What you map is

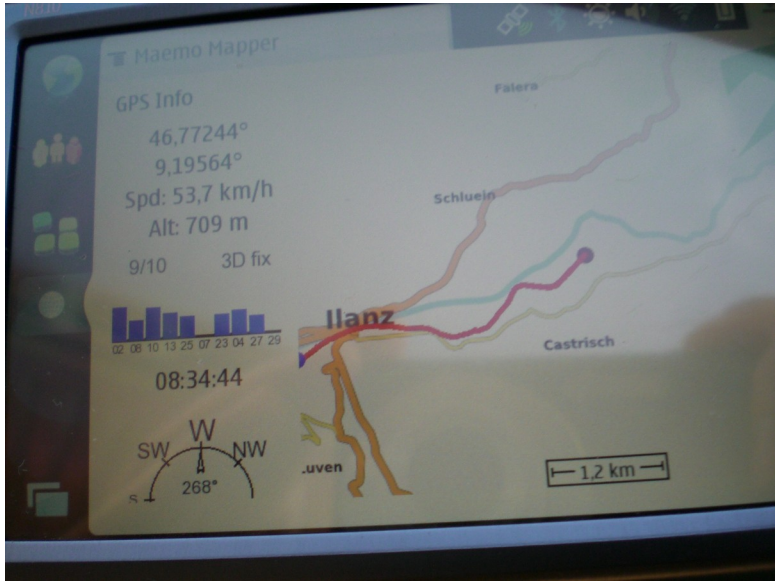
not always

what you get

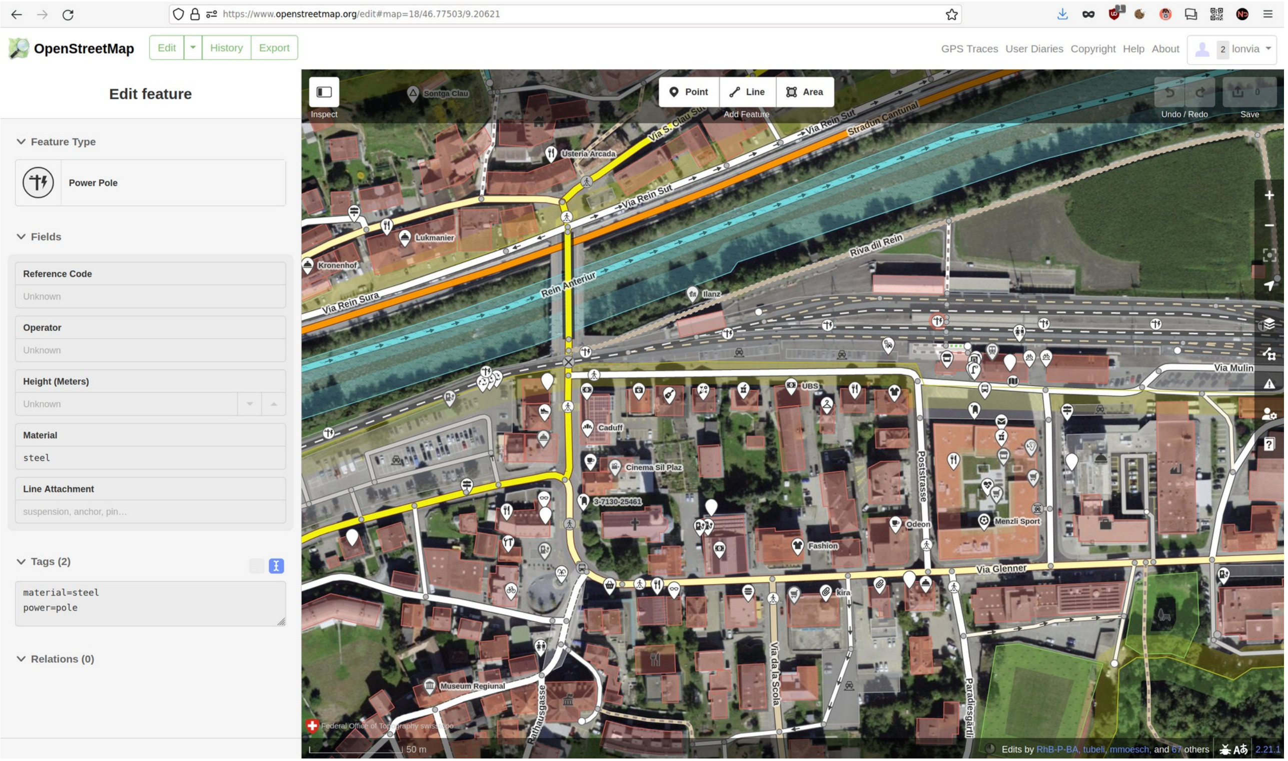
Richard Fairhurst, Sarah Hoffmann

State of the Map 2022, Firenze

Once upon a time



And today



https://www.openstreetmap.org/edit#map=18/46.77503/9.20621
OpenStreetMap
Edit
History
Export
GPS Traces
User Diaries
Copyright
Help
About
lonvia
Edit feature
Feature Type
Power Pole
Fields
Reference Code
Unknown
Operator
Unknown
Height (Meters)
Unknown
Material
steel
Line Attachment
suspension, anchor, pin...
Tags (2)
material=steel
power=pole
Relations (0)
Inspect
Point
Line
Area
Add Feature
Undo / Redo
Save
Via Rein Sura
Rein Anteriur
Riva dil Rein
Via Mulin
Via Glenner
Poststrasse
Via da la Scola
Paradiesgärtli
Rathausgasse
Museum Regiunal
Cinema Sil Plaz
Menzli Sport
Fashion
Odeon
UBS
Caduff
Lukmanier
Kronenhof
Ilanz
Usteria Arcada
50 m
Edits by RhB-P-BA, tubell, mmoesch, and 67 others
2.21.1

The customer side

How do data users see OSM data?

What can mappers do to make it useful?

(We all want our mapping to be *used*, right?)

About Sarah

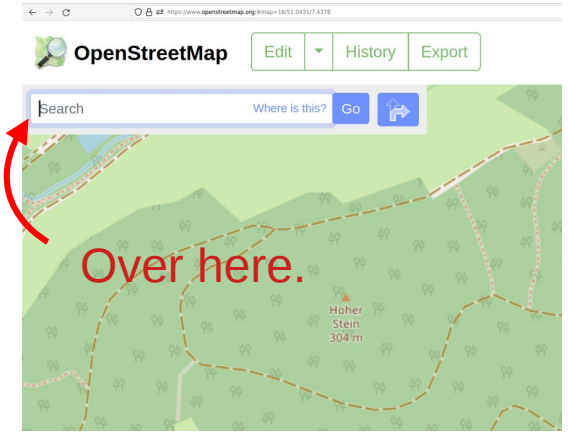
- 2008: first OSM edit
- 2009: first self-made OSM map for N900
- 2010: started waymarkedtrails
- 2012: first Nominatim server setup
- now: full-time OSM software development addict

About Nominatim

- started in 2010
- Search tool for OSMers:
  - Finds what you have mapped.
  - Always be fresh.
  - We speak all languages.
- Biggest challenge:

  Places and boundaries.
- Favorite problem:

  OSMers mapping for the geocoder.



About Richard

- October 2004: “Guess I probably ought to join the list then”
- Wrote Potlatch (1/2/3), started iD, coded occasional bits of functionality like the osm.org routing UI, fought in the trenches of the licence change
- Now: vector tiles with tilemaker, routing with cycle.travel
- Likes bikes, boats, Ruby and Lua

About cycle.travel

- Bike directions for Europe, North America, AU/NZ
- Opinionated routing: quiet, scenic routes
- Routing engine based on OSRM (heavily forked)
- Website now, iOS app in beta-testing, Android later
- (Not open source!)

Part I

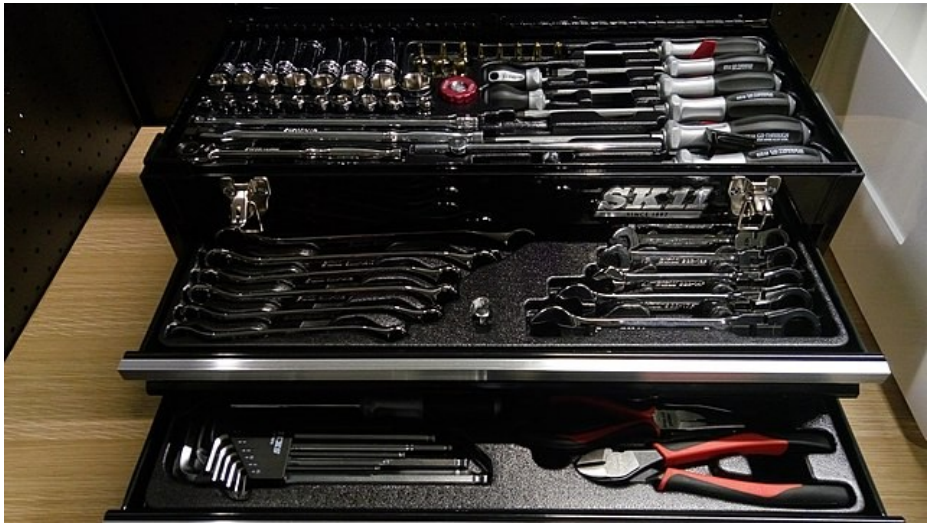
Transforming OSM data in maps

(and routers, and ...)

Where does the OSM data go...

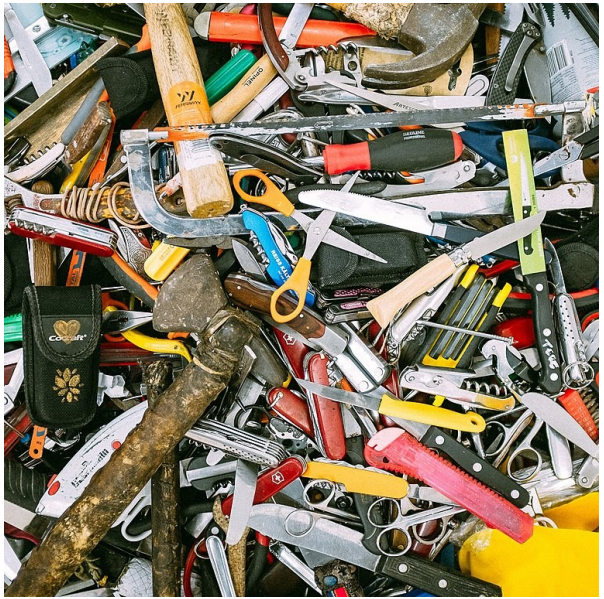
- maps
- search
- routing
- statistics and other exotic uses

What people want



- points, lines, polygons
- properties with documentation

What people get



- nodes, way, relations
- keys, values

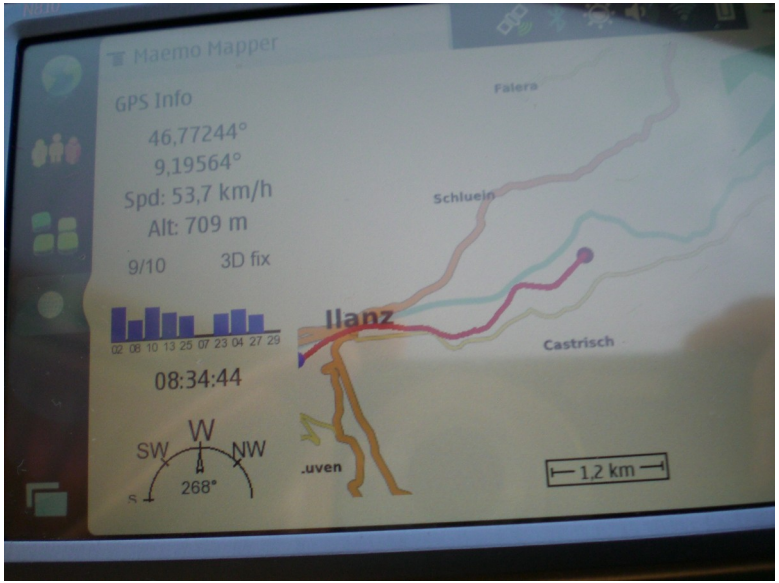
Their solution: go to the preprocessed data

- osm-carto standard style
- OpenMapTiles/Mapbox vector style
- routing standards from OSRM, Graphhopper
- geocoding standards from Nominatim, pelias
- osm2pgsql default style
- shape files from Geofabrik etc.

OSM "core" data

- coastlines, landuse, natural
- boundary=administrative
- highway, railway
- amenity, tourism, shop
- buildings, addr:*

OSM "core" data



Yes, that set of tags from 2008.

OSM "core" data

- Secondary tags are even worse
  - name? Yes.
  - name:it? Maybe.
  - access?
    - vehicle vs. motor_vehicle vs. bicycle vs. foot? Not really.
    - =yes vs =permissive vs =designated vs =official? Definitely not.
  - footway=sidewalk? Nope.

Expanding your data horizon

OSM is not a GIS database.

OSM is a data-mining problem.

Finding tags

- There is no switch2osm for tagging
- OSM’s freeform tagging is not always intuitive to the data consumer – particularly those with a GIS background (who expect tidy thematic layers and a finite number of reliable, complete columns)
- Consequently it’s really hard to ‘correctly’ parse tags unless you’re an OSM expert
- This is a great business model for OSM experts!
- …but may not help your mapping get used

Finding tags

- Taginfo
  - shows actual usage of tags, even correlation
  - does not explain meaning
  - no difference between imports and wide use
- OSM wiki
  - manually curated, explains real-world meaning
  - can be contradictory
  - prone to wiki-fiddling

Finding tags

- discussion forums (mailing lists, forums, etc.)
  - highlight dissenting opinions
  - good for learning about local nuances
  - language barrier
- editor presets (iD, Josm, Vespucci, StreetComplete)
  - ground truth for most mappers
  - machine-readable
  - only machine-readable
- Wikipedia

Following changes

- meaning of tags changes over time (added details, etc)
  - sometimes people quietly “clarify” wiki documentation
  - translations aren’t always in sync
- hard to find places to be notified
  - taginfo is only a snapshot in time, cannot be followed
  - Wiki changes can be tracked but too much noise
  - tagging mailing list also has too much noise
  - editor presets are a maze of twisty little Github issues, all alike
- usually: noticed when the map/router/geocoder breaks

Getting more out of OSM

Handling regional differences

Built-in regional differences

`highway=secondary`

Built-in regional differences

`highway=track`

(Pics CC-BY-SA: Andrew Tatlow · Metrotrekker)

Localised tags

- Germany: motorroad=yes
- UK: designation=*
- Australia: 4wd_only=yes
- New England: highway=path, snowmobile=yes
  (somehow… not actually a path?)

Function follows name

`amenity = cafe`

`cuisine = bistro`

`cuisine = cake`

Pictures by Ninara (CC-by), fotogoocom(CC-by), Crcappuccino (CC-by-SA)

Density differences

place=city
at z7

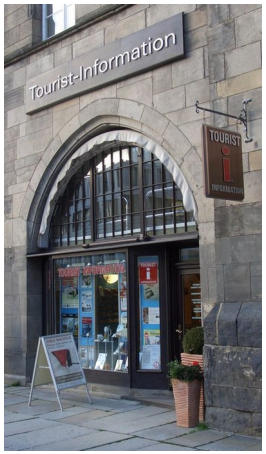
Part ||

How to make most out of mapped data

Unwritten tagging rules

Avoid catch-all primary tags

tourism=information



information=office

information=map

information=guidepost

information=audio_guide

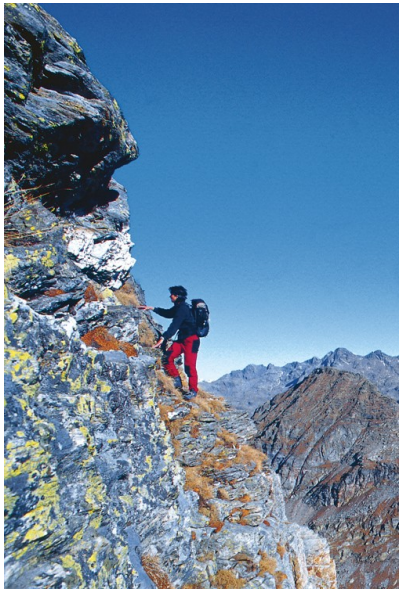
information=trail_blaze

highway=path

`bicycle=designated`

`surface=dirt`
`informal=yes`

`snowmobile=yes`



`sac_scale=difficult_alpine_hiking`

Unwritten tagging rules

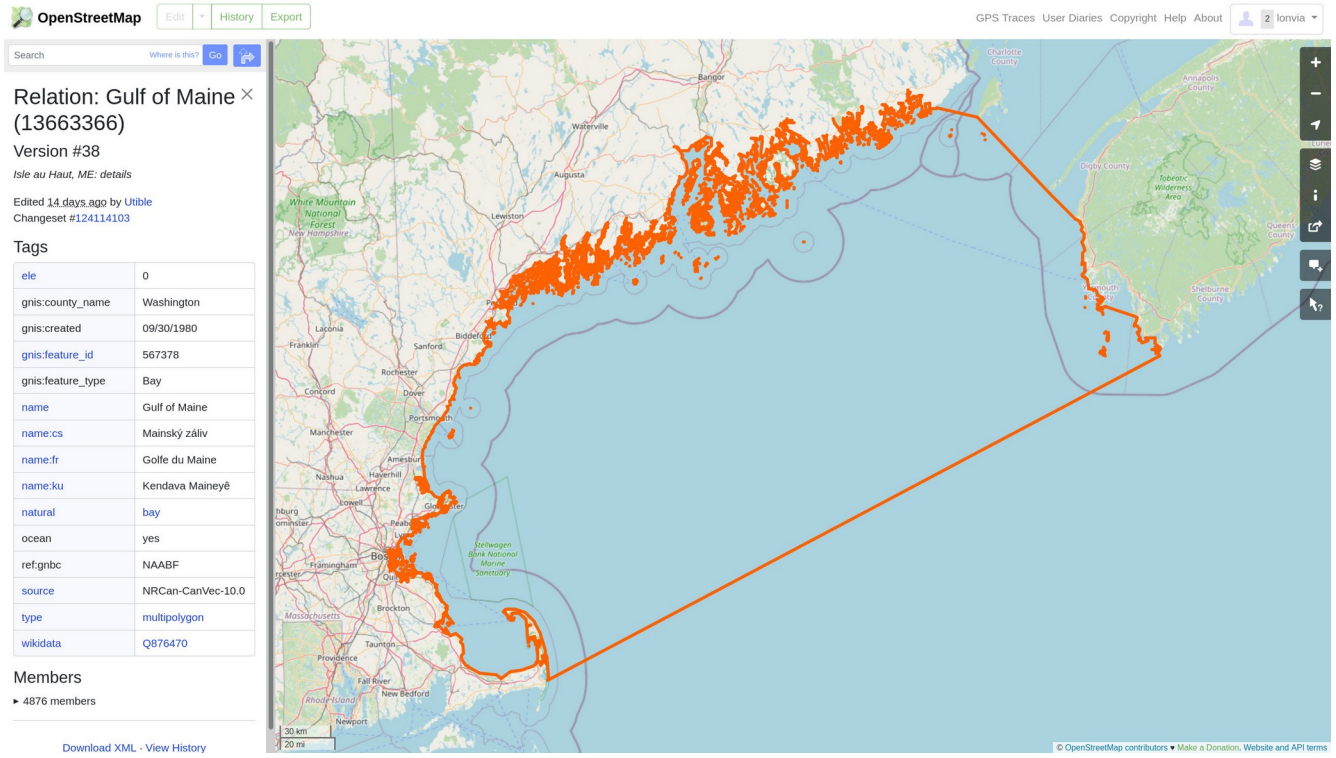
"Do tag for the renderer"

Don' t do that!

Picture by Daniel Capilla (CC-by-SA)

- natural = water
- water = parking_spot

Don' t do that either!



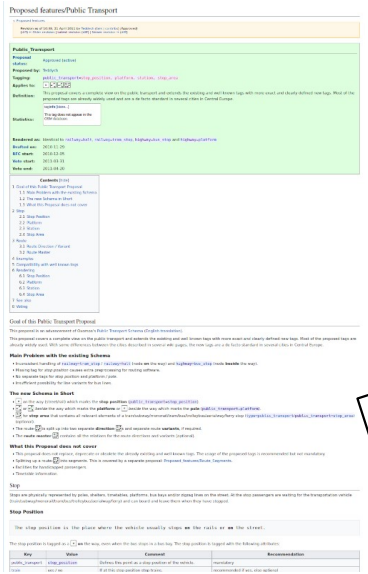
Optimise for the (average) mapper

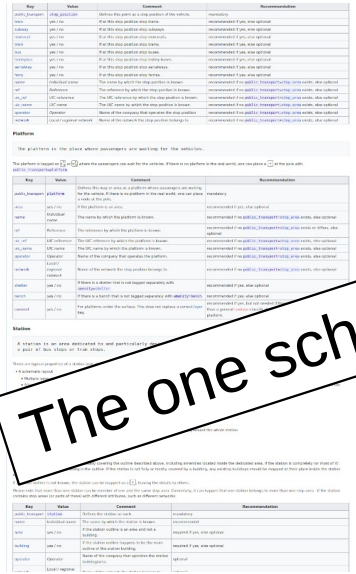
- opt for the tagging with least work for mapper
- complexity should always be optional add-on

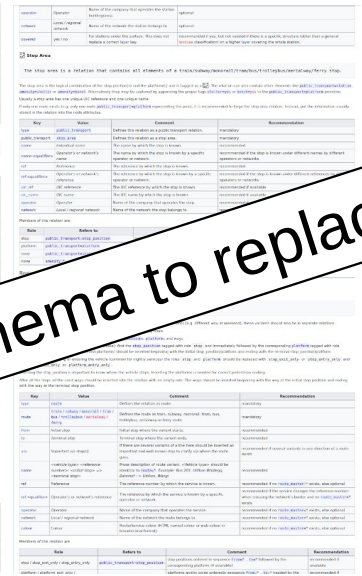
The bad example: solar panels

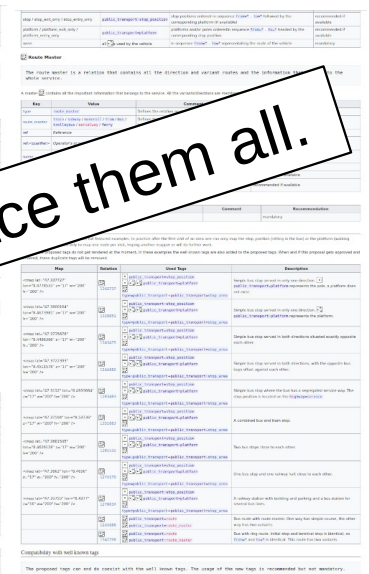
- power=generator
- location=roof *(I guess they don't work so well in the basement?)*
- generator:type=solar_photovoltaic_panel
- generator:source=solar *(wait didn't we just say that)*
- generator:method=photovoltaic *(am I going mad here)*
- generator:output:electricity=yes

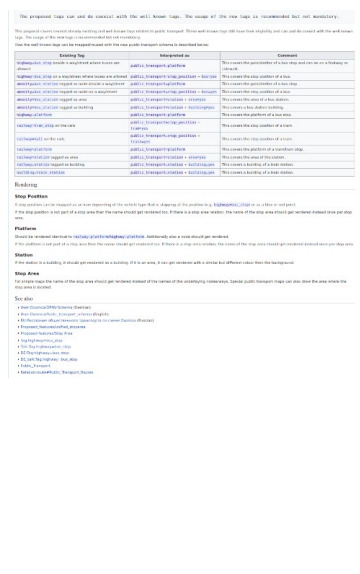
The bad example: PTv1.5





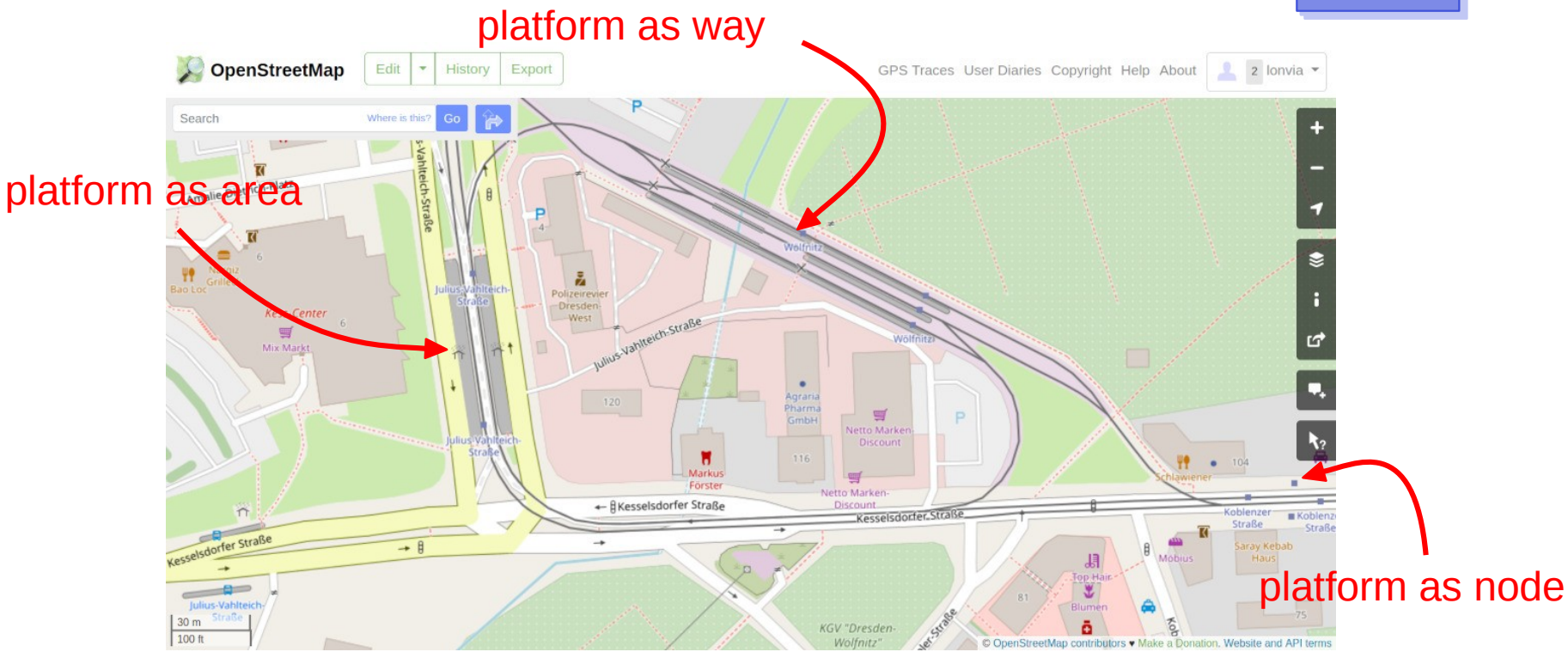




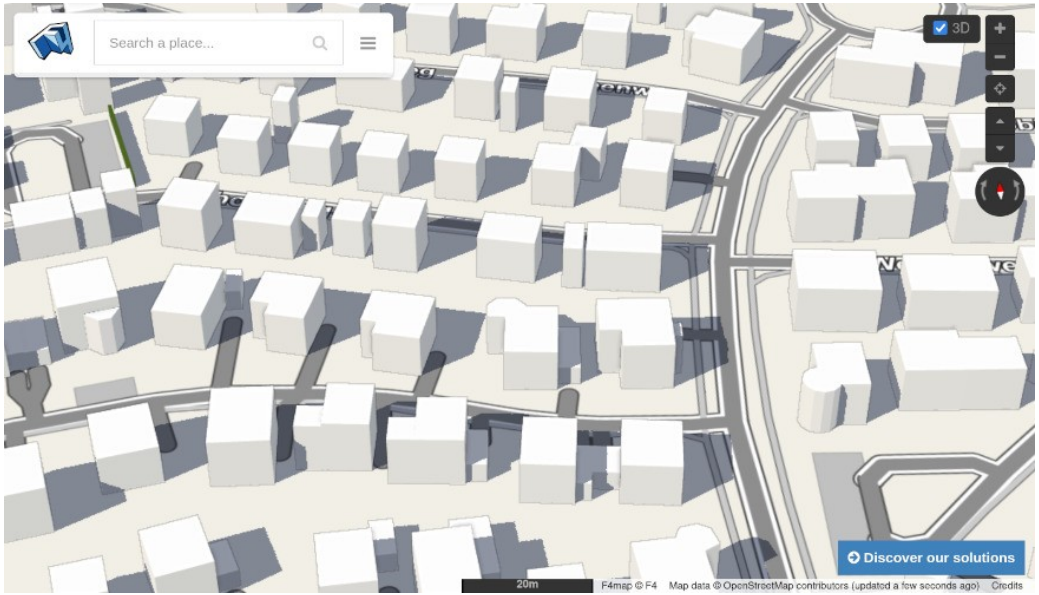


The one schema to replace them all.

The bad example: PTv1.5



The good example: simple 3D



`building = yes`

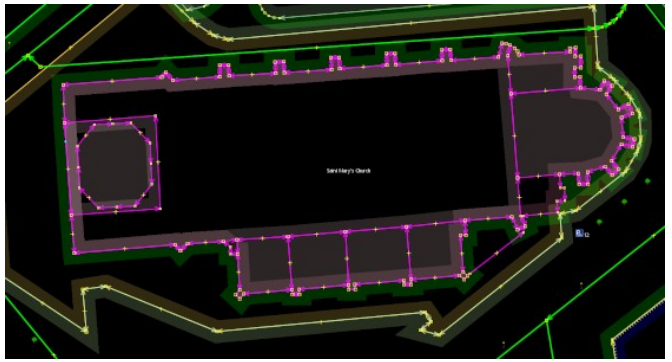
I don't care about 3D buildings.

The good example: simple 3D


```
building = yes
building:levels = 3
roof:shape = hipped
```

This 3D buildings stuff looks interesting...

The good example: simple 3D



3D building expert level achieved.

Checklist: How complex is too complex?

- Does your tagging discourage (new) mappers from editing?
- Is it consistent with OSM tagging idioms?
- Is there a reasonable chance to keep the data up to date?
- Does your tagging need external tool support to be understood? (And is that external tool available for all different platforms?)
- Will your tagging make it harder for “people to use OSM in creative, productive, or unexpected ways”?

Evolving tagging

- build upon existing tagging by adding not changing
  - good example: highway=trunk + motorroad=yes
  - bad example: highway=footway + footway=sidewalk
- degrade gracefully
- don't change tags in the name of tidiness
  - bad example: waterway=riverbank
  - bad example: phone → contact:phone
- be very careful if trying to second-guess data consumers

Be mindful of the preprocessing toolbox

works well

- create points, lines, polygons from single OSM objects
- filtering tags (simple number, boolean, categories)

works less well

- relations beyond routes, turn restrictions and (sane) multi-polygons
- complex processing with more than one OSM object
- data with different interpretations in an area
- lists of values (with semicolon)

New general primitives?

- complex tagging schemes emerge
  - parallel roads and lanes
  - indoor tagging
  - sites
- Do we need data model support?
  - reaching the limits of the centreline model?
  - watch Jochen's talk.

Questions

Sarah – lonvia@denofr.de

Richard – richard@systemed.net · @richardf