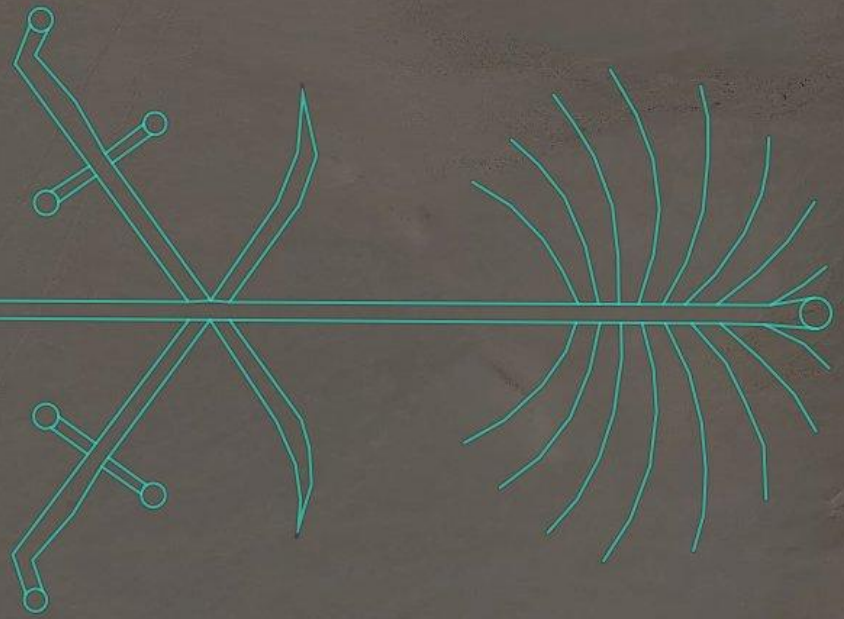
mapbox

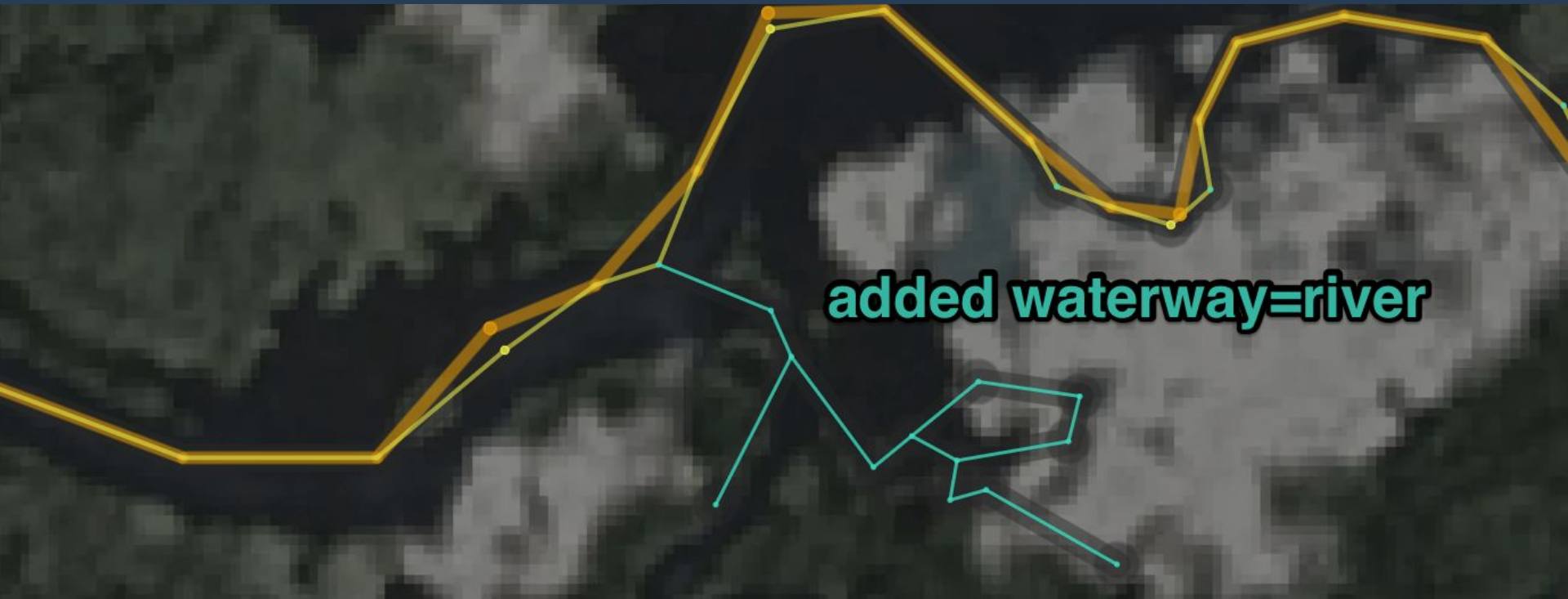
Validating every change



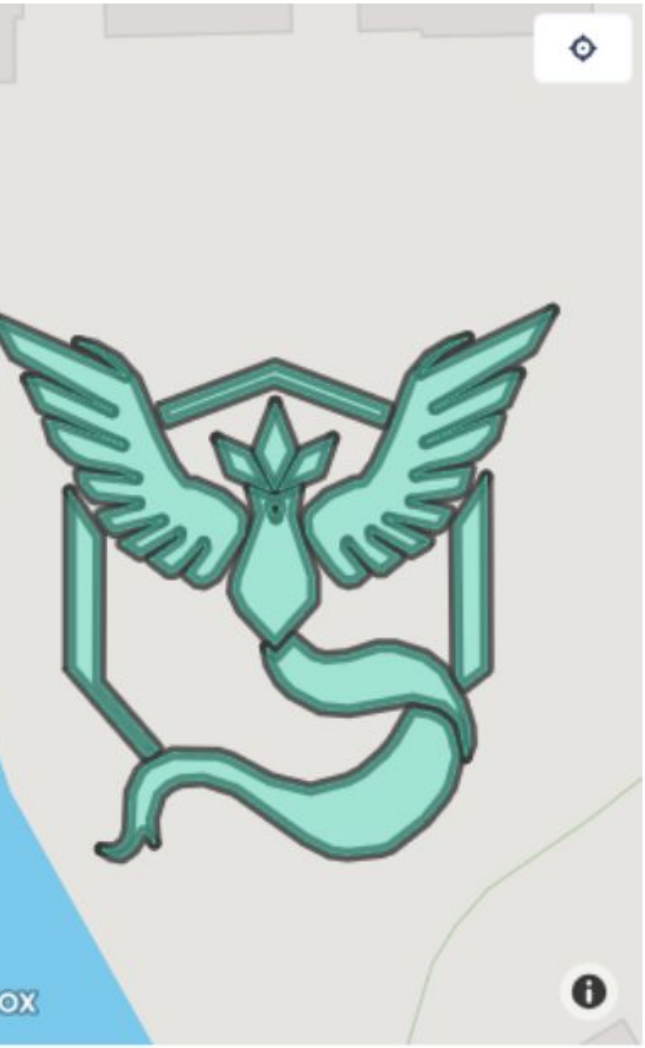
Vandalism

- As online communities grow, destructive actors increase
- OSM is vulnerable
- Mapbox protects the users from harmful data

Incorrect/poor quality



Harmful data


N A St
N D St
W 1st St
E 1st St
W 2nd St
E 2nd St
W 3rd St
W 4th St
74
Perris
74
W 5th St
5th St
W 6th St
W 7th St
D St
Perris Blvd
Bellamo
Park
11th St

Way: I've replaced the batteries but my smoke detector won't stop beeping! (550734402)

Added houses, garages, roads, foot paths, apartments and more Modified streets and more

Edited 5 months ago by

Version #1 · Changeset #55141960

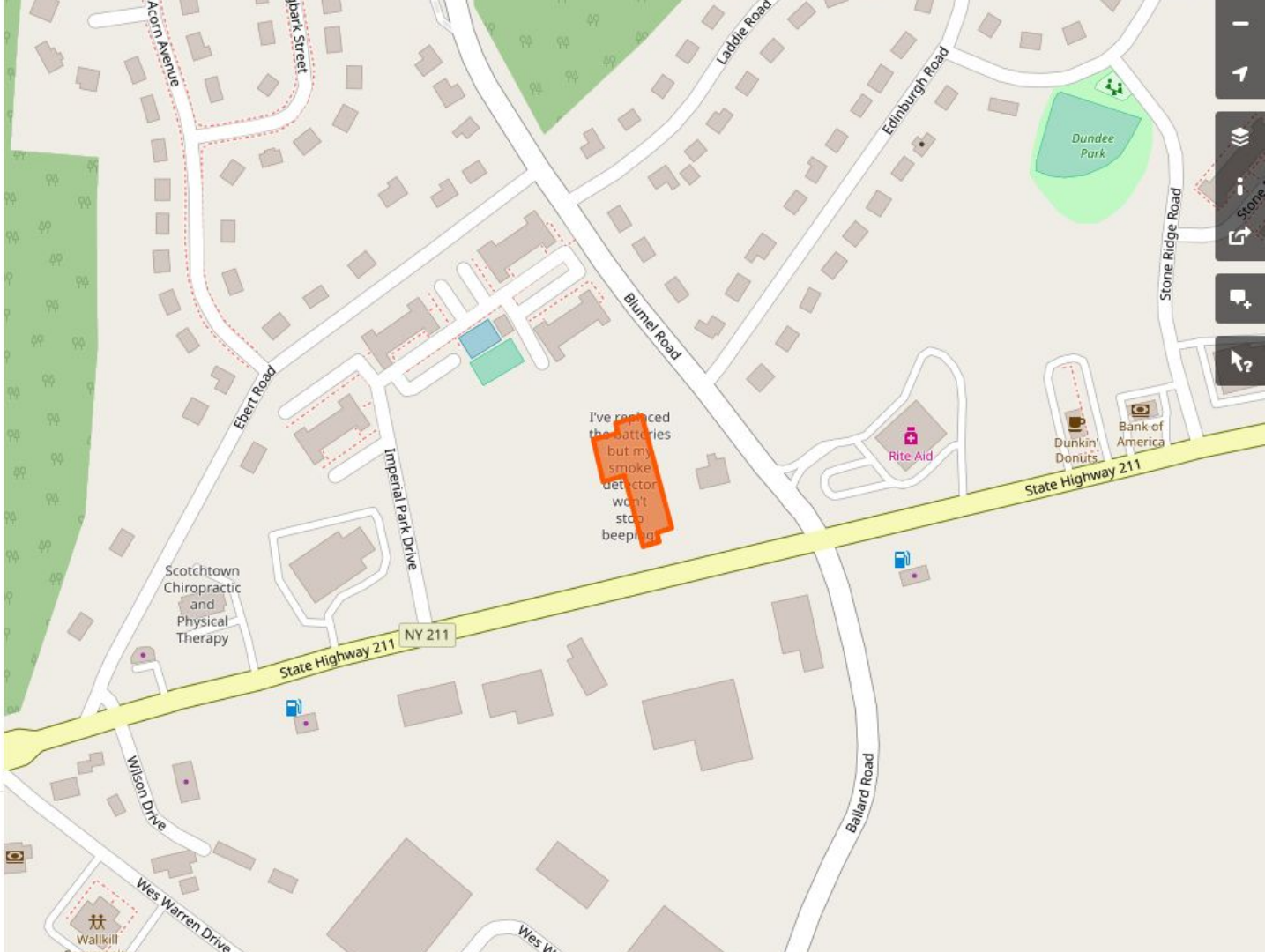
Tags

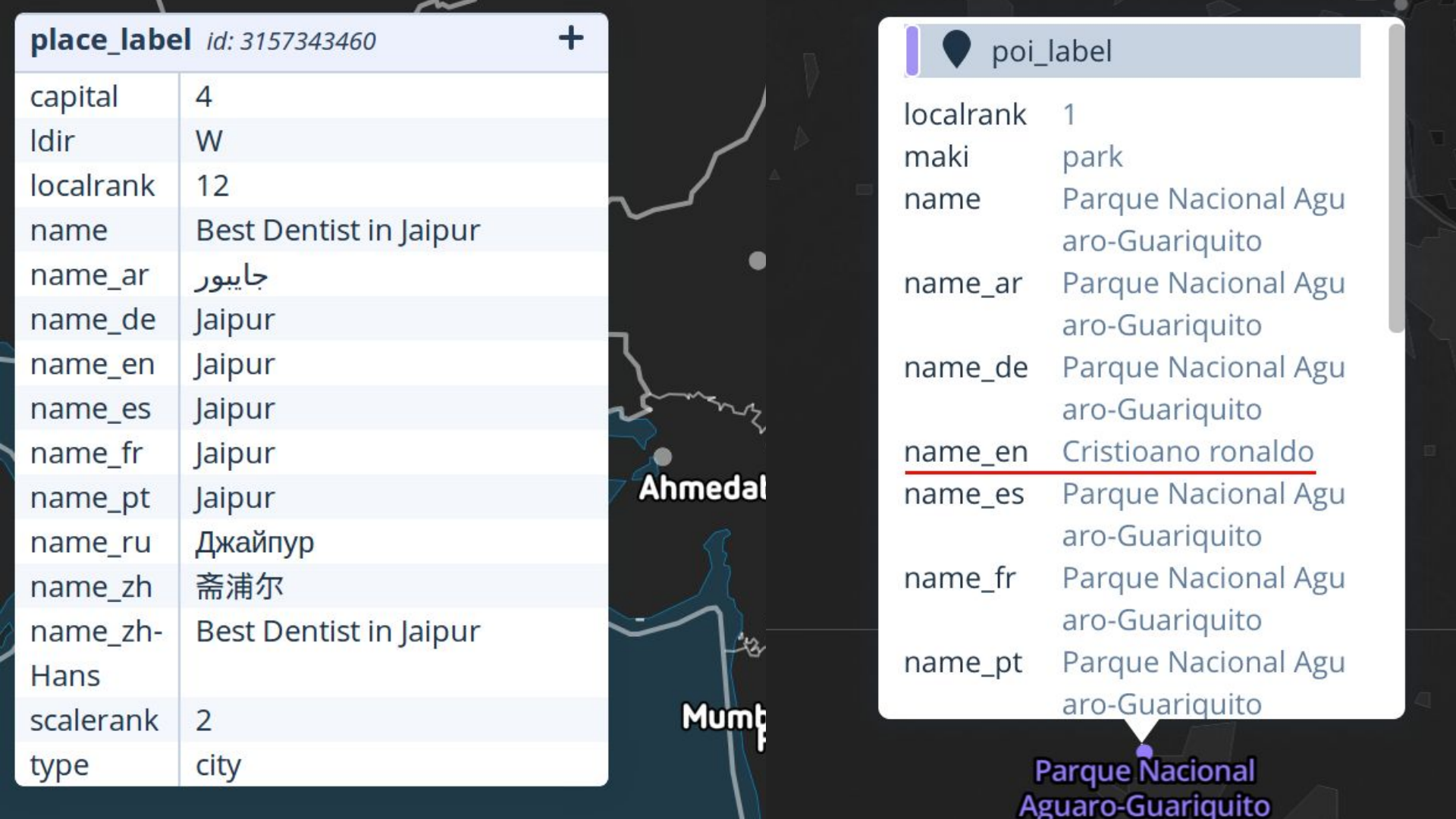
building	commercial
name	I've replaced the batteries but my smoke detector won't stop beeping!

Nodes

5319183869
5319183870
5319183871
5319183872
5319183873
5319183874
5319183875
5319183876
5319183877
5319183878
5319183869

Download XML · View History



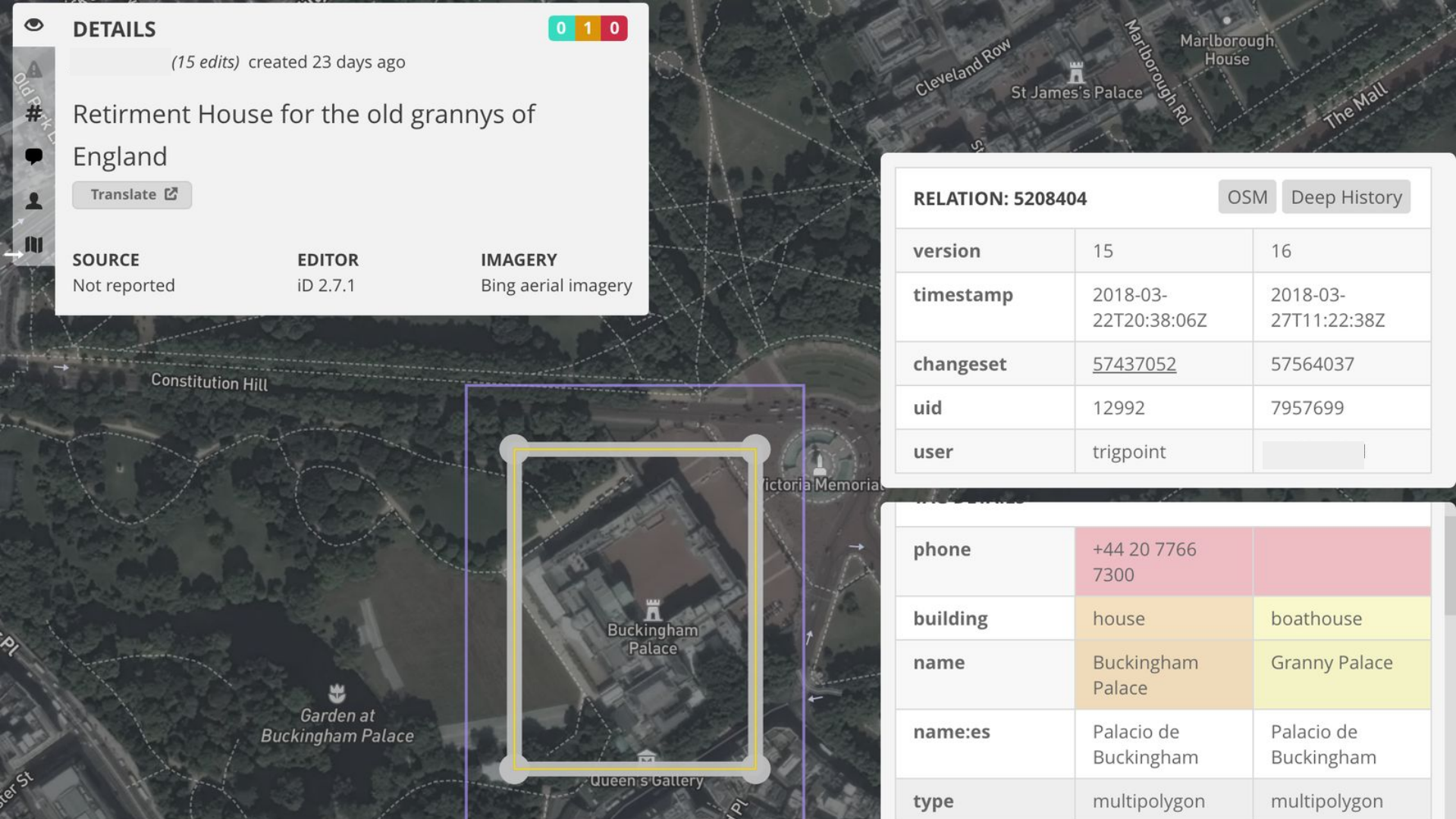


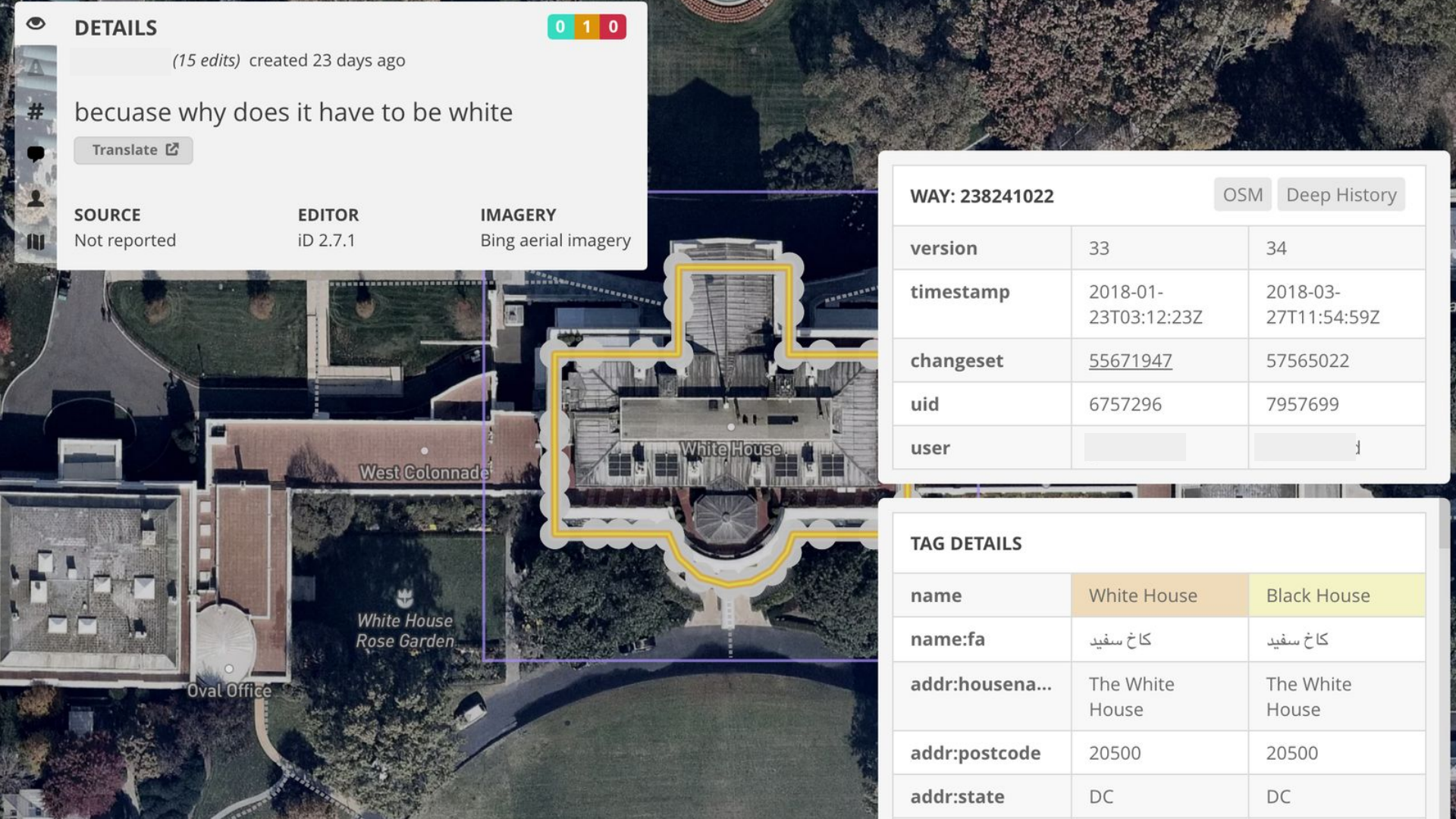
place_label *id: 3157343460*

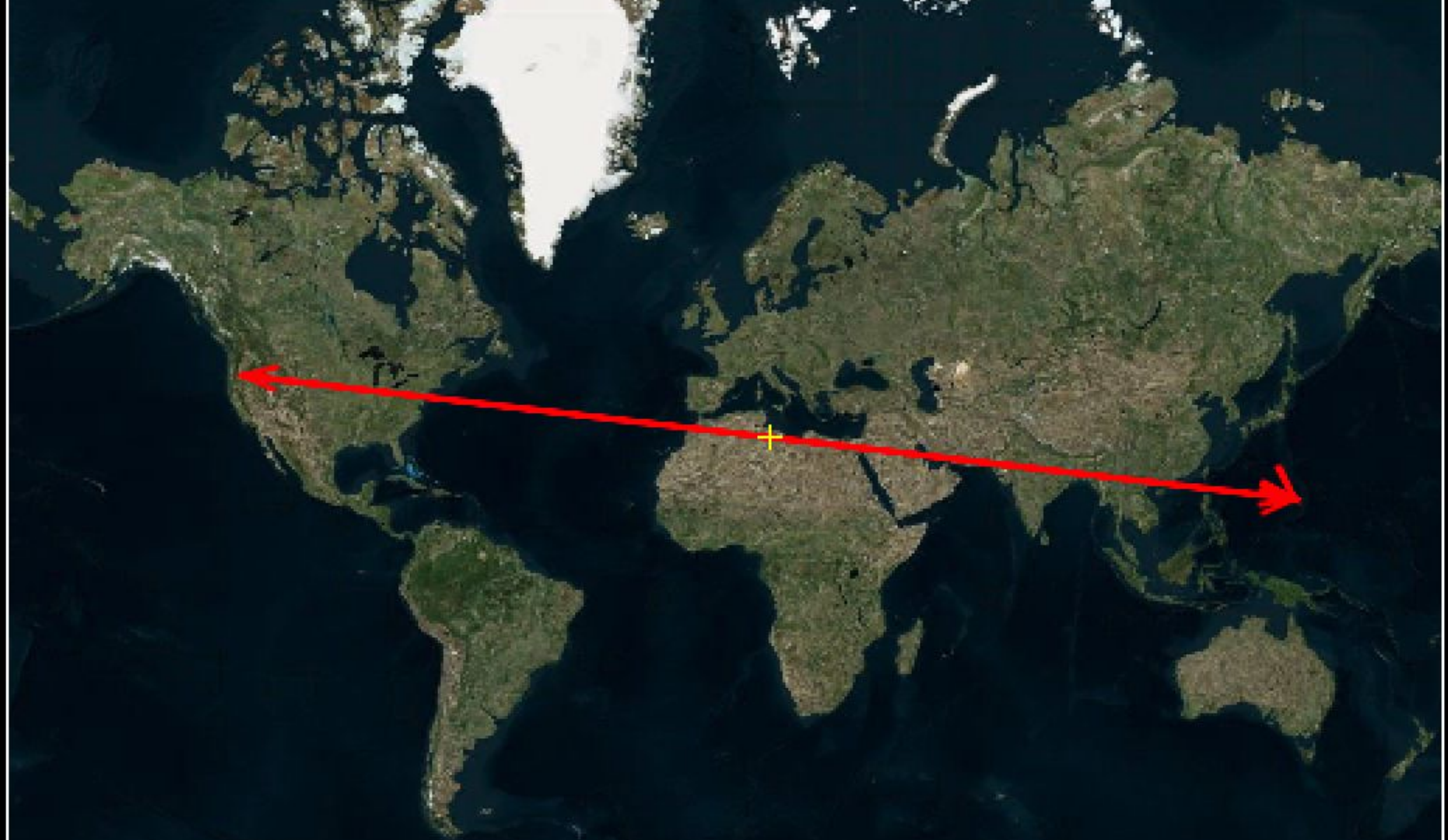
capital	4
ldir	W
localrank	12
name	Best Dentist in Jaipur
name_ar	جايبور
name_de	Jaipur
name_en	Jaipur
name_es	Jaipur
name_fr	Jaipur
name_pt	Jaipur
name_ru	Джайпур
name_zh	斋浦尔
name_zh-Hans	Best Dentist in Jaipur
scalerank	2
type	city

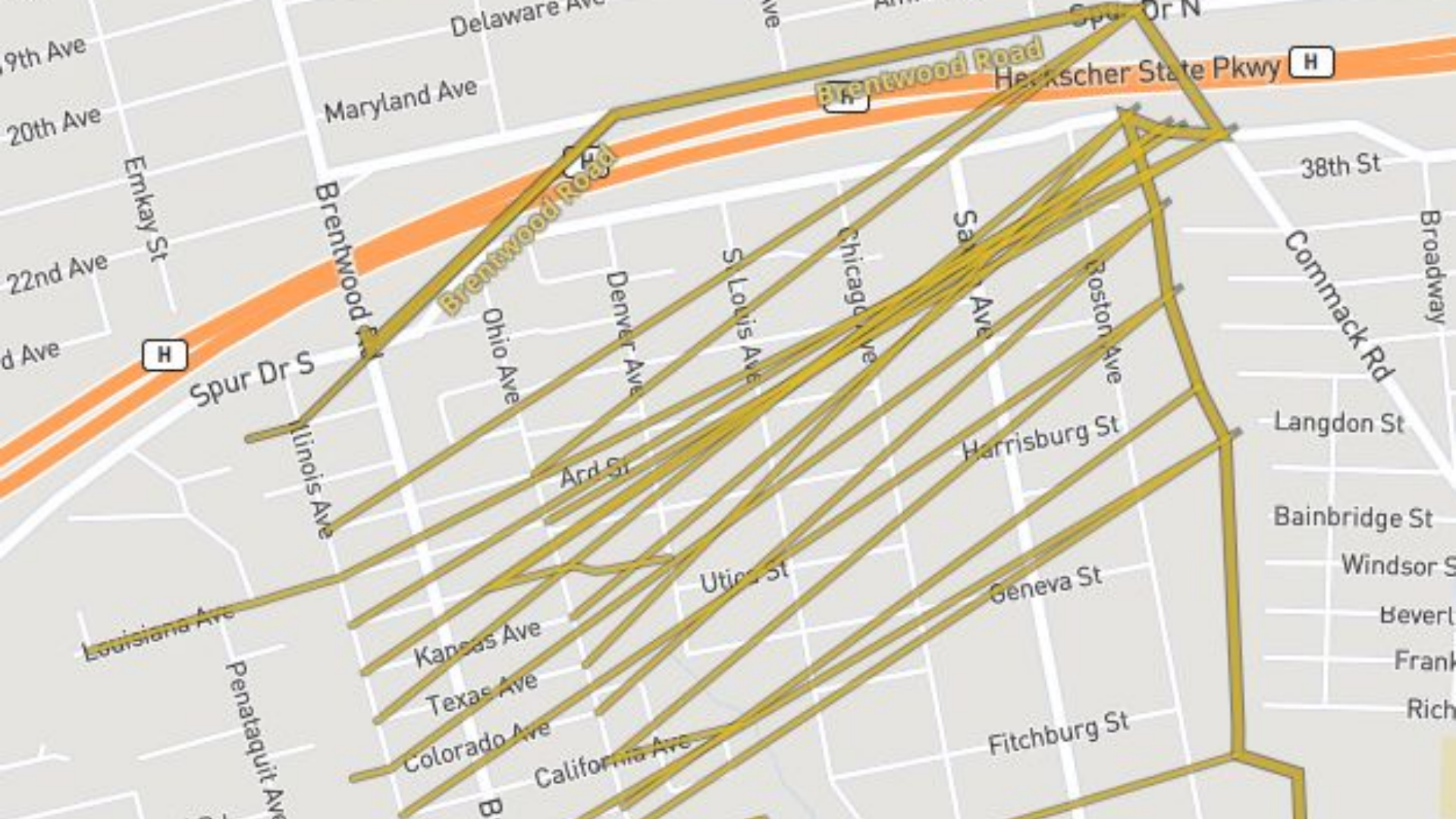
poi_label

localrank	1
maki	park
name	Parque Nacional Aguaro-Guariquito
name_ar	Parque Nacional Aguaro-Guariquito
name_de	Parque Nacional Aguaro-Guariquito
name_en	Cristioano ronaldo
name_es	Parque Nacional Aguaro-Guariquito
name_fr	Parque Nacional Aguaro-Guariquito
name_pt	Parque Nacional Aguaro-Guariquito

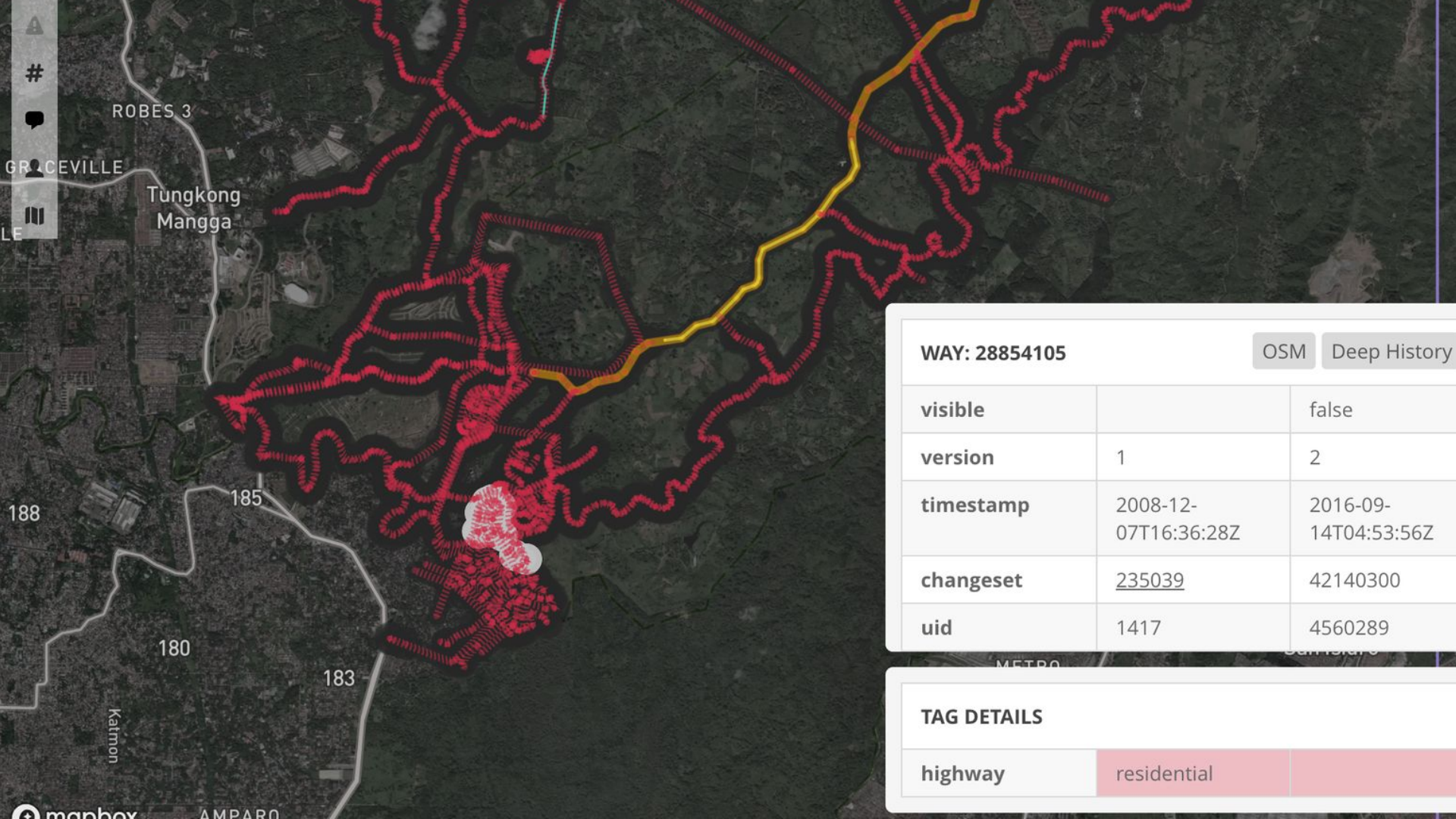

DETAILS
0 1 0
(15 edits) created 23 days ago
Retirment House for the old grannys of England
Translate
SOURCE
Not reported
EDITOR
iD 2.7.1
IMAGERY
Bing aerial imagery
Cleveland Row
St James's Palace
Marlborough Rd
Marlborough House
The Mall
Constitution Hill
Buckingham Palace
Garden at Buckingham Palace
Queen's Gallery
Victoria Memorial
RELATION: 5208404
OSM
Deep History
version 15 16
timestamp 2018-03-22T20:38:06Z 2018-03-27T11:22:38Z
changeset 57437052 57564037
uid 12992 7957699
user trigpoint
phone +44 20 7766 7300
building house boathouse
name Buckingham Palace Granny Palace
name:es Palacio de Buckingham Palacio de Buckingham
type multipolygon multipolygon


DETAILS
0 1 0
(15 edits) created 23 days ago
becuase why does it have to be white
Translate
SOURCE
Not reported
EDITOR
iD 2.7.1
IMAGERY
Bing aerial imagery
West Colonnade
White House
White House
Rose Garden
Oval Office
WAY: 238241022
OSM
Deep History
version 33 34
timestamp 2018-01-23T03:12:23Z 2018-03-27T11:54:59Z
changeset 55671947 57565022
uid 6757296 7957699
user
TAG DETAILS
name White House Black House
name:fa کاخ سفید کاخ سفید
addr:housena... The White House The White House
addr:postcode 20500 20500
addr:state DC DC





Delaware Ave
Spur Dr N
Brentwood Road
Heckscher State Pkwy
H
Maryland Ave
20th Ave
Emkay St
Brentwood Rd
38th St
22nd Ave
Commack Rd
Broadway
Spur Dr S
Ohio Ave
Denver Ave
St Louis Ave
Boston Ave
Langdon St
Harrisburg St
Ard St
Illinois Ave
Bainbridge St
Windsor S
Geneva St
Louisiana Ave
Kansas Ave
Texas Ave
Colorado Ave
California Ave
Penataquit Ave
Fitchburg St

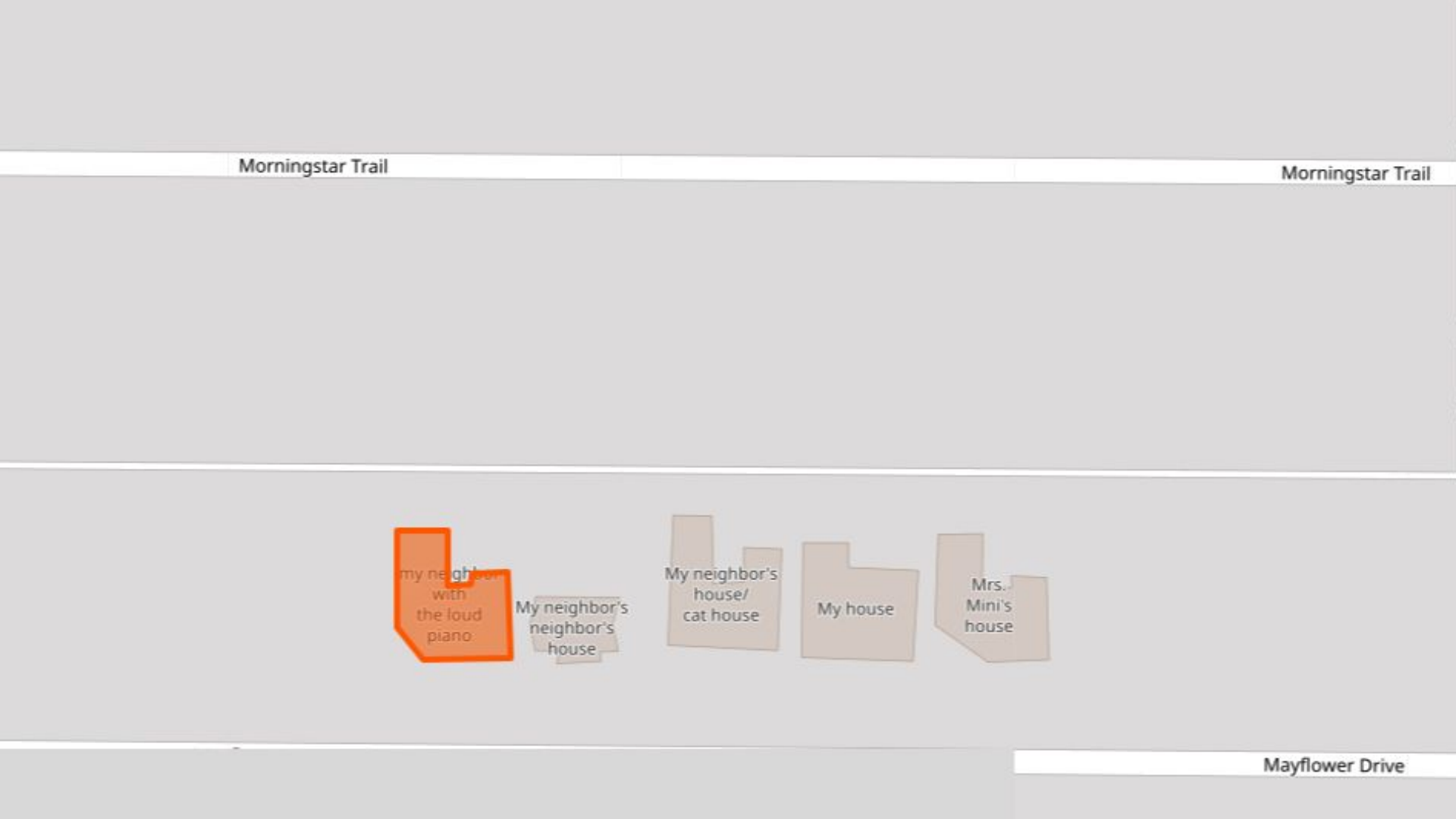


WAY: 28854105 OSM Deep History

visible		false
version	1	2
timestamp	2008-12-07T16:36:28Z	2016-09-14T04:53:56Z
changeset	235039	42140300
uid	1417	4560289

TAG DETAILS

highway	residential


Morningstar Trail
Morningstar Trail
my neighbor with the loud piano
My neighbor's neighbor's house
My neighbor's house/ cat house
My house
Mrs. Mini's house
Mayflower Drive

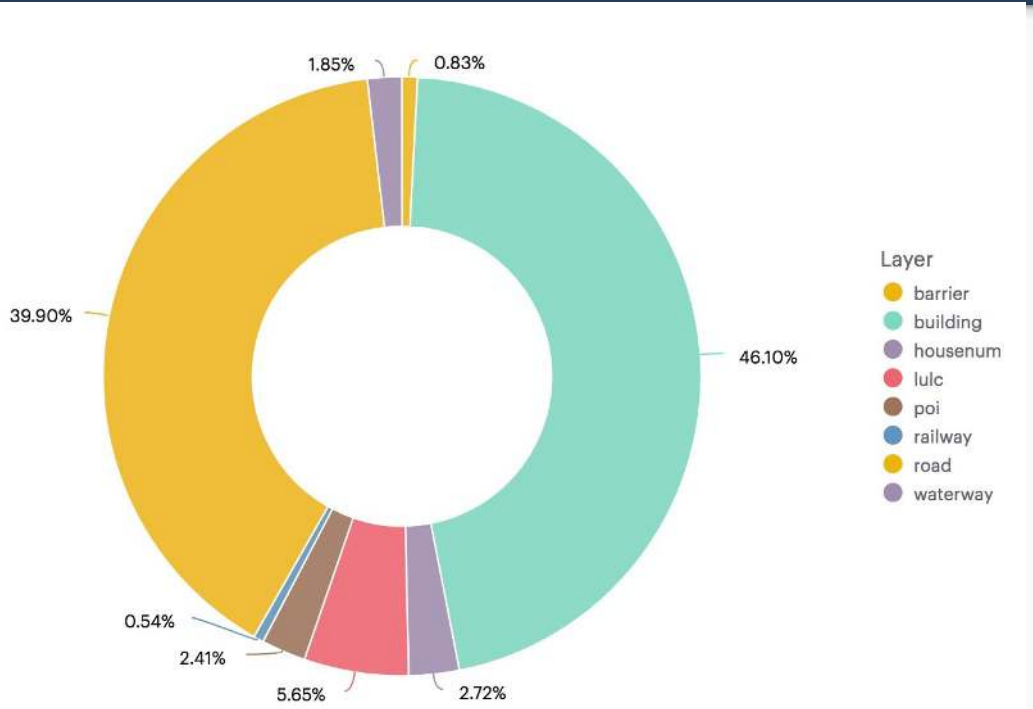
Factory
of Small
Children
Meat

Statistics (per million changes)

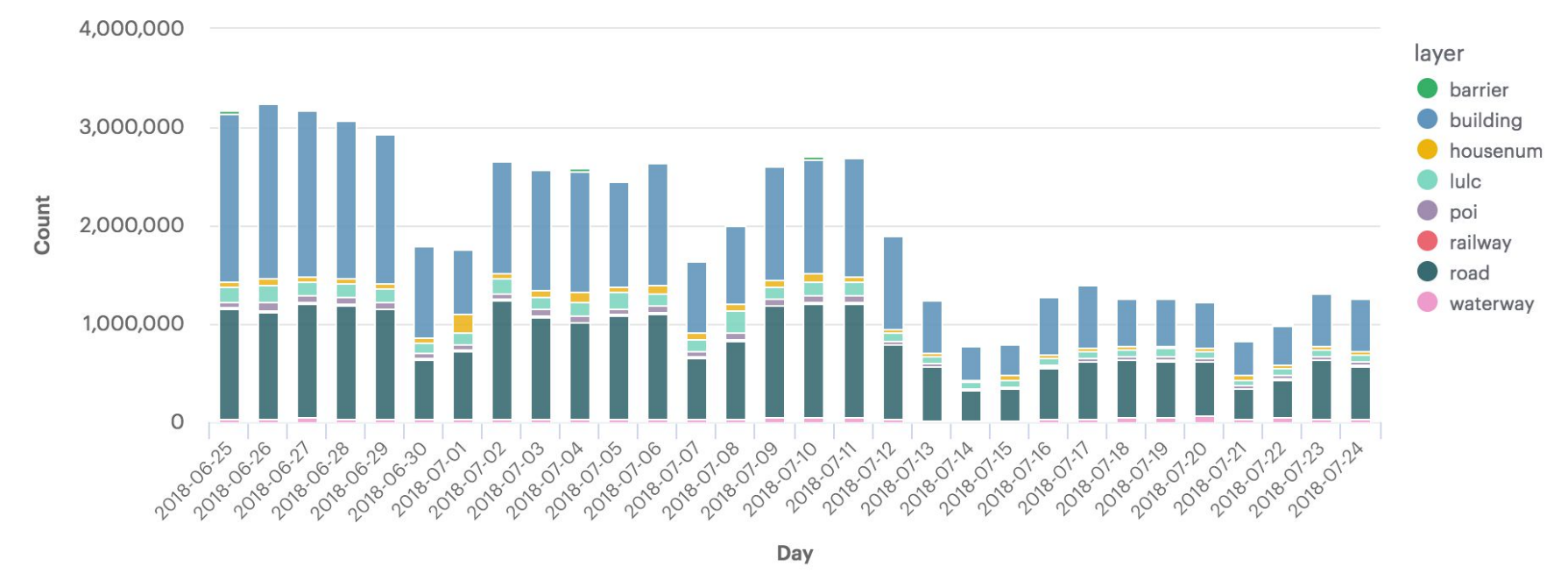
- 570 incorrect labels
- 300 editing failures (dragged nodes)
- 160 spam incidents
- 100 harmful deletions
- 50 obscene labels
- 20 graffiti
- ...

Daily change statistics

- 2 million features get touched
- 10k label edits
- 30k changesets
  - 0.2% is vandalism
  - 2% are low quality
- 20k new contributors join monthly
  - 30% of new users make a mistake in their first 10 edits



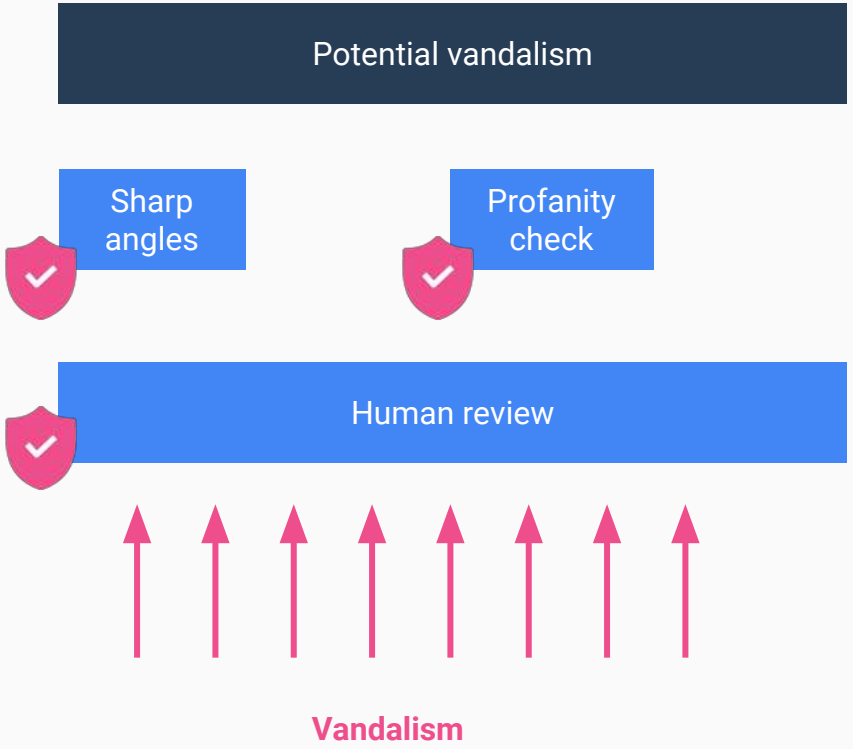
Daily touched features by data layer



Past approaches at Mapbox

- Validating changesets
- Relying only on algorithms
- Monitor new users
- Building blacklists

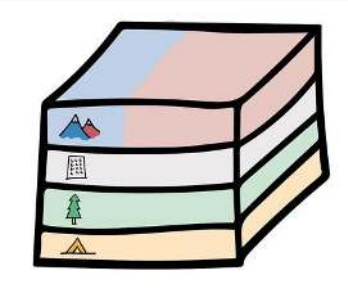
One approach does not address all cases of vandalism.



Approach

Step 1

Split the OSM mono layer into data layers



Step 2

Diff the changes per day.

Step 3

Cluster daily changes into deltas.

A new unit of change

Approach

Step 4

Review the daily changes

Step 5

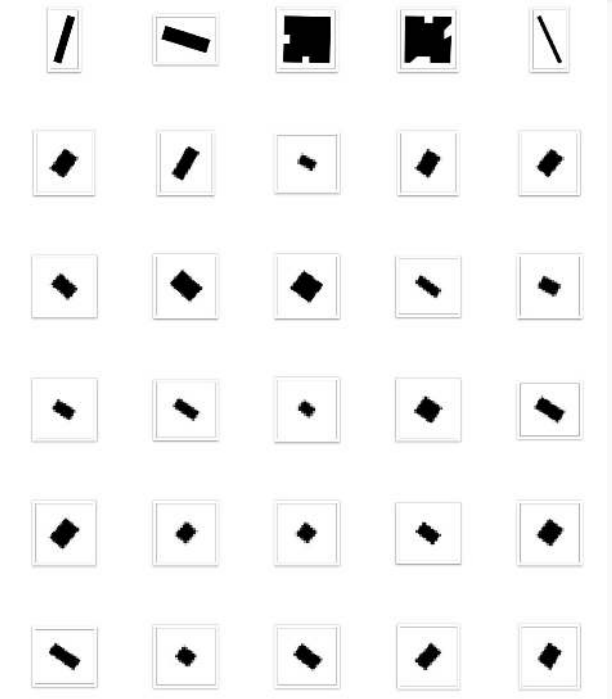
Apply the updates to the map and protect from harmful changes

Step 6

Share harmful changes and fix them

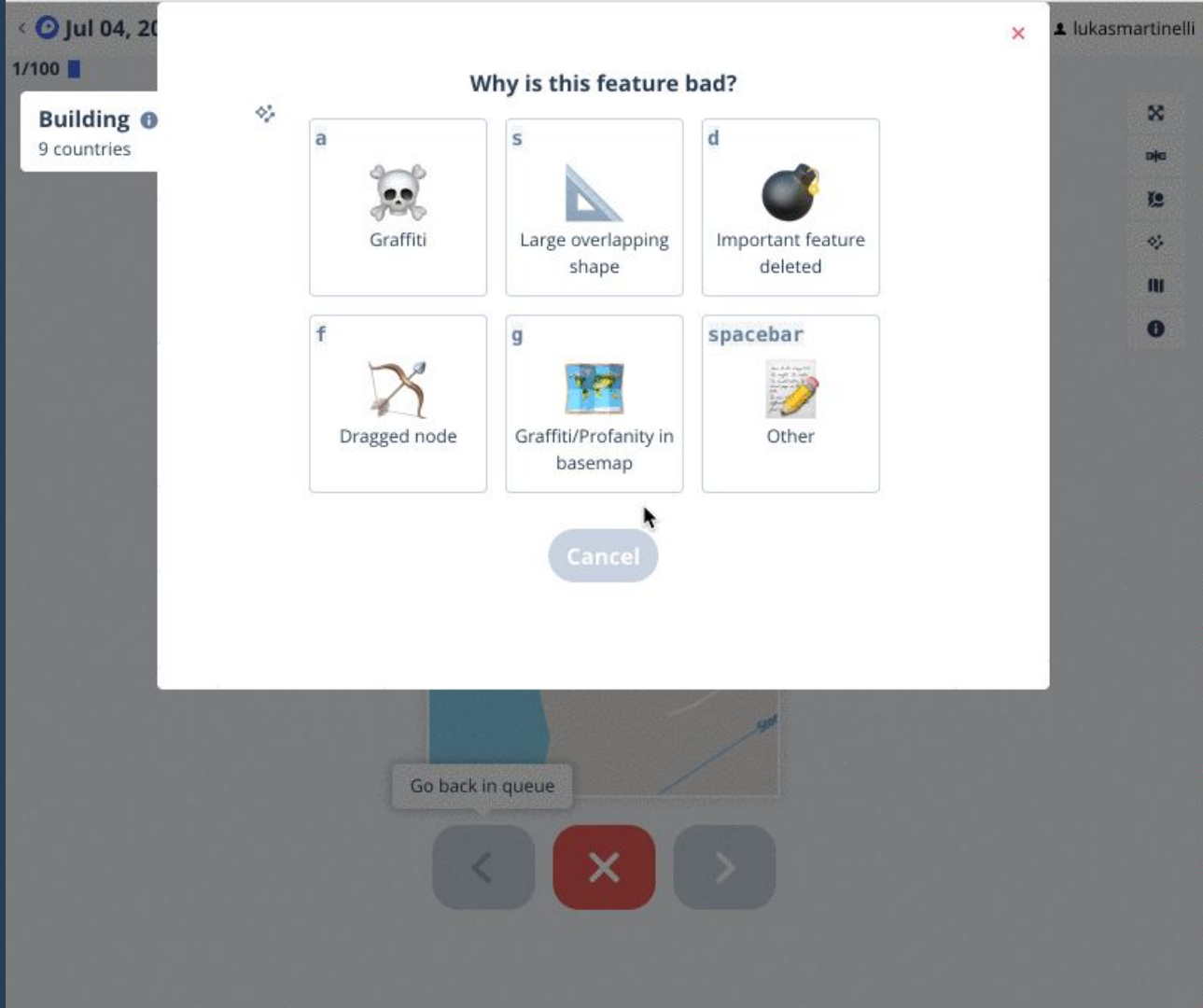
Machine review

- Profanity checking in 100 languages for labels
- Use NLP to determine how likely a label is a place name
- Shape classifiers for likelihood of a shape being a building
- Drastic changes to stable features
- …



Human review

- Review changes in
  - geometry
  - labels
  - hierarchy
  - primary tags
- Classify harmful changes

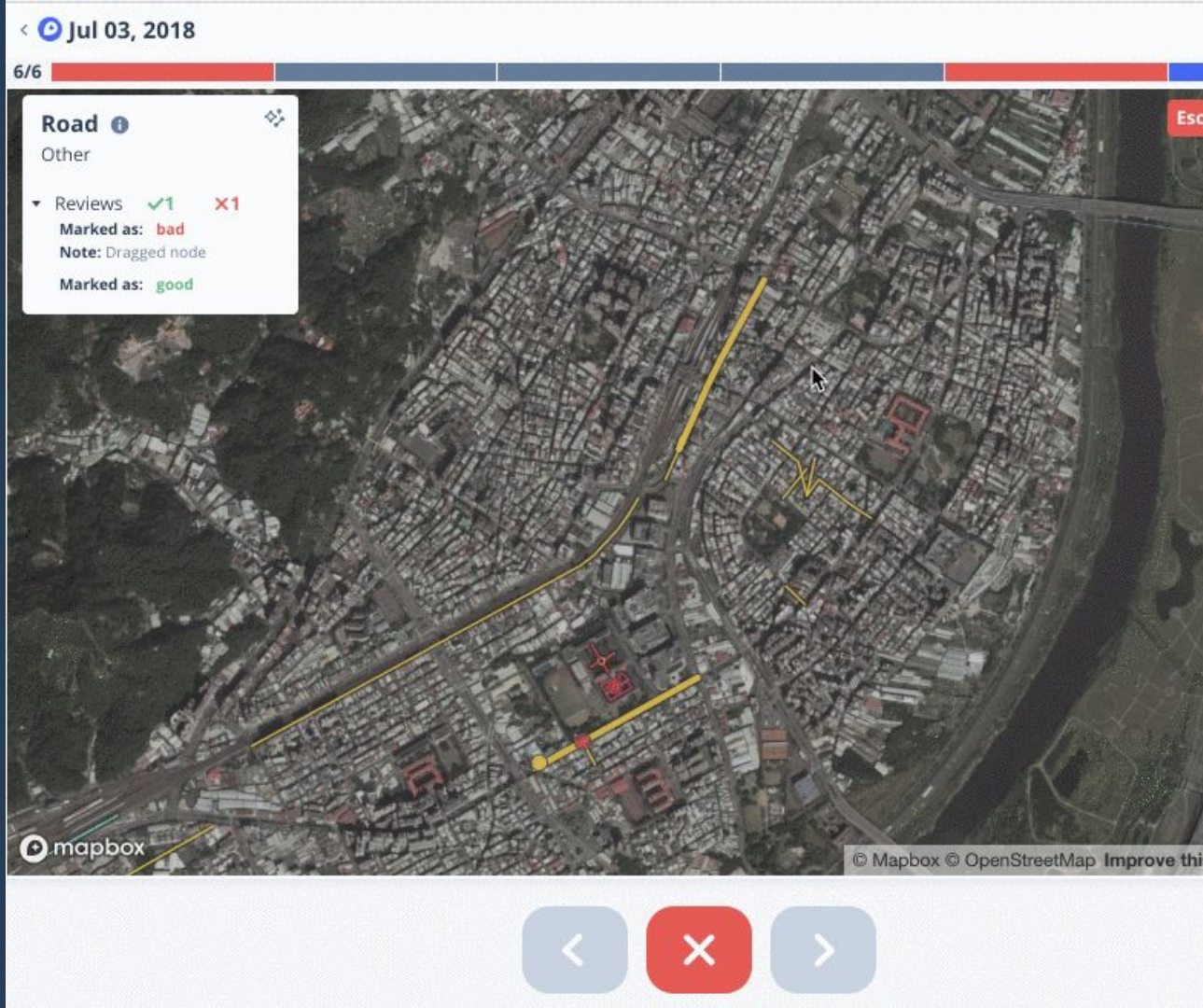


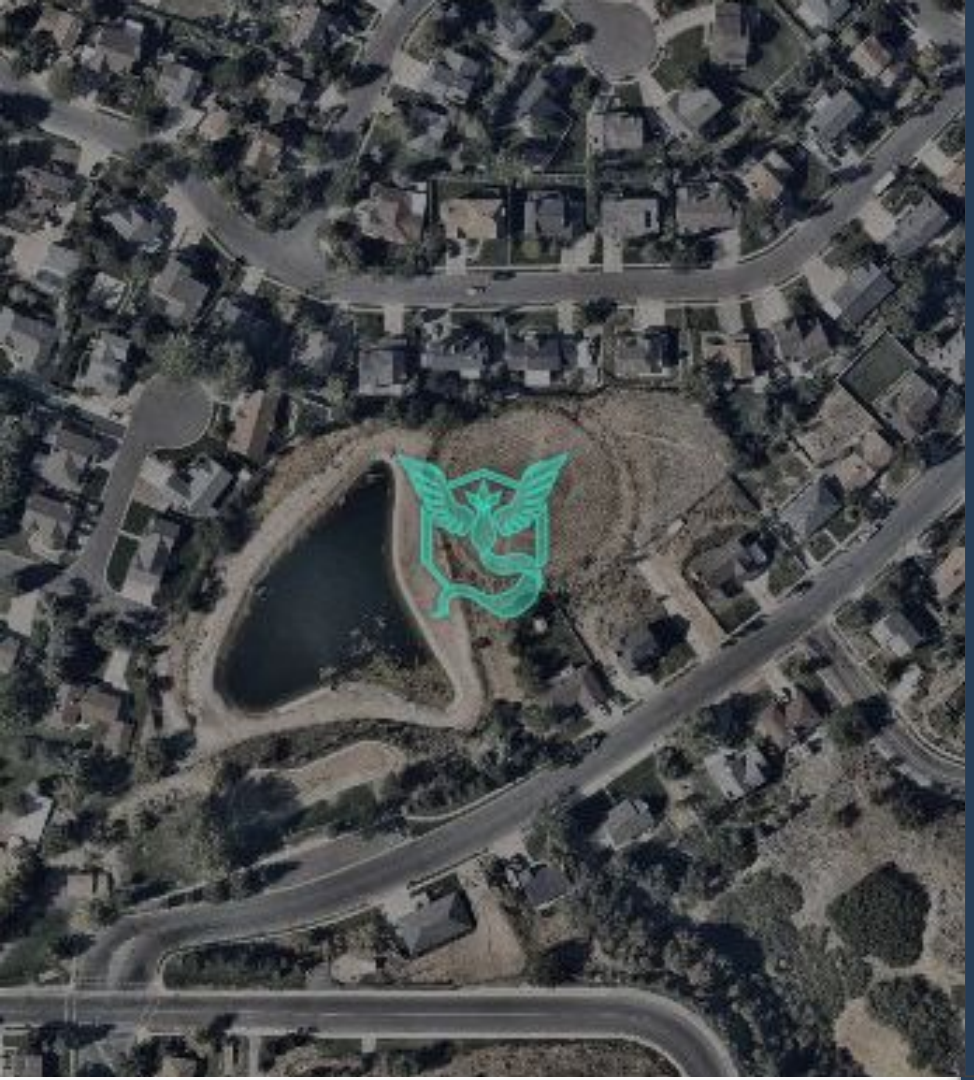
Water

4 countries

QA of reviews

- Review team regularly gets sampled
- >99% accuracy for selected cases
- Expert mappers double check each review and single out the problematic features





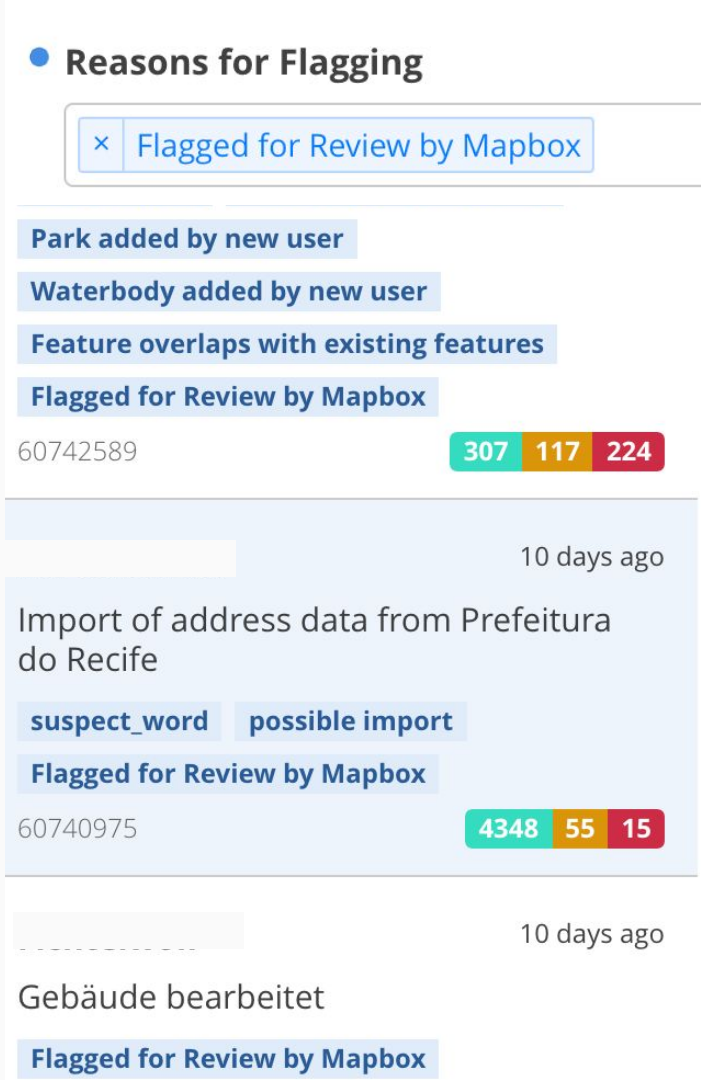
Review statistics from Mapbox

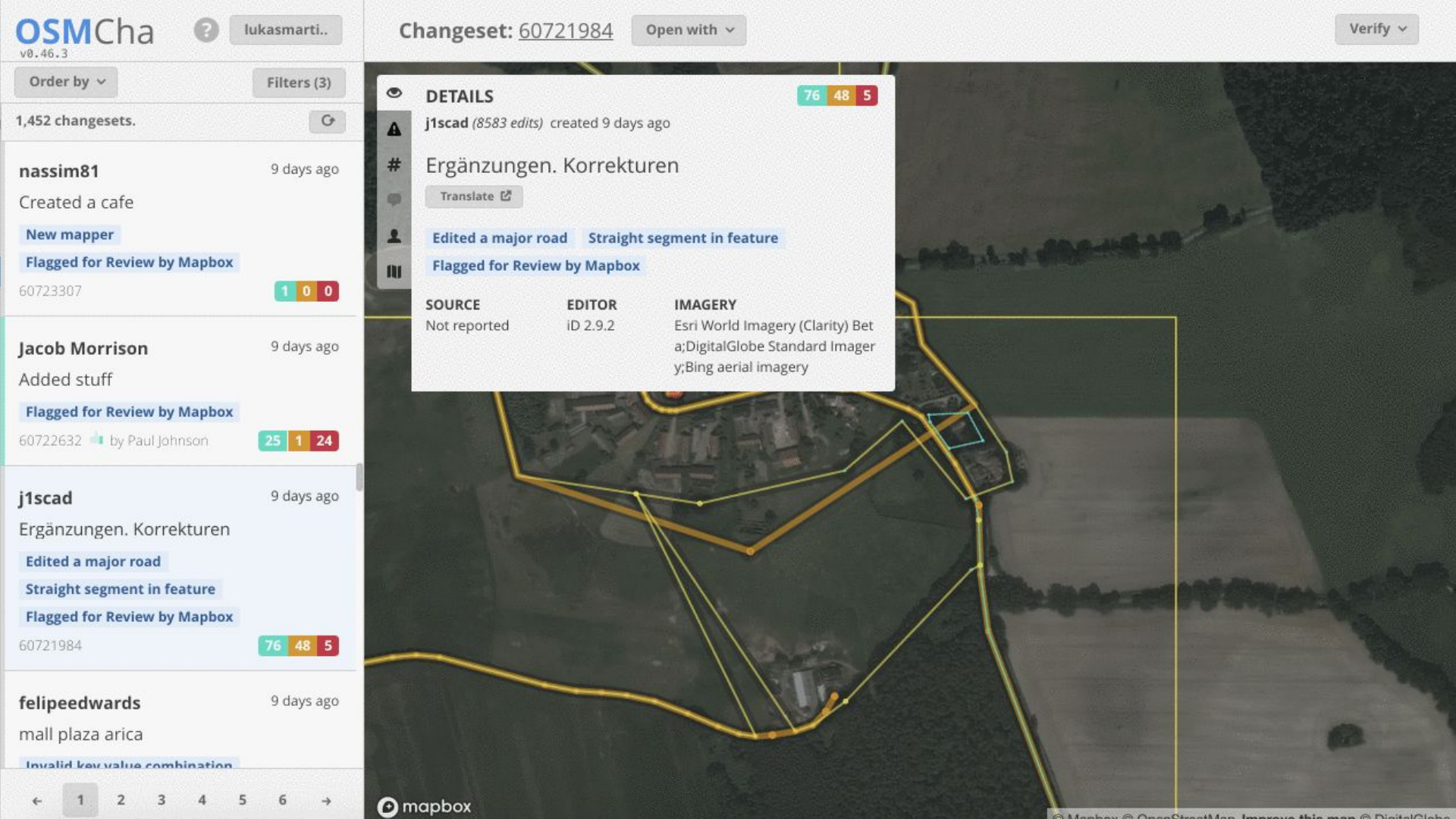
- Our review team reviews all 80’000 changes on a daily basis
- We flag around 1000-2000 changes a day
- We fix >200 defects on a daily basis
- 50% of issues are fixed by OSM

Daily catch

Sharing vandalism detections

- **osmcha.mapbox.com** is the one stop shop for OSM validation
- All our harmful detections are made public
- Mapbox regularly fixes harmful data




OSMCha
v0.46.3
lukasmarti..
Changeset: 60721984
Open with
Verify
Order by
Filters (3)
1,452 changesets.
nassim81
9 days ago
Created a cafe
New mapper
Flagged for Review by Mapbox
60723307
1 0 0
Jacob Morrison
9 days ago
Added stuff
Flagged for Review by Mapbox
60722632
by Paul Johnson
25 1 24
j1scad
9 days ago
Ergänzungen. Korrekturen
Edited a major road
Straight segment in feature
Flagged for Review by Mapbox
60721984
76 48 5
felipeedwards
9 days ago
mall plaza arica
Invalid key value combination
1 2 3 4 5 6
DETAILS
76 48 5
j1scad (8583 edits) created 9 days ago
Ergänzungen. Korrekturen
Translate
Edited a major road
Straight segment in feature
Flagged for Review by Mapbox
SOURCE
Not reported
EDITOR
iD 2.9.2
IMAGERY
Esri World Imagery (Clarity) Beta;DigitalGlobe Standard Imagery;Bing aerial imagery
mapbox

Takeaways

- Only 0.2% of edits are vandalism
- OSM is eventually consistent
- Mapbox provides you a validated view of OSM
- Let’s protect the future of OSM before vandalism becomes a bigger problem
- We need better shared monitoring efforts