags

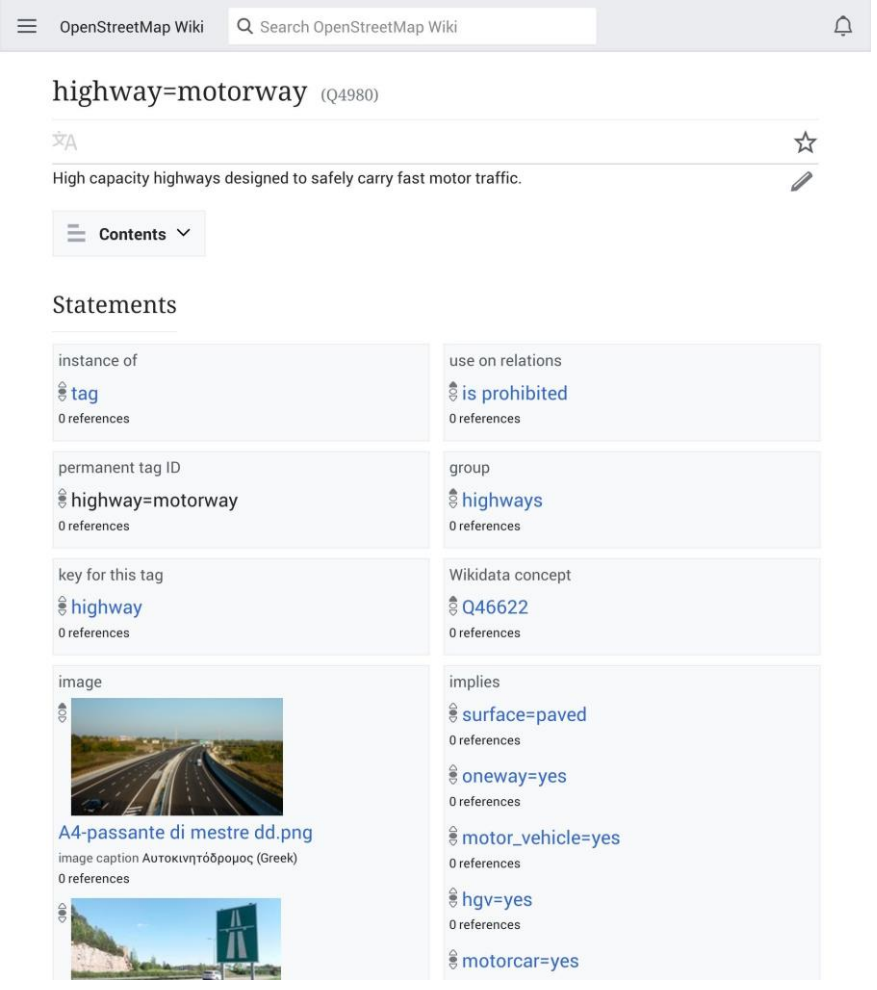
name	Guaiacum officinale
natural	tree
species:wikidata	Q2051175
species:wikipedia	en:Guaiacum officinale

Download XML · View History

Addendum: The OSM Wiki has the Wikibase extension installed. This is intended to provide better multilingual machine-

readable data on OSM's

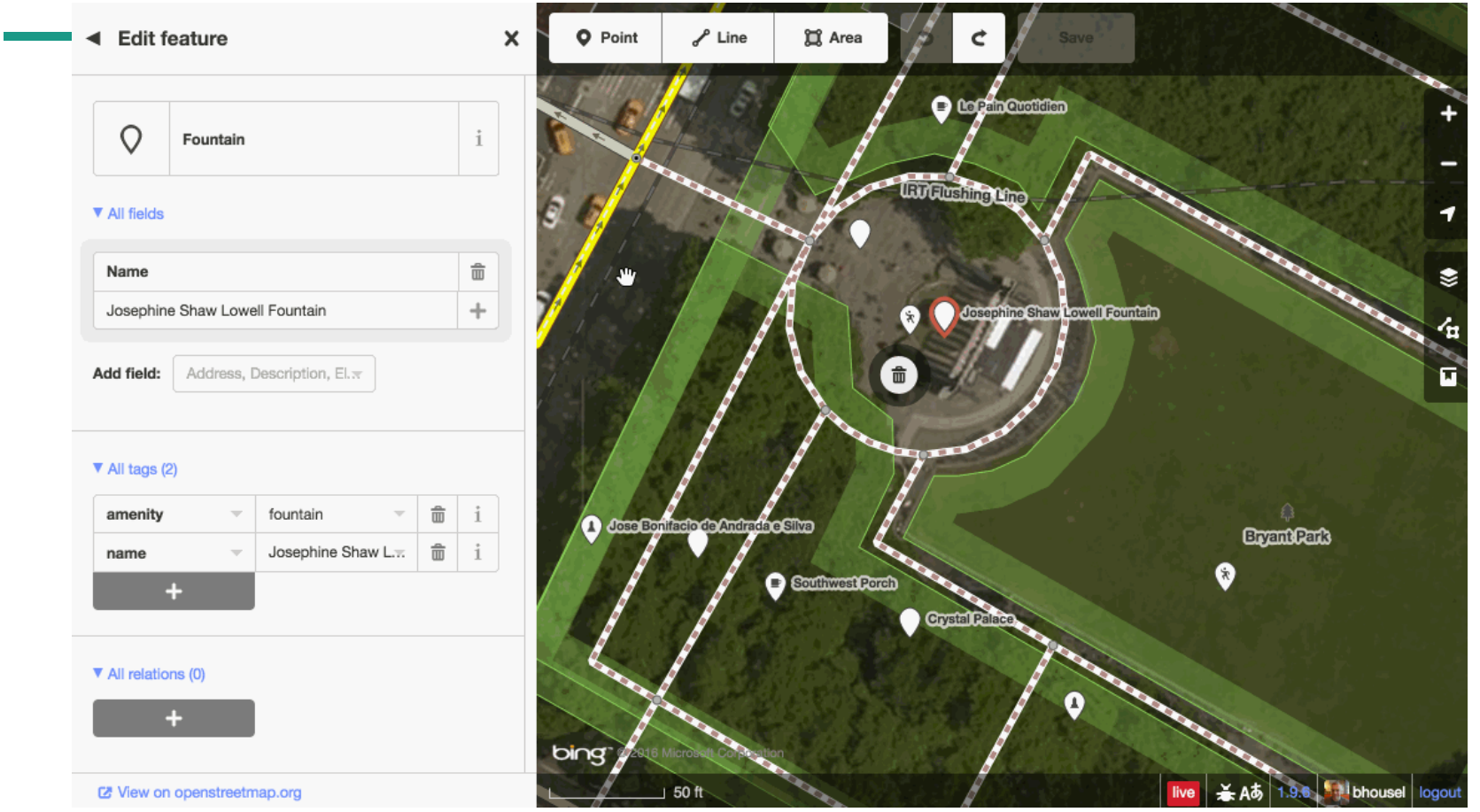
tags.



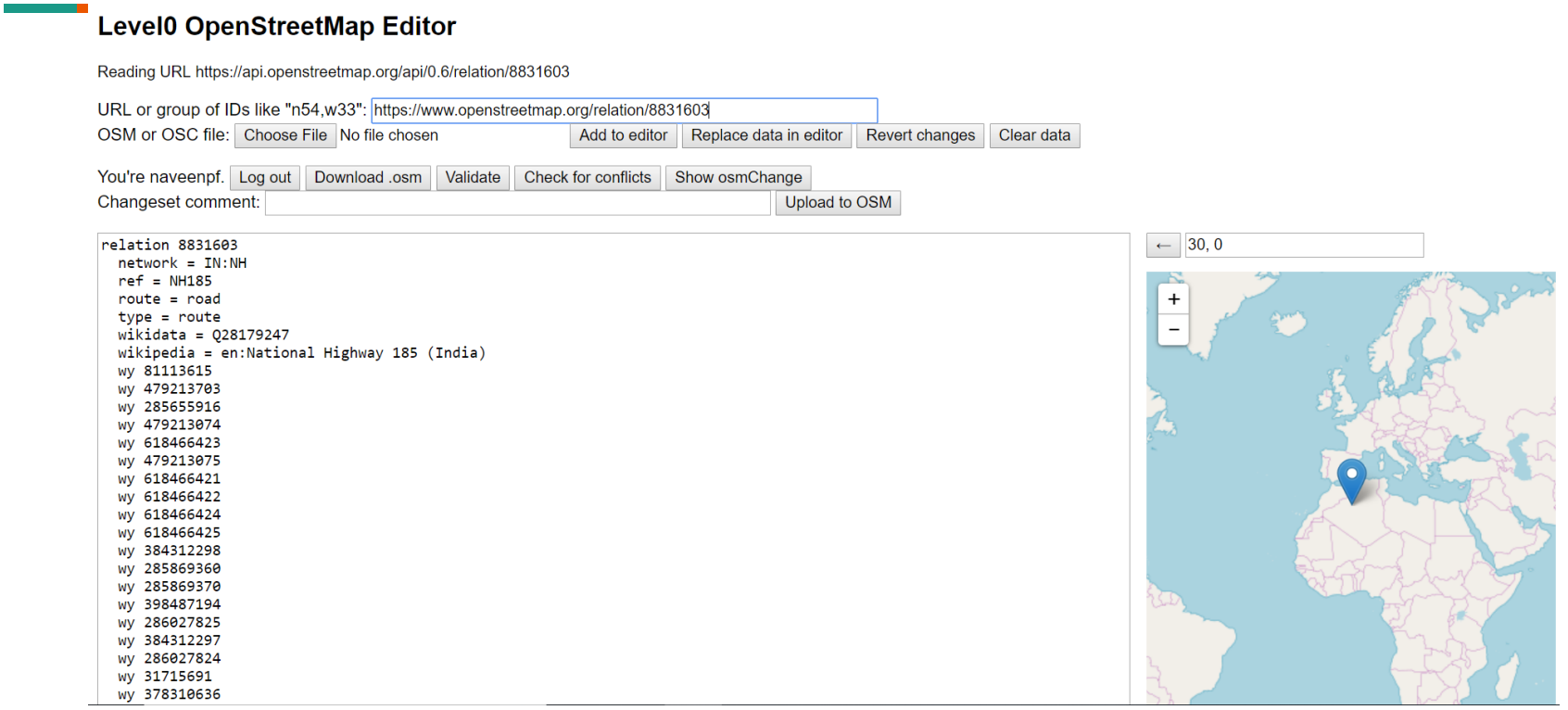
CHAPTER 3

How to add Wikidata to OSM

ID Editor


Edit feature
Fountain
All fields
Name
Josephine Shaw Lowell Fountain
Add field:
Address, Description, El...
All tags (2)
amenity
fountain
name
Josephine Shaw L...
All relations (0)
View on openstreetmap.org
Point
Line
Area
Save
Le Pain Quotidien
IRT Flushing Line
Josephine Shaw Lowell Fountain
Jose Bonifacio de Andrada e Silva
Southwest Porch
Crystal Palace
Bryant Park
bing
50 ft
live
bhousel
logout

Level0

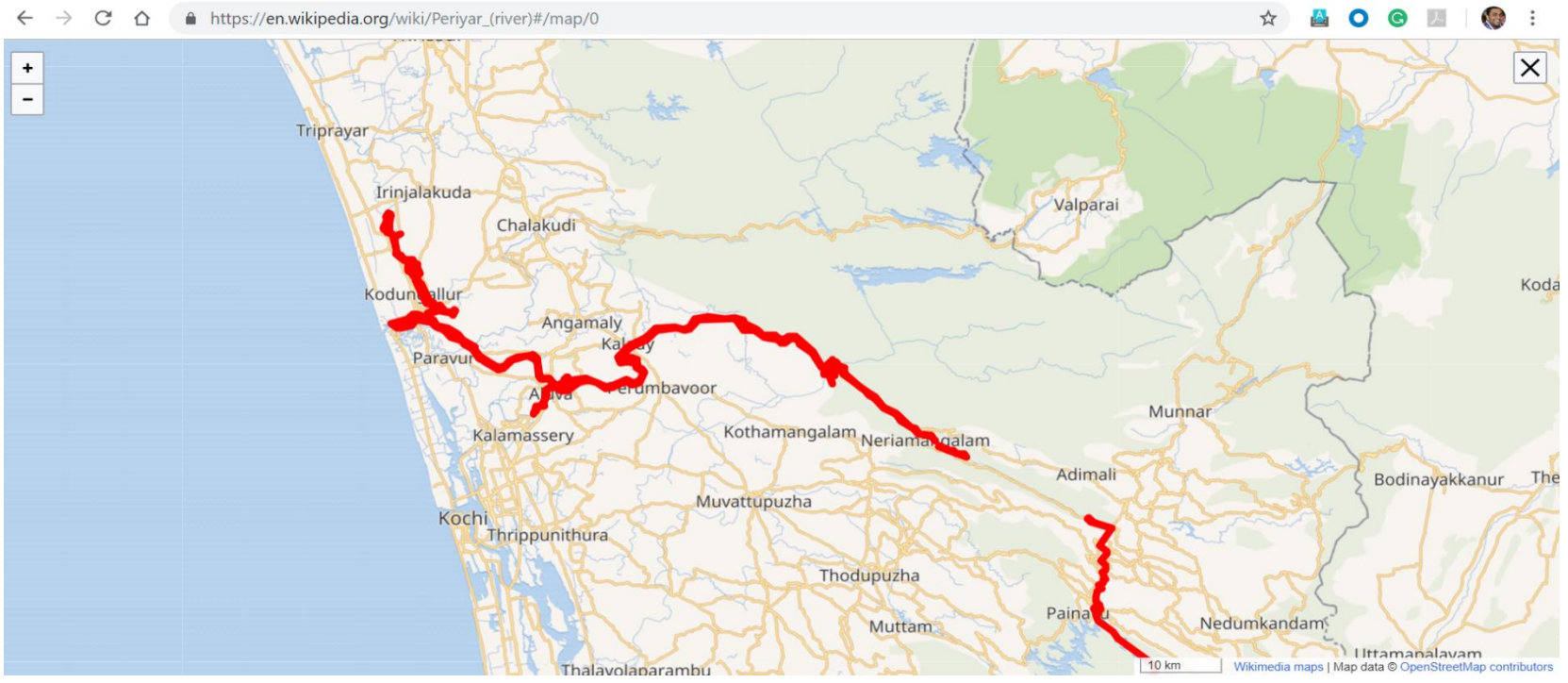


http://level0.osmz.ru/index.php

Linking OpenStreetMap and Wikidata

- A semi-automated, user-assisted editing tool
- https://osm.wikidata.link/

Challenges: River is line or shape ?



Challenges: Changing of tags

- Changing of tags in Wikidata and OSM causes issues.
- Very difficult to find what tags has been changed.

Challenges

Wikidata cannot wholesale import coordinates from OSM because OSM is licensed under the Open Database License (ODbL).

(We have a messy situation regarding EU database rights and the repeated extraction of insubstantial extracts.)

Challenges

OSM will *not* import coordinates from Wikidata despite the CC0 license because of provenance issues related to Google Maps and others.

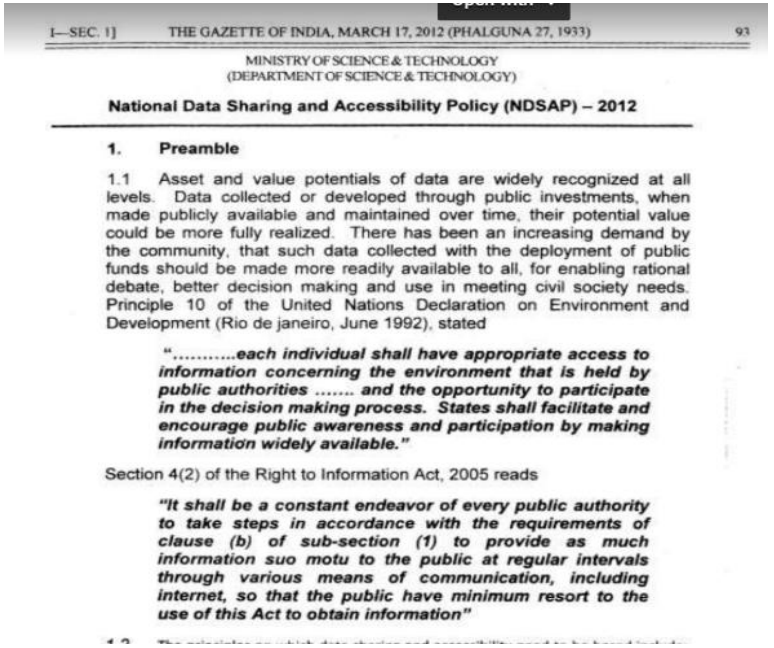
("It is an established principle in OpenStreetMap that we don't import geodata from Wikipedia.")

Challenges

There is some push back from the OSM community against bulk adding more wikidata=* tags into OSM. The comments are eerily similar to Wikipedians'.

(Example: "Could we just pause any wikidata edits for a month or two?" —OSM talk mailing list thread)

kerala.data.gov.in
Open Government Data Portal Kerala



I—SEC. 1] THE GAZETTE OF INDIA, MARCH 17, 2012 (PHALGUNA 27, 1933) 93

MINISTRY OF SCIENCE & TECHNOLOGY
(DEPARTMENT OF SCIENCE & TECHNOLOGY)

National Data Sharing and Accessibility Policy (NDSAP) – 2012

1. Preamble

1.1 Asset and value potentials of data are widely recognized at all levels. Data collected or developed through public investments, when made publicly available and maintained over time, their potential value could be more fully realized. There has been an increasing demand by the community, that such data collected with the deployment of public funds should be made more readily available to all, for enabling rational debate, better decision making and use in meeting civil society needs. Principle 10 of the United Nations Declaration on Environment and Development (Rio de janeiro, June 1992), stated

> ***".....each individual shall have appropriate access to information concerning the environment that is held by public authorities ..... and the opportunity to participate in the decision making process. States shall facilitate and encourage public awareness and participation by making information widely available."***

Section 4(2) of the Right to Information Act, 2005 reads

> ***"It shall be a constant endeavor of every public authority to take steps in accordance with the requirements of clause (b) of sub-section (1) to provide as much information suo motu to the public at regular intervals through various means of communication, including internet, so that the public have minimum resort to the use of this Act to obtain information"***

Topics of the most popular open datasets

1		**Police and crime:** Police incidents; jail bookings; police station locations; crime statistics
2		**Transportation:** Taxi licenses; transit data; traffic counts; road infrastructure data; parking data
3		**Emergency calls:** Police, fire, and EMS responses; 911 calls; response times; incident reports
4		**Development:** Commercial developments; housing developments; property data; housing affordability
5		**Building safety:** Building permits; safety permits; certificates of occupancy
6		**Finance:** Revenue; spending; employee salaries; capital budgets; payments
7		**Elections:** Election results; polling locations; campaign finance reports
8		**Businesses and licenses:** Business licenses; liquor licenses; vendor, contract, and procurement data
9		**Inspections and service requests:** Restaurant health inspections; 311 requests; code violations
10		**Education:** Schools information; student health data; after-school programs; library locations

https://kerala.data.gov.in/

kerala.data.gov.in
Open Government Data Portal Kerala

- Admin Boundaries:- Districts,Taluk, Village
- Political Boundaries:- Lokshaba Constituency, Assembly Constituency, Local Government Boundaries.

- LDF Keralam (*ldfkeralam.org* website) images to CC release
- Sarvavijnanakosham
- Schoolwiki
- Panchayatwiki
- Wikipedia article CD (KITE – IT @ School)
- Public release of Wikisource article CD
- Wikisource – Digitalisation effort with IT @ School

WIKIMEDIANS OF KERALA USER GROUP

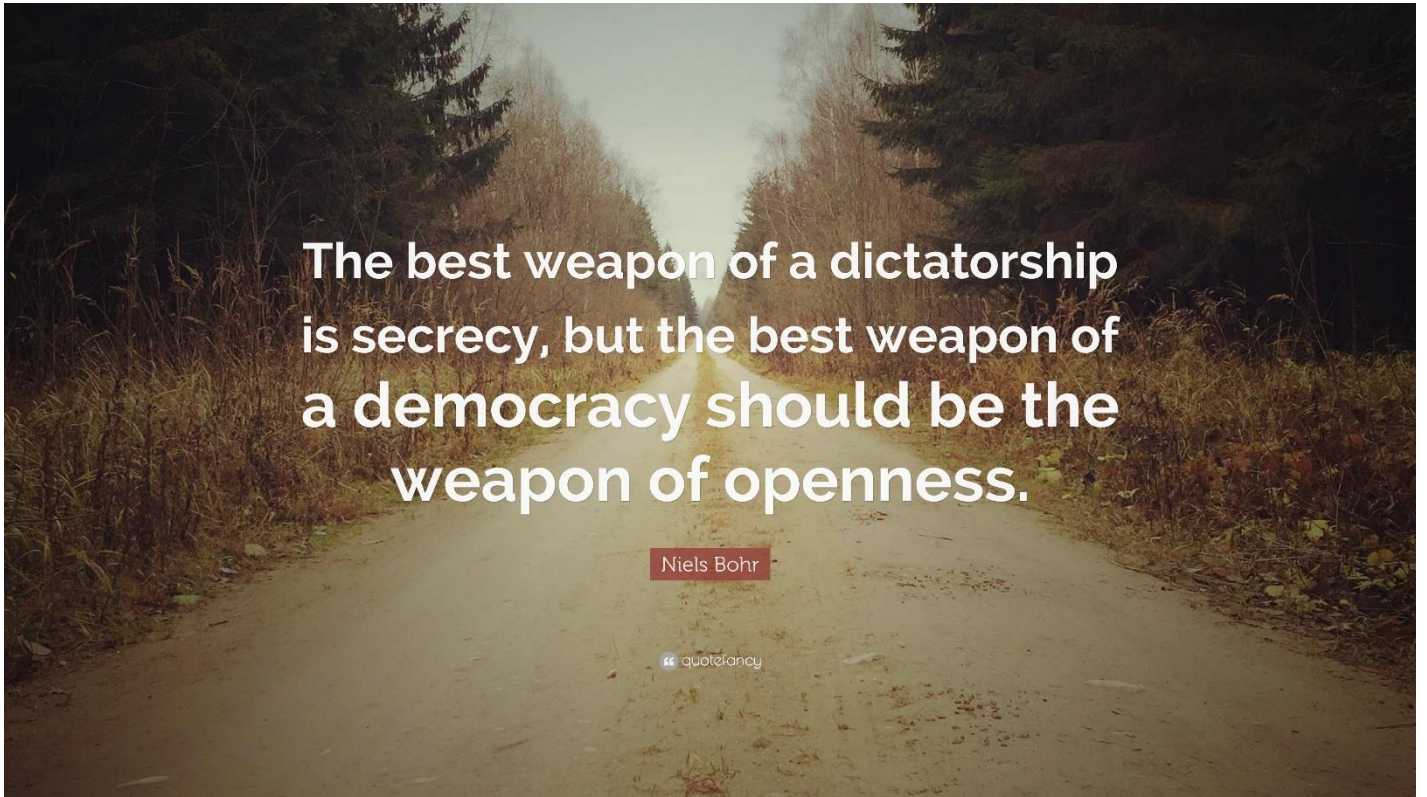
- Activities in partnering with SMC
- Partnering with HOTOSM – Remote Mapping during 2018 Floods
- Govt opendata licence vetting by LWG
- Partnering with Facebook –Mapping with AI during 2018 Floods
    - Helping in Hospital data import – Facebook/RMSI
- IT Dept contributing to OpenStreetMap – Mapathon Kerala
- Intervention to start - Kerala Opendata Portal

- Data used
    - Kerala Rescue
    - Covid19Kerala
    - CoronaSafe Network

SABHA TV
നാം ഒറ്റയ്ക്ക് നേടുന്നത് എത്ര കുറച്ചാണ്

SABHA TV
ഒരുമിച്ചു നമുക്ക് നേടാവുന്നത് എത്ര ഏറെയാണ്
TV

#opendata is the future
#opendata will win



Thank you

Thank you
Slides adapted :Eugene Alvin Villar
Inputs from : Manoj Karingamadathil