

# Linking OpenStreetMap and Wikidata

A semi-automated, user-assisted editing tool

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

Edward Betts

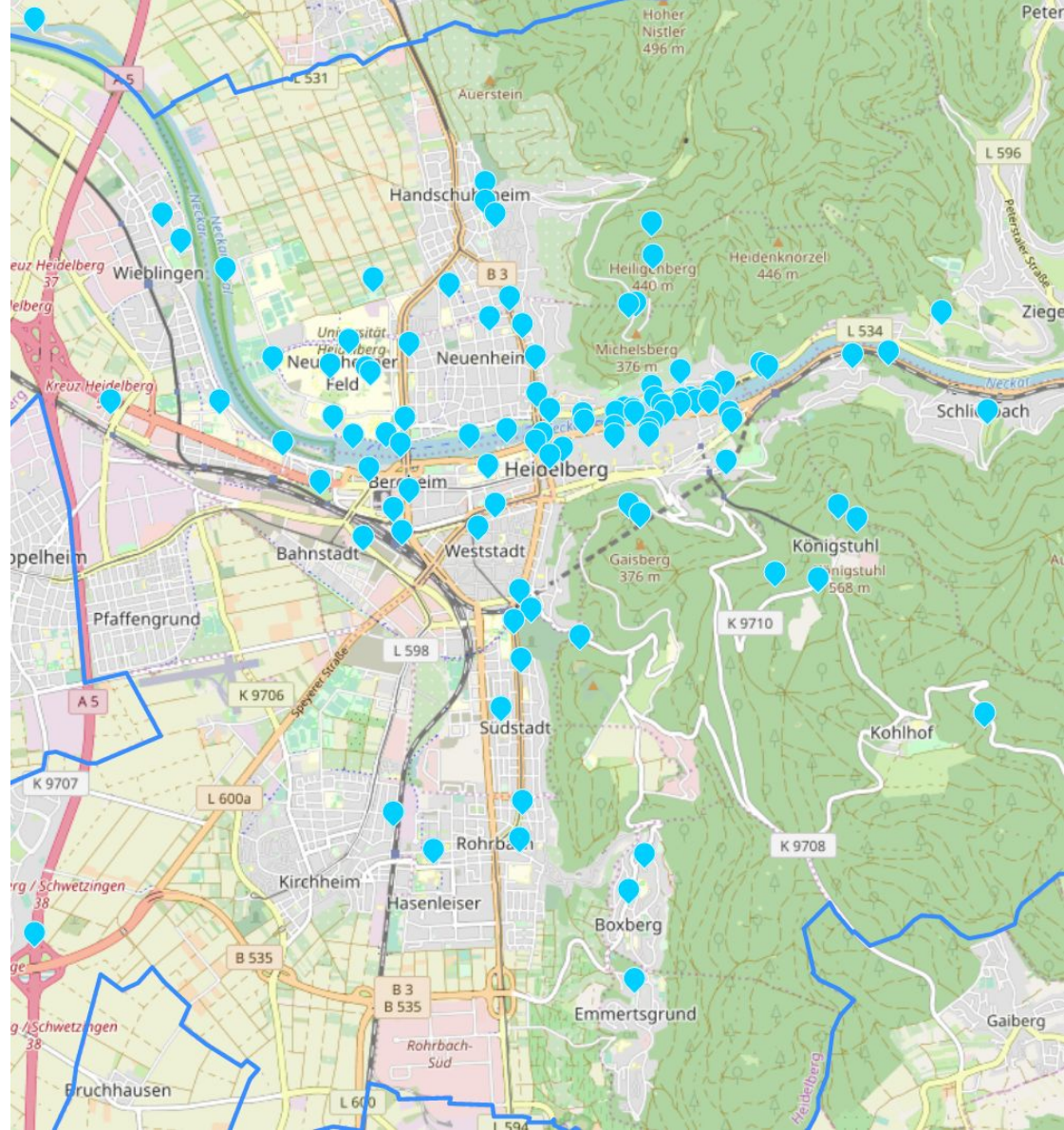
Twitter: @EdwardBetts

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?



Photo by Edward Betts // CC 0

# 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



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

# 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



Photo by Edward Betts // CC 0



# 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

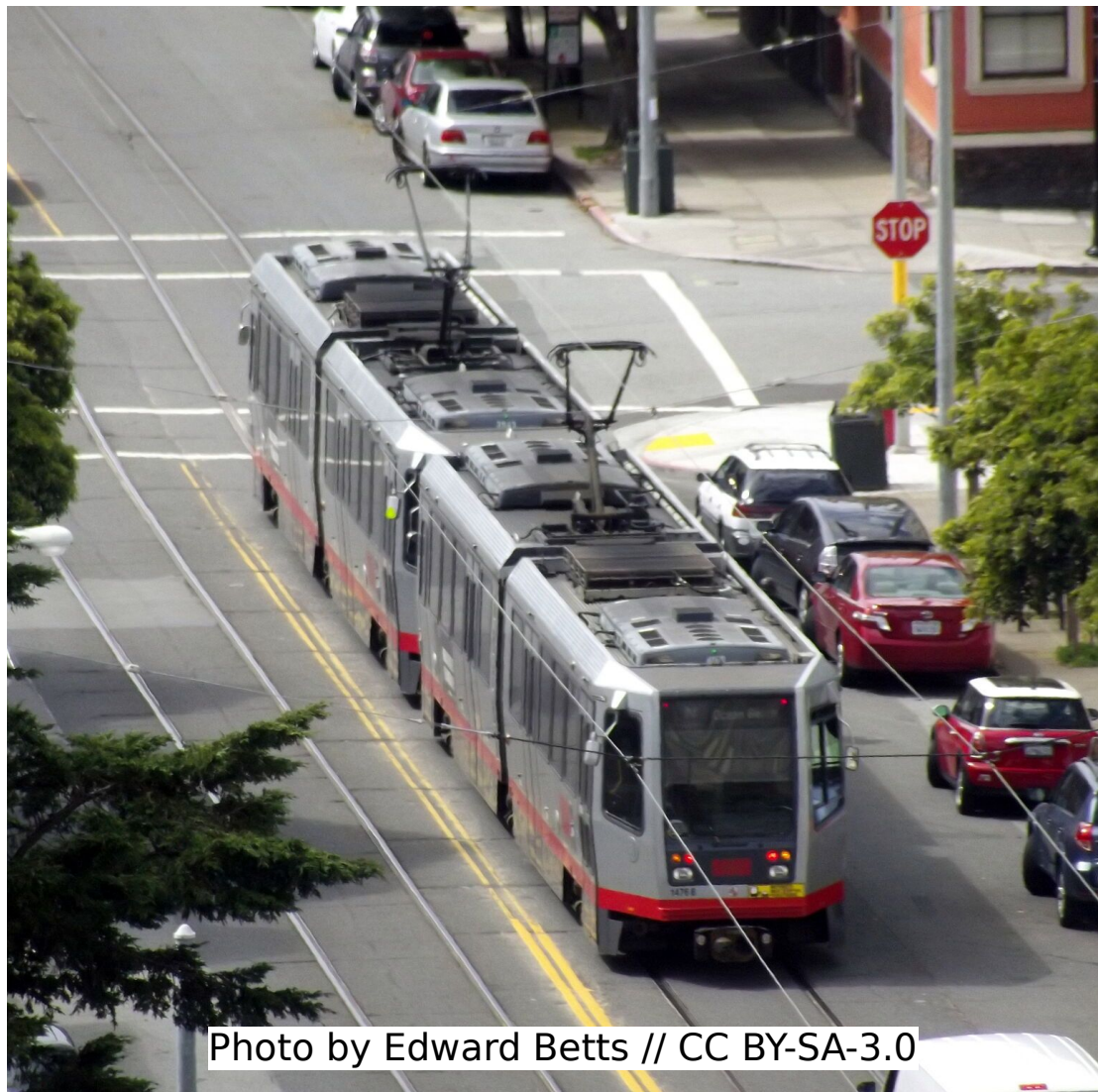


Photo by Edward Betts // CC BY-SA-3.0

# 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

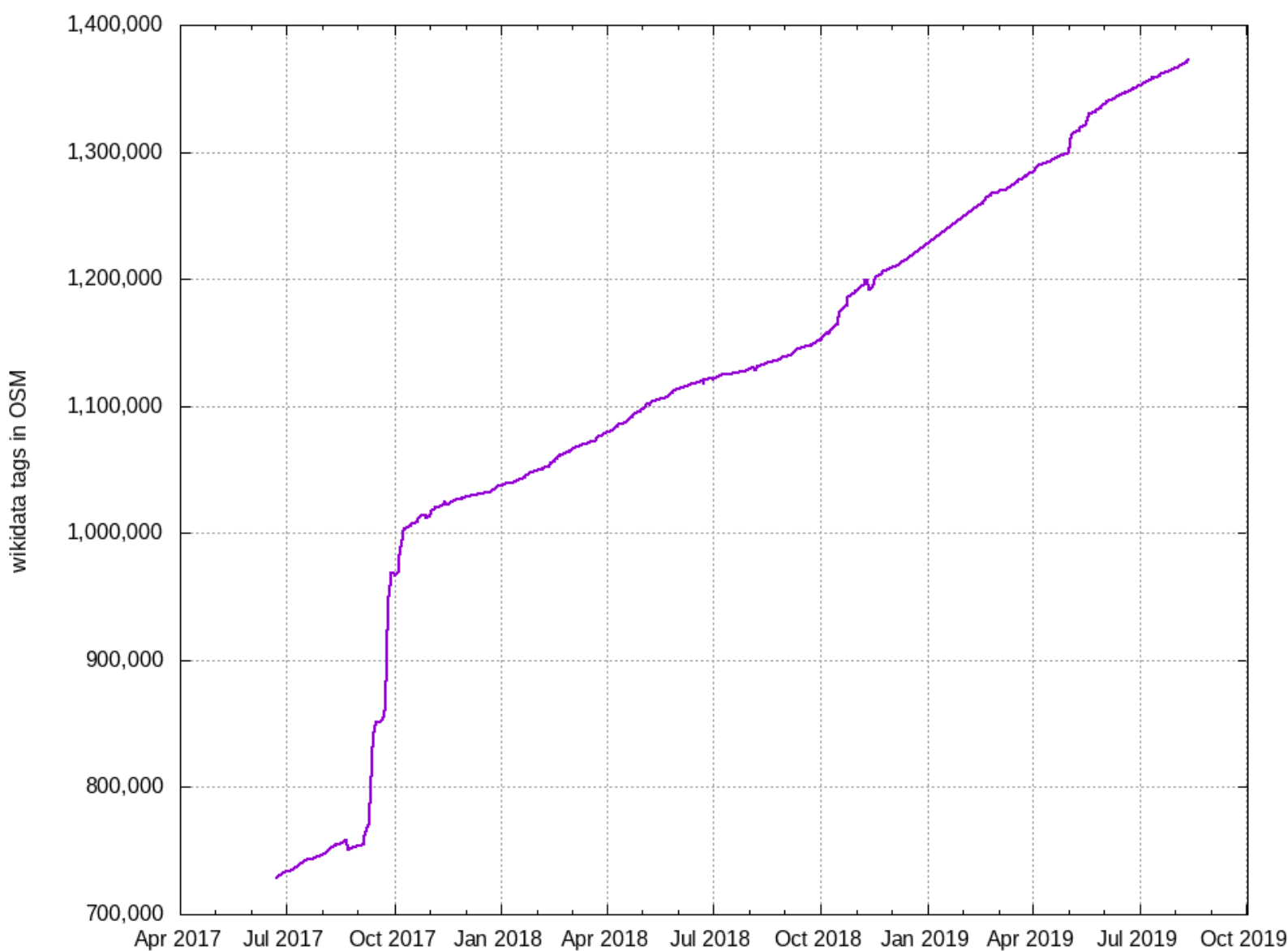
- OSM Overpass
- OSM Nominatim
- OSM map API
- Wikibase API
- Wikipedia API
- Wikidata Query service

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

Photo by Edward Betts // CC 0