

# OpenStreetMap ♥ Wikidata

Edward Betts  
Eugene Alvin Villar



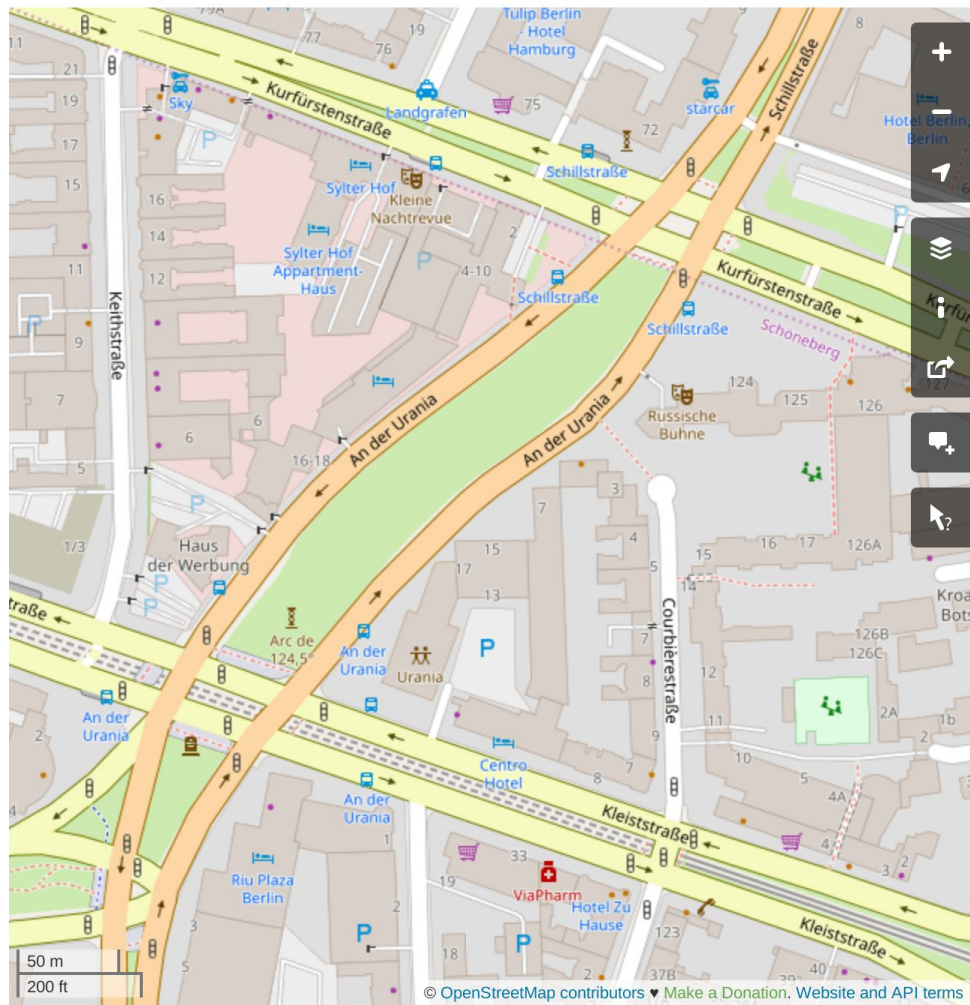
# OpenStreetMap and its data model

# OSM?



**OpenStreetMap** is a crowdsourced project to map the whole world!

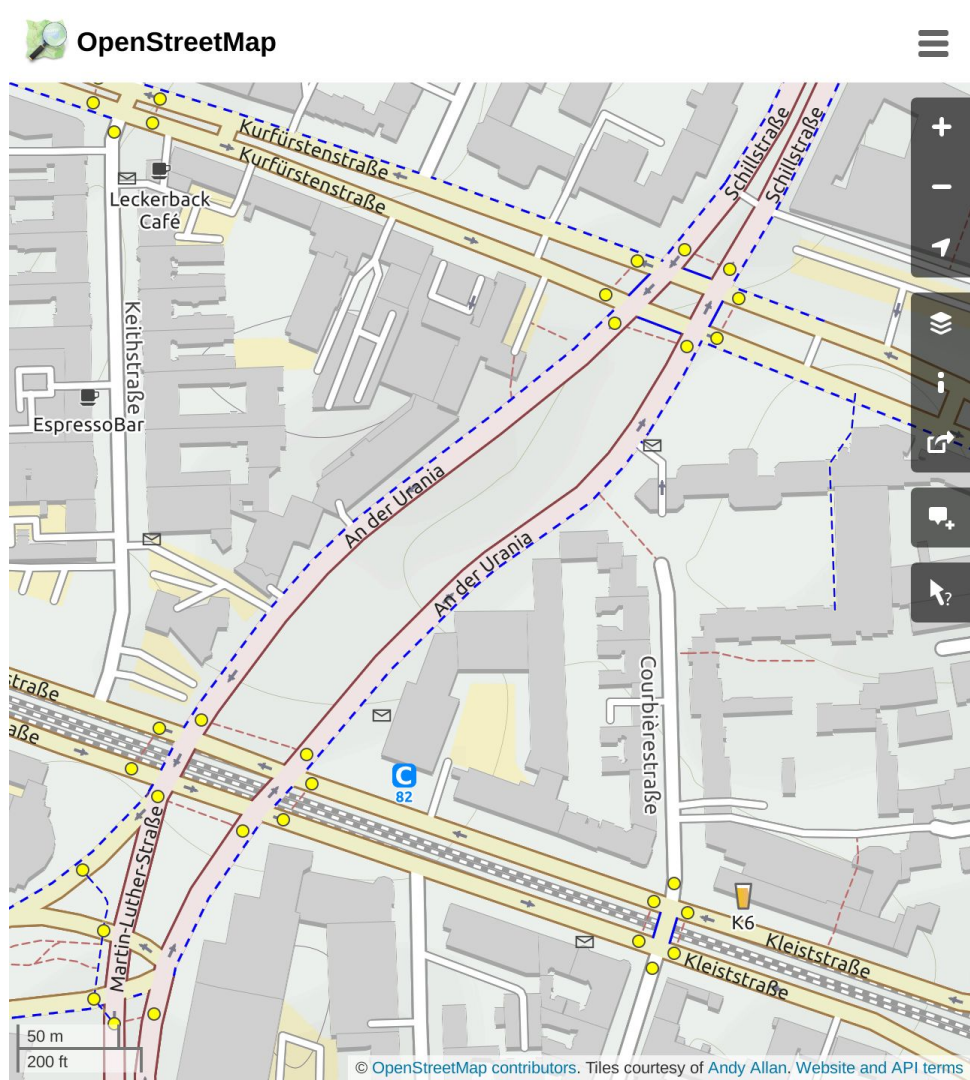
OpenStreetMap



# OSM?



Usual way to introduce OSM:  
*“OpenStreetMap is like Wikipedia for maps”*



# OSM?

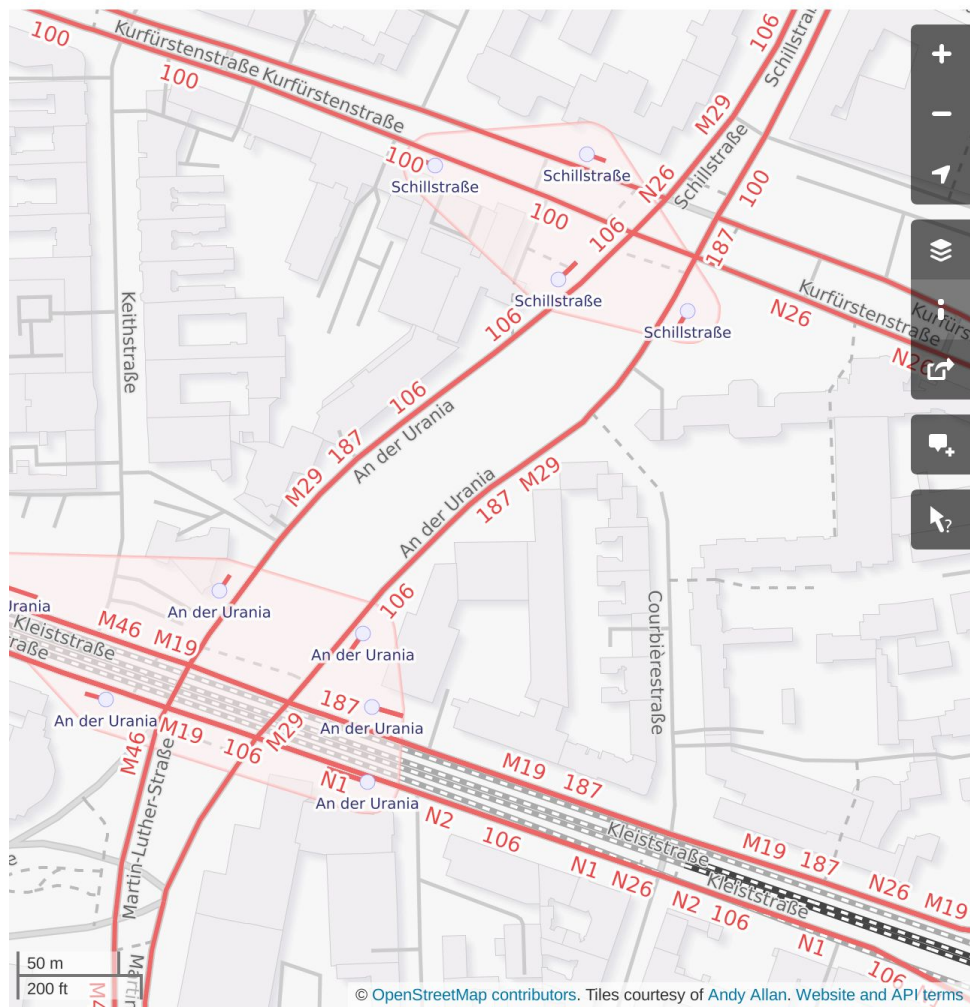


More accurate way  
to introduce OSM:

*“OpenStreetMap is  
like Wikidata for  
geographical data”*

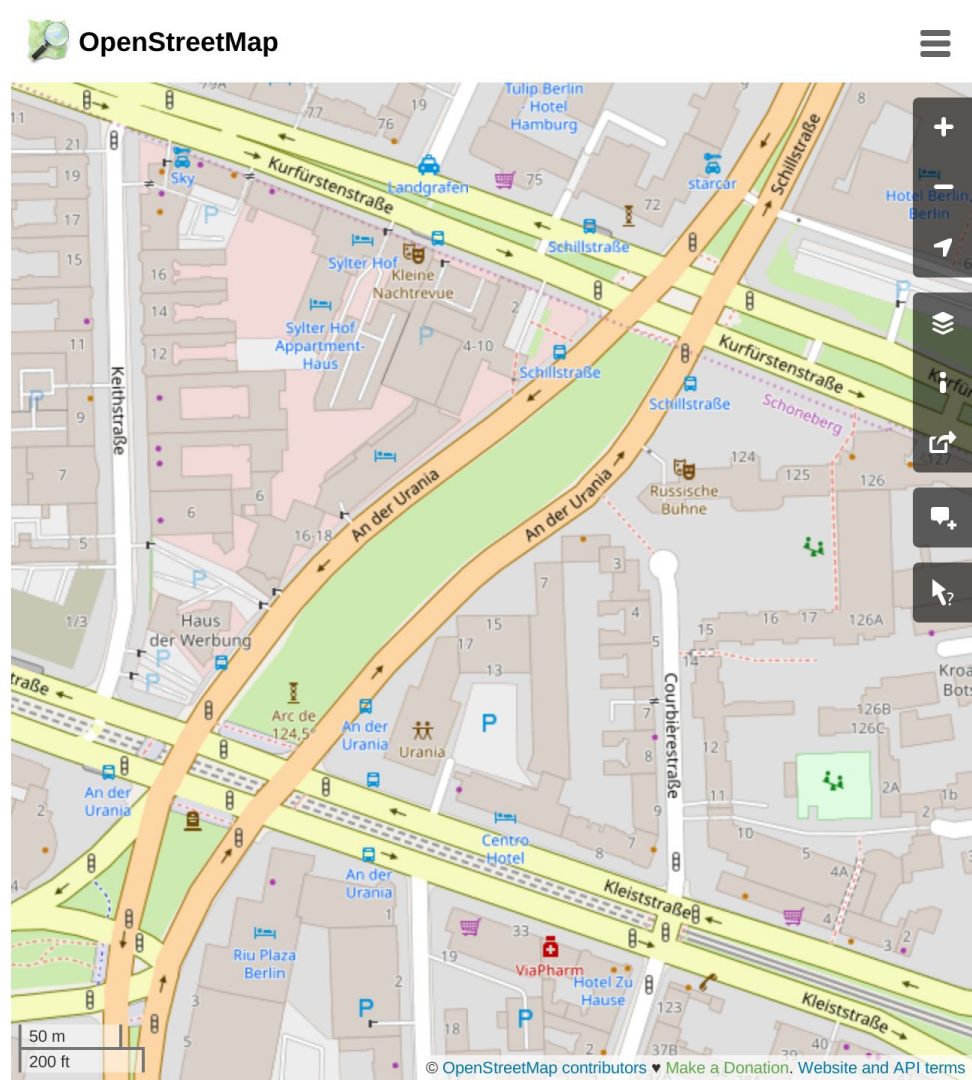
*(But only if you are already  
familiar with Wikidata!)*

 OpenStreetMap



# Data model

OSM and Wikidata have quite a lot in common, both being crowdsourced, open data projects.



# Data model

While Wikidata has items, statements, properties, values, qualifiers, references, labels, descriptions, etc., ...

The image shows a Wikidata item page for Marie Curie (Q7186) with several annotations explaining its data model components:

- label**: Points to the name "Marie Curie" and the identifier "Q7186".
- description**: Points to the text "Polish-French physicist and chemist" and the aliases "Maria Skłodowska-Curie | Maria Salomea Skłodowska | Madame Curie".
- property**: Points to the "award received" property.
- value**: Points to the value "Nobel Price in Physics".
- qualifiers**: Points to the table of qualifiers for the Nobel Prize, including "point in time" (1903), "together with" (Henry Becquerel, Pierre Curie), and "prize money" (35,339 Swedish krona).
- rank**: Points to the "rank" column in the table.
- statement group**: Points to the entire "Statements" section.
- opened references**: Points to the expanded reference for the Nobel Prize, showing details like "reference URL", "retrieved", "publisher", "language of work or name", and "title".
- collapsed reference**: Points to the collapsed reference for the "Willard Gibbs Award" (1921), which shows "1 reference".

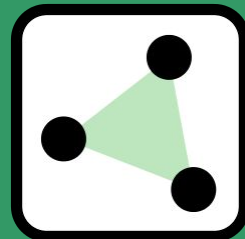
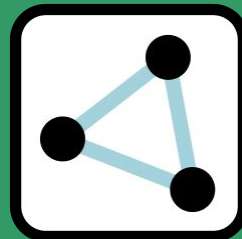
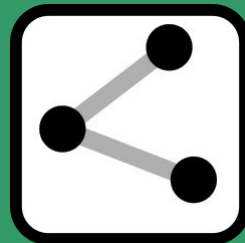
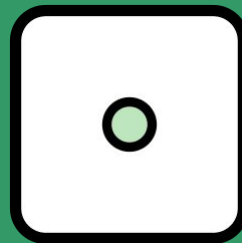
award received	Nobel Price in Physics
point in time	1903
together with	Henry Becquerel Pierre Curie
prize money	35,339 Swedish krona

rank	opened references												
	<p>▼ 2 references</p> <table border="1"><tbody><tr><td>reference URL</td><td>http://www.nobelprize.org/nobel_prizes/physics/laureates/1903/index.html</td></tr><tr><td>retrieved</td><td>3 August 2015</td></tr><tr><td>publisher</td><td>Nobel Foundation</td></tr><tr><td>language of work or name</td><td>English</td></tr><tr><td>title</td><td>The Nobel Prize in Physics 1903 (English)</td></tr><tr><td>reference URL</td><td>https://www.nobelprize.org/nobel_prizes/about/amounts/</td></tr></tbody></table> <p>+ add reference</p>	reference URL	http://www.nobelprize.org/nobel_prizes/physics/laureates/1903/index.html	retrieved	3 August 2015	publisher	Nobel Foundation	language of work or name	English	title	The Nobel Prize in Physics 1903 (English)	reference URL	https://www.nobelprize.org/nobel_prizes/about/amounts/
reference URL	http://www.nobelprize.org/nobel_prizes/physics/laureates/1903/index.html												
retrieved	3 August 2015												
publisher	Nobel Foundation												
language of work or name	English												
title	The Nobel Prize in Physics 1903 (English)												
reference URL	https://www.nobelprize.org/nobel_prizes/about/amounts/												

rank	collapsed reference		
	<p>Willard Gibbs Award</p> <table border="1"><tbody><tr><td>point in time</td><td>1921</td></tr></tbody></table> <p>► 1 reference</p> <p>+ add value</p>	point in time	1921
point in time	1921		

# Data model

... OSM has nodes, ways, relations (that have members and roles), tags, keys, values, etc.





# Data model

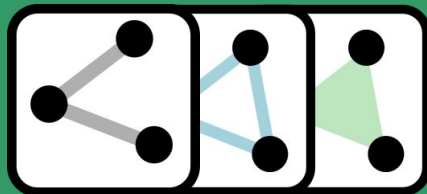
Nodes, ways, and relations model the geometry and topology of objects.

Tags (key-value strings) describe the actual “things” those objects represent.

nodes



ways



relations



---

tags



# Data model

OSM tags are quite similar to Wikidata statements:

**OSM key**  $\approx$

**Wikidata property**

*OSM's tags are not strictly controlled—users can invent and add any tag they like! There is a tagging proposal process for common tags, though.*



## Berlin item (Q64):

**population (P1082)** = 3611222

**point in time (P585)** = 2017-11-30

**reference URL (P854)** =

<https://www.statistik-berlin-brandenburg.de/>



## Berlin relation (62422):

**population** = 3440441



**source:population** =

[http://www.statistik-berlin-brandenburg.de/Publikationen/Stat\\_Berichte/2010/SB\\_A1-1\\_A2-4\\_q01-10\\_BE.pdf](http://www.statistik-berlin-brandenburg.de/Publikationen/Stat_Berichte/2010/SB_A1-1_A2-4_q01-10_BE.pdf) 2010-10-01

# Data model

Data modeling discussions on the [Wikidata:Project chat](#) page are actually quite similar to discussions on OSM's [tagging](#) mailing list.

*(Folksonomic ontology is hard!)*

## ^ Building that has changed uses



Are there any good models of how to deal with a building that has changed uses over time? This came up with [Mercy Friary \(Q47766776\)/Museum of Fine Arts of Seville \(Q2163496\)](#). I'm not sure I found a satisfactory solution -- it seems here like what I've done associates the physical building with [Mercy Friary \(Q47766776\)](#) and there is no item for the religious community that happened to reside in the building (whereas the museum that happens to reside in the building gets its own item). Is there a better way to do this? [Calliopejen1 \(talk\)](#) 05:33, 10 October 201

- Looking for a case I hoped would be handled correctly, I discovered [Post Office \(Q5532118\)](#) and [Hotel Monaco \(Q5911750\)](#), despite their buildings have no linkage at all beyond both being given as



## [Tagging] Guard booth building type

**Mateusz Konieczny** [matkoniecz at tutanota.com](mailto:matkoniecz@tutanota.com)  
Mon Oct 7 07:02:03 UTC 2019

- Previous message (by thread): [\[Tagging\] Feature Proposal - Voting - traffic calming=dynamic bump](#)
- Next message (by thread): [\[Tagging\] Guard booth building type](#)
- **Messages sorted by:** [\[date\]](#) [\[thread\]](#) [\[subject\]](#) [\[author\]](#)

What would be the best tagging for guard booth minibuildings?

I think that `building=guard_booth` would be the best one for objects constructed as guard booth.

Such objects are housing guard, usually at entrance to parking lot, industrial, closed private residential complex, military base etc.

Is there better name than "guard booth"? "guard shed" was a previous version that I had

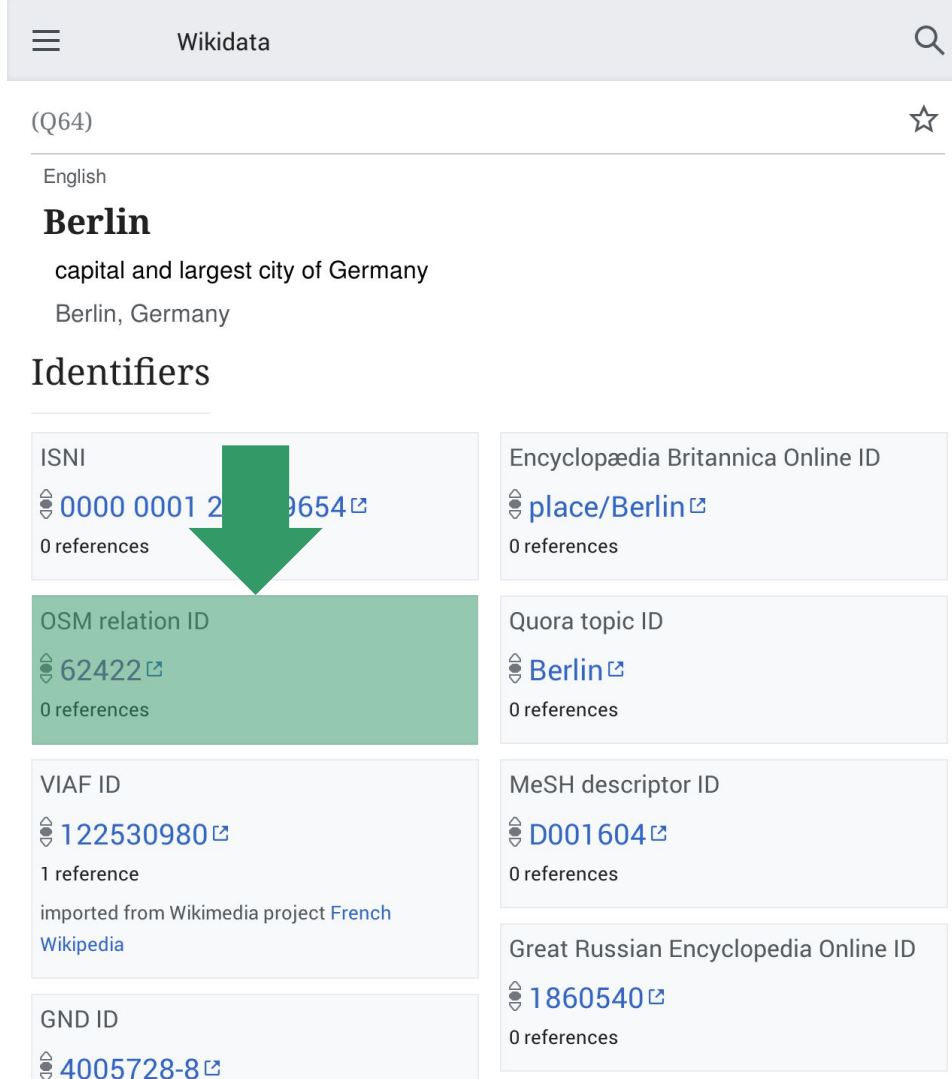




# WD ⇒ OSM

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

*(Why only relations?  
Because OSM IDs are not stable, but relations have the most stable IDs.)*



Wikidata

(Q64)

English

**Berlin**

capital and largest city of Germany

Berlin, Germany

Identifiers

ISNI	0000 0001 2...9654	0 references
Encyclopædia Britannica Online ID	place/Berlin	0 references
OSM relation ID	62422	0 references
Quora topic ID	Berlin	0 references
VIAF ID	122530980	1 reference imported from Wikimedia project <a href="#">French Wikipedia</a>
MeSH descriptor ID	D001604	0 references
GND ID	4005728-8	0 references
Great Russian Encyclopedia Online ID	1860540	0 references

# WD ⇒ OSM

*Ontology:* Wikidata items/properties for geographical features can link to “equivalent” OSM classes using the **OSM tag or key** (P1282) property.

0 references

image



Leuchtturm in Westerheversand  
crop.jpg

3,606 × 2,894; 8.36 MB

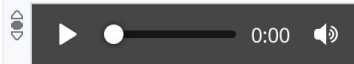
0 references

studied by

 [pharology](#)

0 references

pronunciation audio



[LL-Q188 \(deu\)-Sebastian Wallroth-Leuchtturm.wav](#)

1.4 s; 131 KB

language of work or name [German](#)

0 references

0 references

 [Brockhaus and Efron](#)

[Encyclopedic Dictionary](#)

statement is subject of [Q24467653](#)

0 references

 [Small Brockhaus and Efron](#)

[Encyclopedic Dictionary](#)

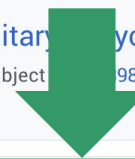
statement is subject of [Q24772093](#)

0 references

 [Sytin Military Encyclopedia](#)

statement is subject of [Q198057](#)

0 references




OSM tag or key

 [Tag:man\\_made=lighthouse](#)

0 references

equivalent class

 <http://dbpedia.org/ontology/Lighthouse>

described at URL <http://mappings.dbpedia.org/index.php/OntologyClass:Lighthouse>

retrieved 16 June 2015

1 reference

imported from Wikimedia project [DBpedia](#)

has part

# OSM ⇒ WM

OSM objects can link to corresponding Wikipedia articles and Wikidata items using the [wikipedia=\\*](#) and [wikidata=\\*](#) tags respectively.


Relation: Berlin (62422)

"capital=yes" is the characteristic of Berlin "place=city" node

Edited about 2 months ago by [azakh-world](#)

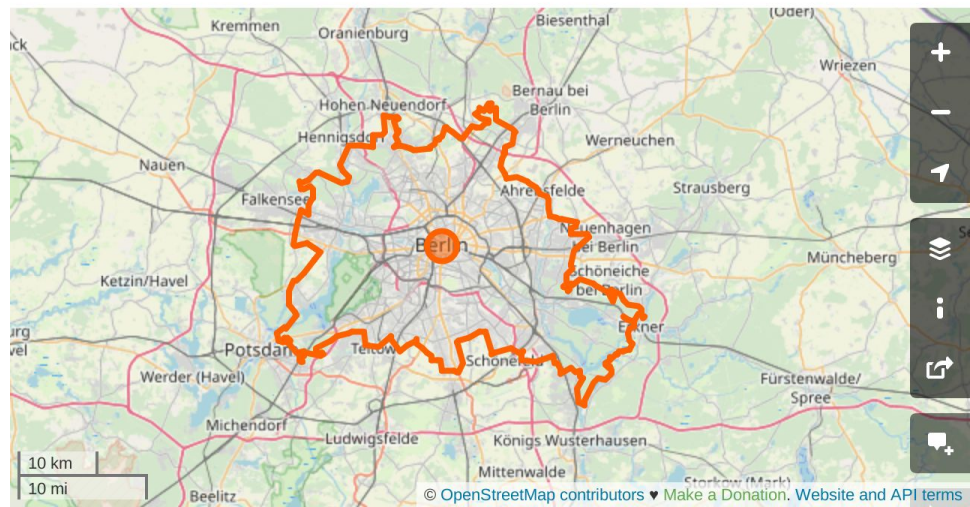
Version #265 · Changeset #73620162

Tags



wikidata	Q64
wikipedia	de:Berlin

Part of



# OSM ⇒ WM

There are also several Wikidata secondary tags such as:

**brand:wikidata=\***

**architect:wikidata=\***

**artist:wikidata=\***

**name:etymology:**

**wikidata=\***

## OpenStreetMap

Location: [52.5313942](#), [13.3429669](#)

Tags



<a href="#">artist:wikidata</a>	<a href="#">Q28406050</a>
<a href="#">artist_name</a>	Herakut
<a href="#">artwork_type</a>	mural
<a href="#">inscription</a>	As Long As You Are Standing, Give A Hand To Those Who Have Fallen.
<a href="#">tourism</a>	<a href="#">artwork</a>
<a href="#">wheelchair</a>	yes

## OpenStreetMap

**Node: Europcar (1719500135)**

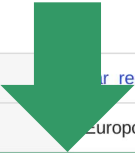
Modified via [wheelmap.org](#)

Edited [about 2 months ago](#) by [wheelmap\\_visitor](#)

Version #6 · Changeset [#73869547](#)

Location: [52.5045062](#), [13.3424821](#)

Tags



<a href="#">amenity</a>	<a href="#">car rental</a>
<a href="#">brand</a>	Europcar
<a href="#">brand:wikidata</a>	<a href="#">Q1376256</a>
<a href="#">brand:wikipedia</a>	en:Europcar



# OSM ⇒ WM

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



- Main Page
- The map
- Map Features
- Contributors
- Help
- Blogs
- Shop
- Donations
- Recent changes

- Tools
- What links here
- Related changes
- Special pages
- Printable version
- Permanent link
- Page information
- OpenStreetMap Wiki item
- Cite this page

## man\_made=lighthouse

The **man\_made=Lighthouse** tag is used to identify a **lighthouse** or former lighthouse. Although the tower structure is very common, a lighthouse building can be any shape. The common factor is that these buildings have a lamp room at the top. Many lighthouses are no longer operational, so there may not be any form of light in the lamp room.

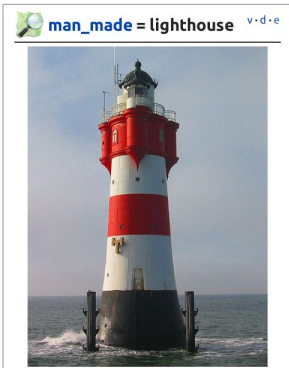
Operational lighthouses show different colors in different sectors (e.g. white, red, green), so the relative direction of the light to the viewer can be approximated. Some lights are positioned behind each other and, when brought to an alignment act as a leading light. Often, lights incorporate radio transmitters, DGPS, radar and other things too.

Note that there are some small navigation aids that emit light that are not lighthouses but beacons (**man\_made=beacon**).

Contents [hide]	
1	How to map
1.1	Building properties
1.2	Navigational light properties
1.3	Re-purposed and multi-purpose lighthouse buildings
2	Rendering
3	Example
4	See also
5	External links

### How to map

Full mapping may consist of one or two OSM objects. The **lighthouse** object describes the characteristics of the tower as a building; the **light** object describes the



**Description**  
Tower that emits light to serve as a navigational aid at sea or on inland waterway

**Rendering in openstreetmap-carto**

**Group:** Man made

**Used on these elements**

A set of icons showing how the tag is rendered in the OpenStreetMap cartographic style, including a lighthouse icon and a red and white striped tower icon.

**Combination**  
=\*

**man\_made=beacon**

**Wikidata**  
Q39715  
**Status:** de facto



# OSM maps and data in Wikimedia

# Interactive Maps

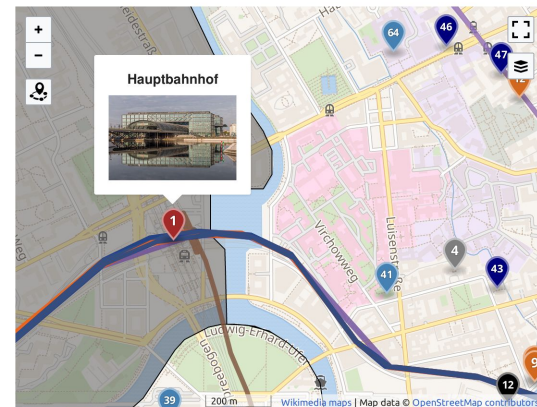
OSM data powers  
WMF's [Kartotherian](#)  
map tile service,  
which is used by the  
[Kartographer](#)  
MediaWiki extension.

... and by the northern part of the area to the west. The eastern part of the area takes its name from the "Barn Quarter" from the move in 1672 by the Great Elector of all the hay barns out of the fire-prone city center. In the late 19th century, the area became a refuge for Jews fleeing persecution and pogrom in Russia and Poland. By then it was the center of Jewish life in Berlin.

- **Potsdamer Platz** – the area around the completely razed Potsdamer Platz became no man's land between East and West Berlin and remained an empty strip of land until the 1990s, when it was rebuilt as a large project including striking highrises of concrete and steel, mixing offices and commercial space.
- **Spreebogen/Regierungsviertel** – "Spreebogen" means "the bow of the river Spree" and in Berlin generally refers to a particular one, where Spree meets the Berlin-Spandau Canal. The area around it houses the German federal government's institutions on the south (or left) bank, called *Regierungsviertel* ("government district"), while directly opposite it you will find Berlin's all-new central train station *Hauptbahnhof*.

^ Get in

Mitte regained its position as the main point of entry to Berlin in June 2006 with the opening of the new central station (**1** [Hauptbahnhof](#)), a giant palace of glass and steel, which is at the border of Mitte and Moabit. Almost all short- and long-haul trains arrive and depart from this station. Hauptbahnhof is also served by a Straßenbahn (tram) line and by the S-Bahn as well as U55, a rather pointless line of the U-Bahn (going only 1.8 km to Brandenburger Tor) until it is connected to the rest of U5 some time before 2020. Other main public transport stations are Friedrichstraße and Alexanderplatz.



Map of Berlin/Mitte

## By S- and U-Bahn

Mitte is served by many S- and U-Bahn lines. The **S1**, **S2** and **S25** go from north (Oranienburg and Gesundbrunnen) to south (Potsdamer Platz and Schöneberg), the Stadtbahn (city S-Bahn, line 5, 7, and 75) goes from west (Charlottenburg) to east (Friedrichshain). They cross at Friedrichstraße. U-Bahn line **U2** connects Mitte with Charlottenburg (west) and Prenzlauer Berg (northeast), the U-Bahn lines **U6** and **U8** go north to Wedding and south to Kreuzberg and Neukölln. The short U-Bahn line **U55** (see infobox) goes from Hauptbahnhof to Brandenburger Tor. An extension linking **U55** to the main **U5** making it a lot more useful is expected to open in 2020 (updated Apr 2019).

The most important stations are:



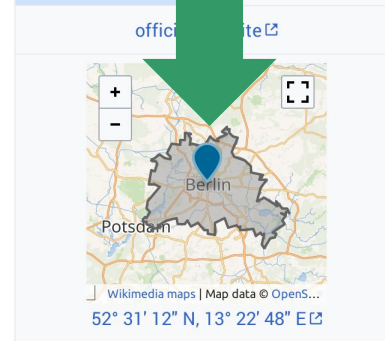
S-Bahn station Hackescher Markt

# Interactive Maps

Kartographer maps can also pull and overlay geometry from OSM.

- ?
  - ▶ [Unidentified locations in Berlin](#) (3 C, 765 F)
- A
  - ▶ [Architecture of Berlin](#) (27 C, 42 F)
- C
  - ▶ [Construction in Berlin](#) (9 C, 20 F)
  - ▶ [Culture of Berlin](#) (51 C, 66 F)
- E
  - ▶ [Economy of Berlin](#) (30 C, 37 F)
  - ▶ [Education in Berlin](#) (8 C, 16 F)
- G
  - ▶ [Geography of Berlin](#) (12 C, 6 F)
  - ▶ [Geology of Berlin](#) (12 C)
- H
  - ▶ [History of Berlin](#) (65 C, 1 P, 45 F)
- I
  - ▶ [Infrastructure in Berlin](#) (2 C)
- M
  - ▶ [Military in Berlin](#) (3 C, 3 F)
- N
  - ▶ [Nature of Berlin](#) (26 C, 138 F)
- O
  - ▶ [Objects of Berlin](#) (15 C, 16 F)
  - ▶ [Organizations of Berlin](#) (38 C, 16 F)
- P

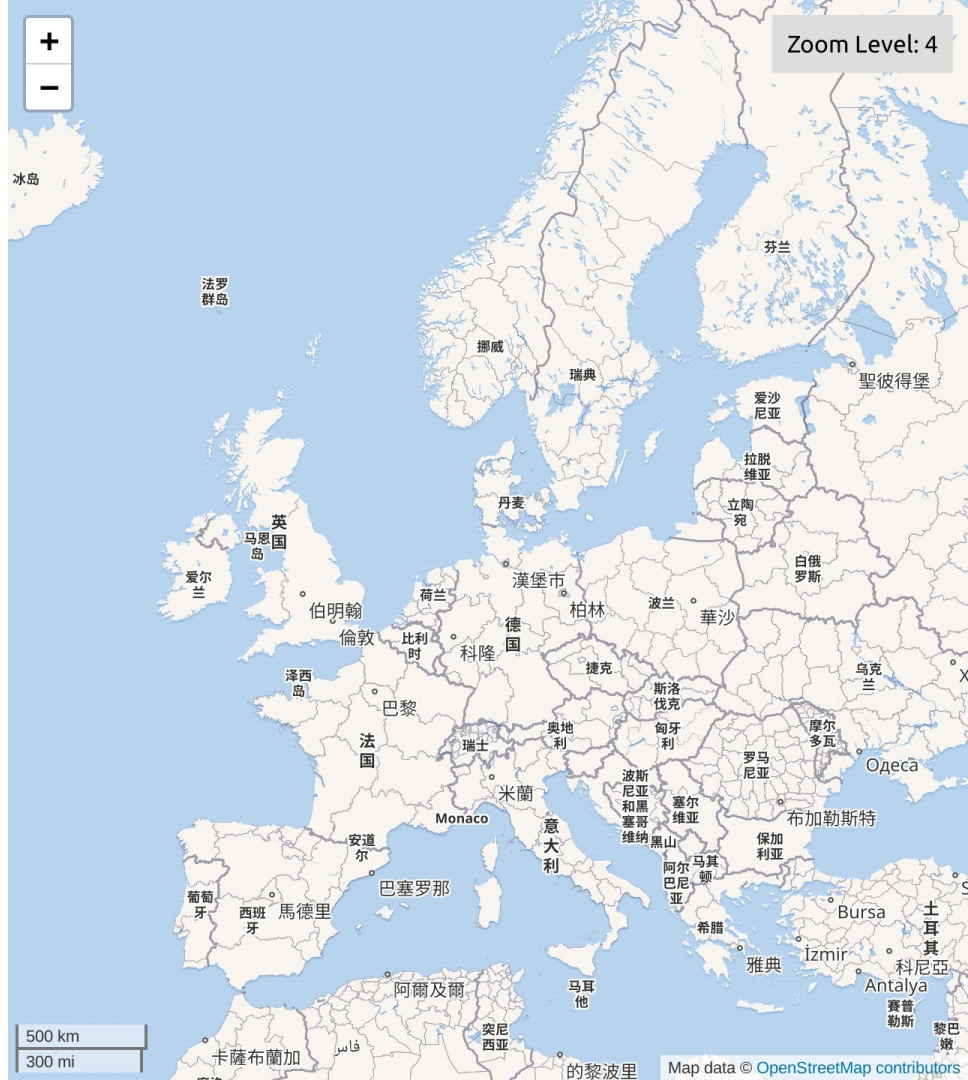
<b>Inception</b>	• 1237
<b>Highest point</b>	• <a href="#">Arkenberge</a>
<b>Population</b>	• 3,613,495 (2017)
<b>Area</b>	• 891.12 km <sup>2</sup> (2016, 2017)
<b>Elevation above sea level</b>	• 34 ±1 m
<b>Budget</b>	• 28,000,000,000 Euro (2018)



<b>Authority control</b>
<span>Q64</span>
ISNI: <a href="#">0000 0001 2341 9654</a>
VIAF ID: <a href="#">122530980</a>
GND ID: <a href="#">4005728-8</a>
Library of Congress authority ID: <a href="#">n79034972</a>
Bibliothèque nationale de France ID: <a href="#">15298132w</a>
National Diet Library Auth ID: <a href="#">00629194</a>

# Interactive Maps

In 2018, WMF [released](#) localised map tiles for Kartotherian, leveraging OSM's **name:\*=\*** tags.



# Wikidata use in OSM

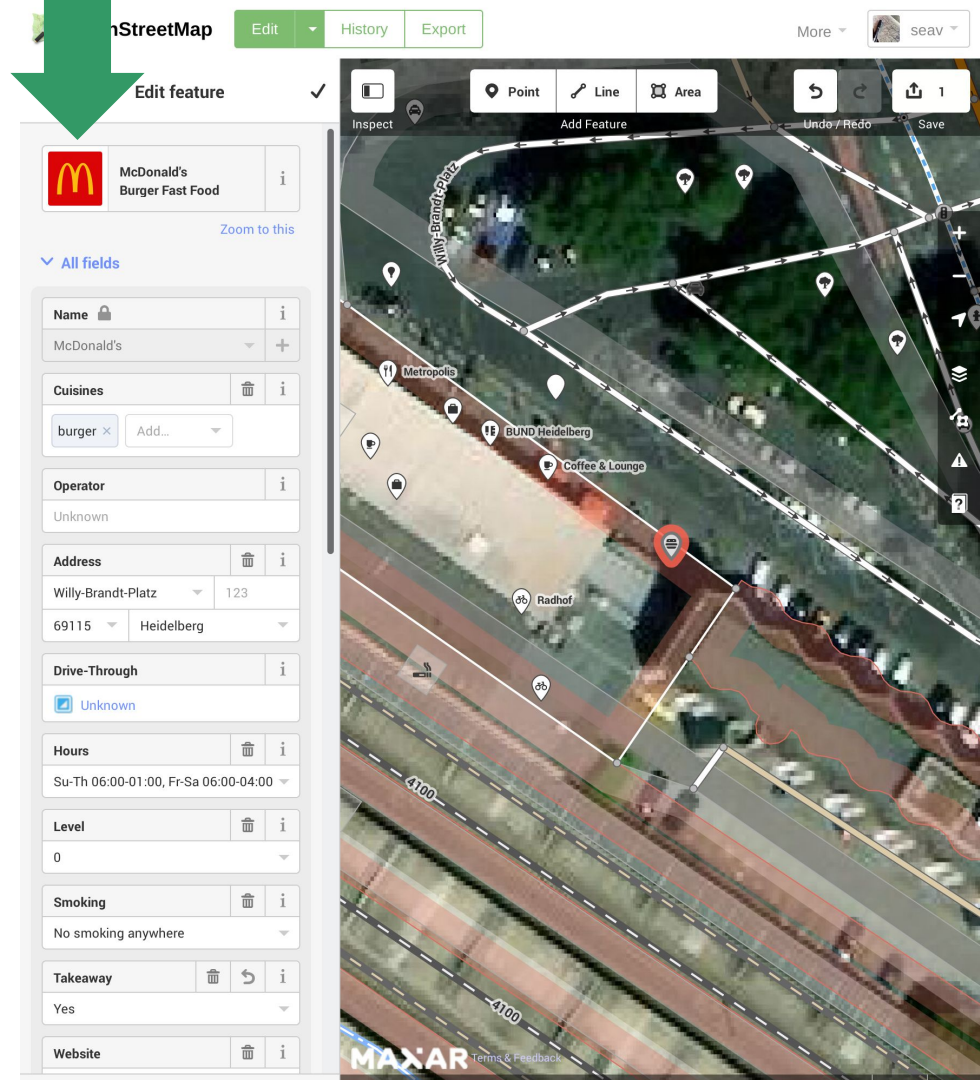


# Brands

OSM's **Name Suggestion Index**

uses Wikidata to provide brand identity and improved tagging.

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

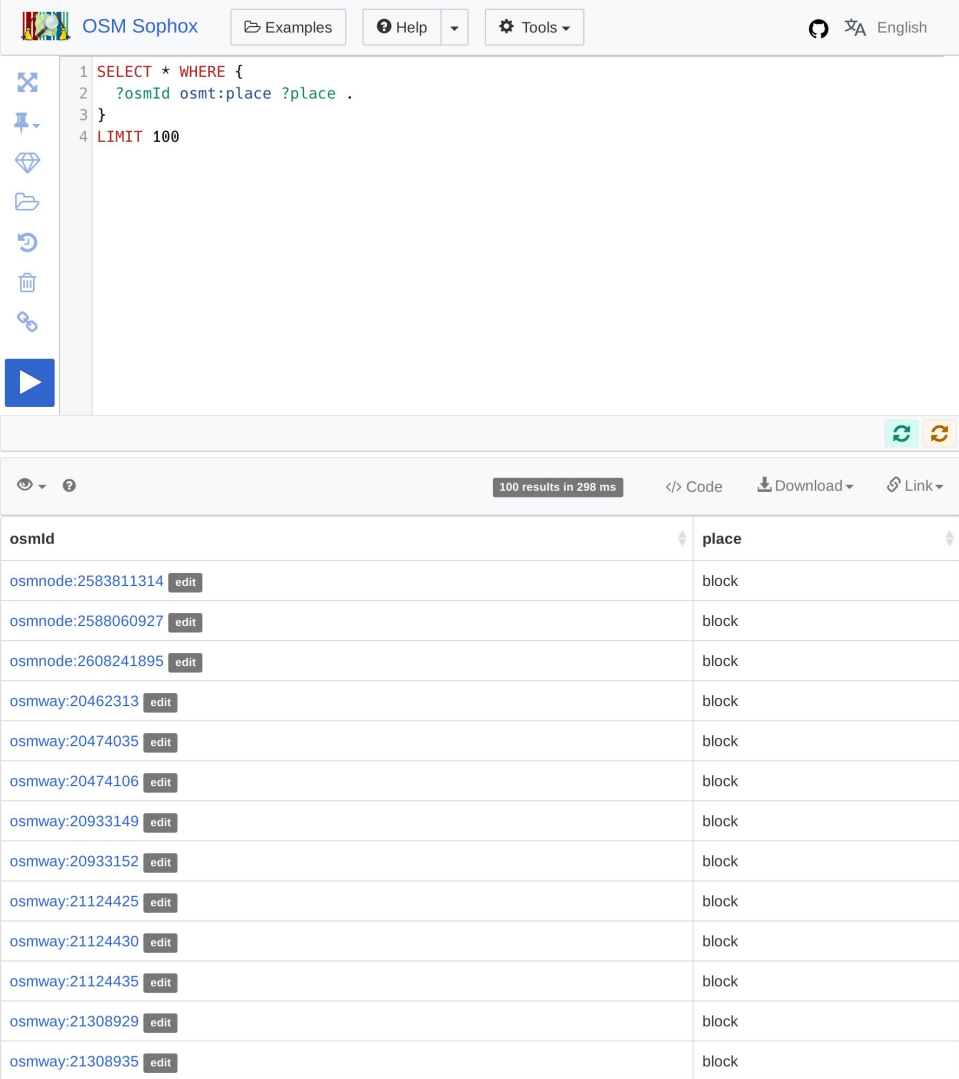


# Federation

**Sophox** is a SPARQL endpoint for OSM data. This service can use RDF federation to also query linked Wikidata items.

<https://wiki.osm.org/Sophox>

<https://sophox.org/>



The screenshot shows the OSM Sophox SPARQL query interface. The query is:

```
1 SELECT * WHERE {  
2   ?osmId osmt:place ?place .  
3 }  
4 LIMIT 100
```

The results table shows 100 results in 298 ms. The table has two columns: **osmId** and **place**. Each row contains an OSM ID and the word "block".

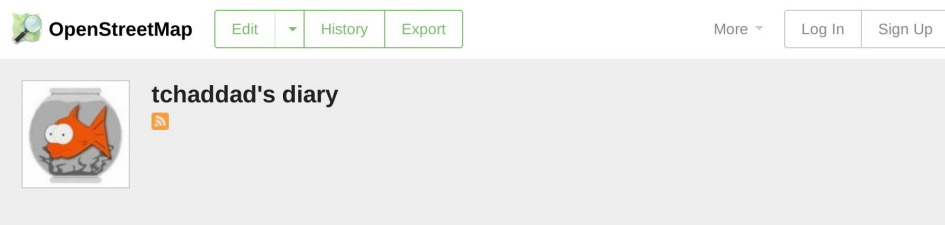
osmId	place
<a href="#">osmnode:2583811314</a> <small>edit</small>	block
<a href="#">osmnode:2588060927</a> <small>edit</small>	block
<a href="#">osmnode:2608241895</a> <small>edit</small>	block
<a href="#">osmway:20462313</a> <small>edit</small>	block
<a href="#">osmway:20474035</a> <small>edit</small>	block
<a href="#">osmway:20474106</a> <small>edit</small>	block
<a href="#">osmway:20933149</a> <small>edit</small>	block
<a href="#">osmway:20933152</a> <small>edit</small>	block
<a href="#">osmway:21124425</a> <small>edit</small>	block
<a href="#">osmway:21124430</a> <small>edit</small>	block
<a href="#">osmway:21124435</a> <small>edit</small>	block
<a href="#">osmway:21308929</a> <small>edit</small>	block
<a href="#">osmway:21308935</a> <small>edit</small>	block



# Geocoding

In 2019, OSM had a GSoC project to use **wikidata=\*** tags in **Nominatim**, OSM's default geocoder, for better search results.

*Note: Nominatim already uses **wikipedia=\*** tags.*



The screenshot shows the top of a diary entry on the OpenStreetMap website. At the top left is the OpenStreetMap logo. To its right are three buttons: 'Edit', 'History', and 'Export'. Further right is a 'More' dropdown menu, and at the far right are 'Log In' and 'Sign Up' buttons. Below this navigation bar is the header for the diary entry, which includes a profile picture of a fish in a bowl and the title 'tchaddad's diary' with a small orange icon.

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:



# Structured tag info

The OSM Wiki has the **Wikibase** extension installed. This is intended to provide better machine-readable data on OSM's tags.



OpenStreetMap Wiki

## highway=motorway (Q4980)

High capacity highways designed to safely carry fast motor traffic.

Contents ▾

### Statements

instance of <a href="#">tag</a> 0 references	use on relations <a href="#">is prohibited</a> 0 references
permanent tag ID <a href="#">highway=motorway</a> 0 references	group <a href="#">highways</a> 0 references
key for this tag <a href="#">highway</a> 0 references	Wikidata concept <a href="#">Q46622</a> 0 references
image  <a href="#">A4-passante di mestre dd.png</a> image caption Αυτοκινητόδρομος (Greek) 0 references	implies <a href="#">surface=paved</a> 0 references <a href="#">oneway=yes</a> 0 references <a href="#">motor_vehicle=yes</a> 0 references <a href="#">hgv=yes</a> 0 references <a href="#">motorcar=yes</a>
	

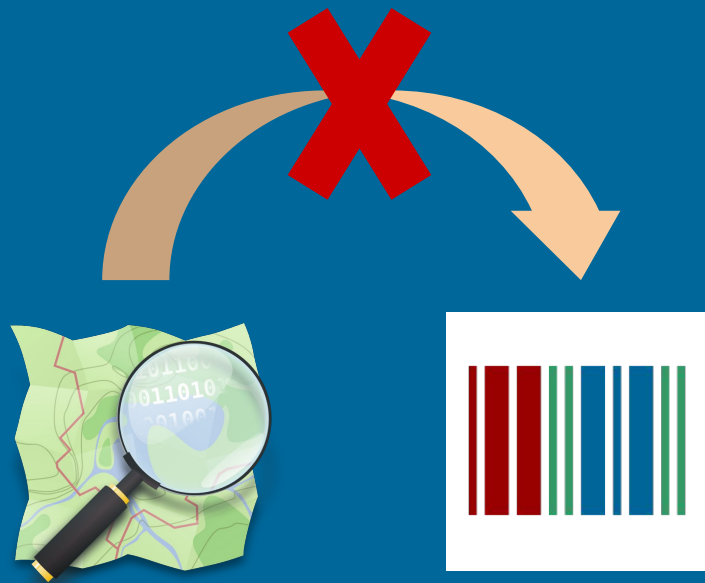
# Copyright and IP issues



# OSM ⇒ WD

Wikidata can't import coordinates from OSM, which is under the Open Database License (ODbL).

*(EU/UK have database rights, conflicting with the U.S. "facts are not copyrightable" doctrine.)*



# WD ⇒ OSM

OSM **will not** import geodata from Wikidata despite CC0 because of data provenance issues.

*(“It is an established principle in OSM that we don’t import geodata from Wikipedia.”)*



# Adding links from OpenStreetMap to Wikidata

# OSM/Wikidata linking tool

A semi-automated,  
user-assisted editing tool

<https://osm.wikidata.link/>



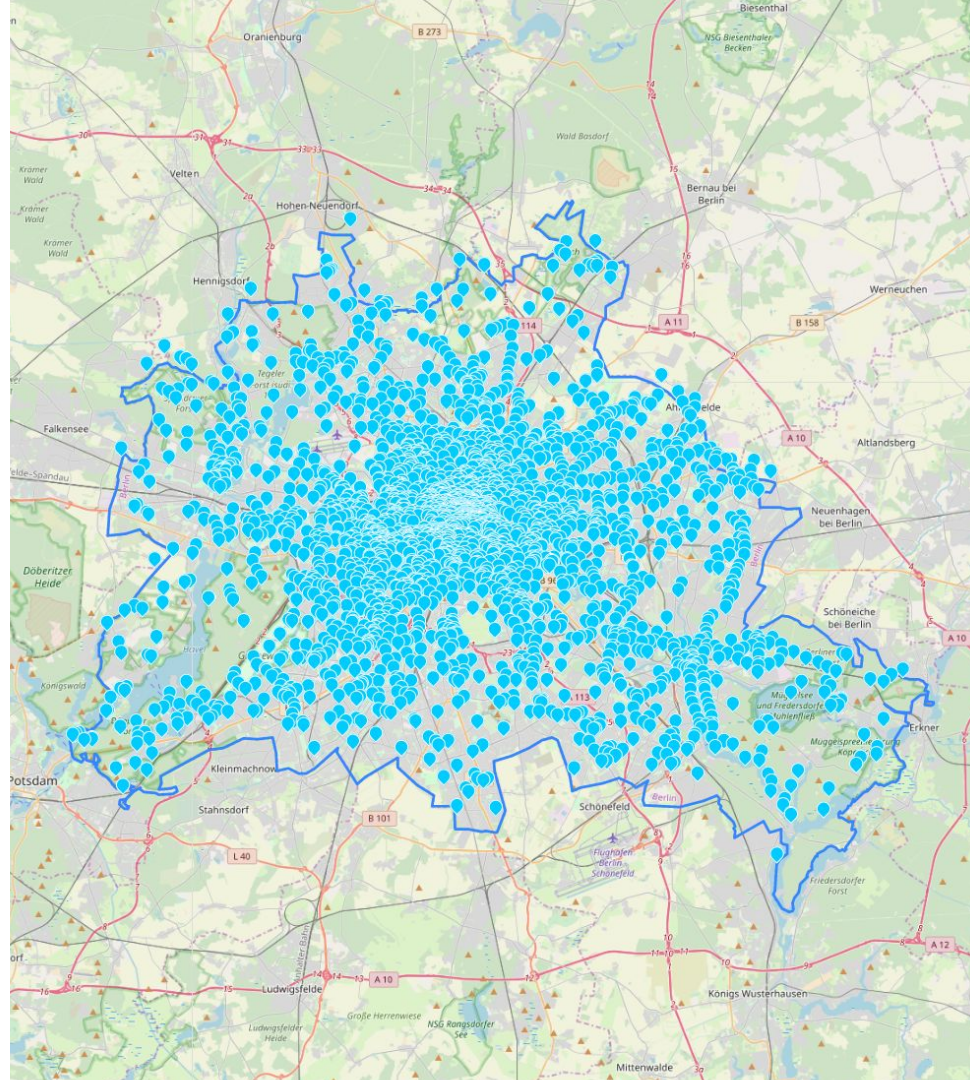
Photo by Edward Betts // CC-0



# Berlin

The software finds Wikidata items with a matching object in OSM.

There are 2,858 matches in Berlin.



# OSM 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

# Robot edits

Adding links individually is laborious.

OSM has strict rules about robot editing.

Candidate matches need to be checked for false positives.

Machine-assisted editing is a good compromise.



Photo by Edward Betts // CC-0

# 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

# Tram stops

Wikidata represents tram stops in both directions as a single item.

OSM represents tram stops as a two nodes, one for each direction, then both are grouped into a single relation.

These relations are unsupported in osm2pgsql, so the matcher fails to identify the relation.

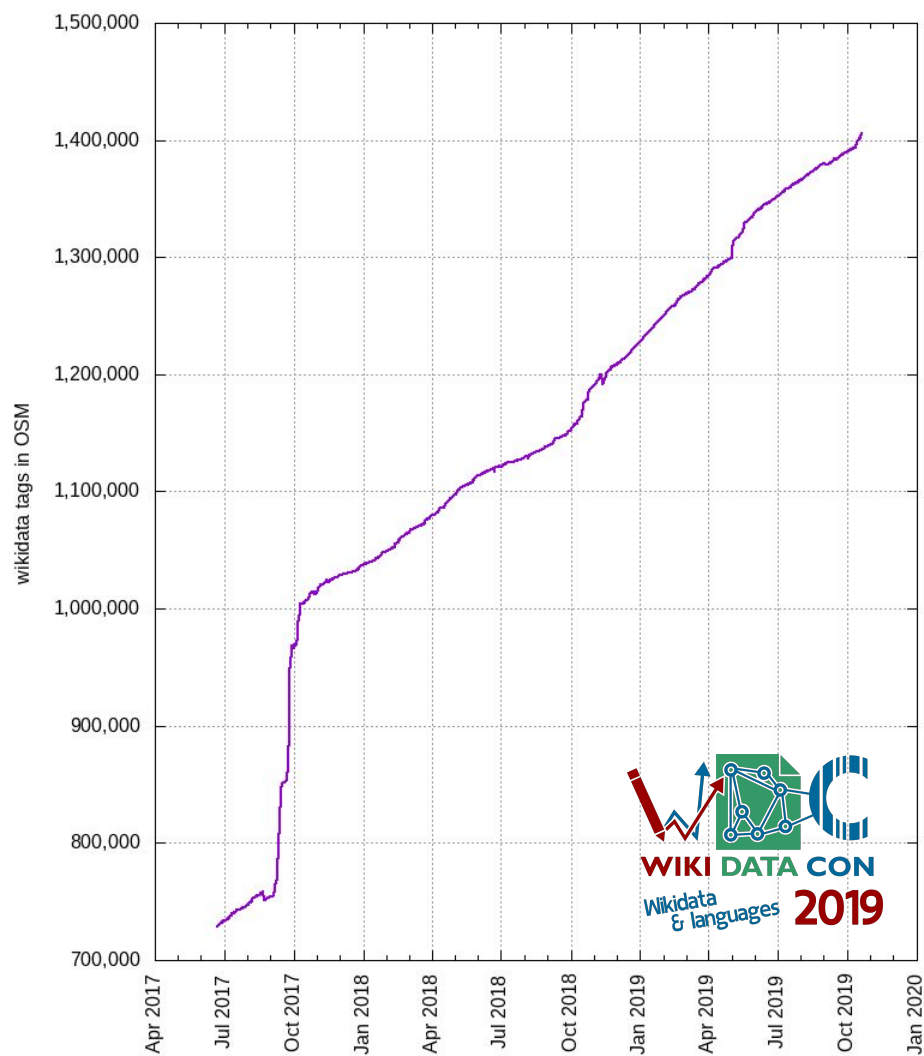


Photo by Edward Betts // CC BY-SA 3.0

# Mappers are using the tool

- Uploads by 225 mappers
- 9,600 changesets uploaded to OpenStreetMap
- 250,000 Wikidata links added

There are now  
over 1.4 million  
OSM objects  
with a  
wikidata=\* tag





# OpenStreetMap ♥ Wikidata

Edward Betts  
Eugene Alvin Villar