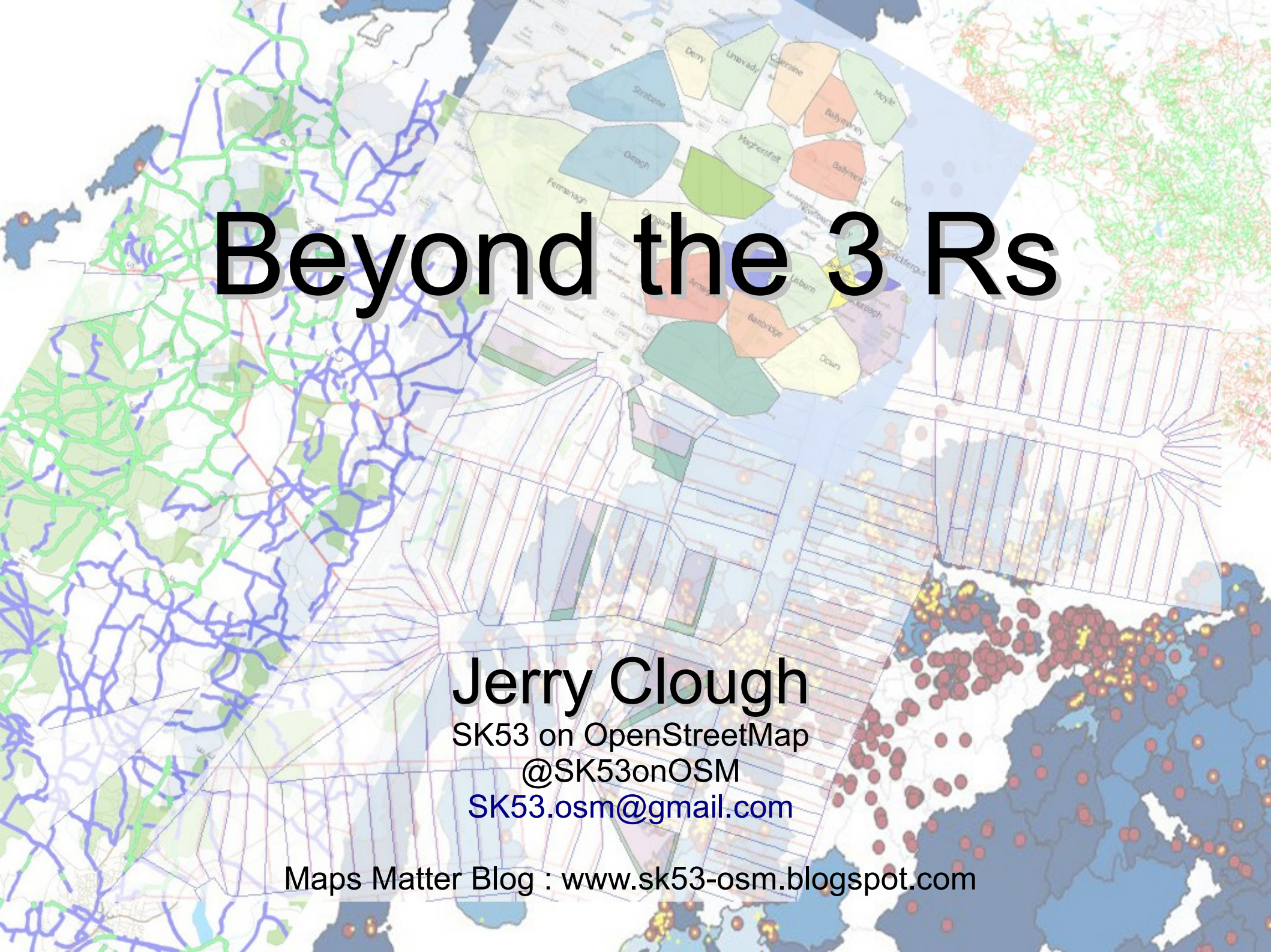


Beyond the 3 Rs

Jerry Clough

SK53 on OpenStreetMap

@SK53onOSM

SK53.osm@gmail.com

Maps Matter Blog : www.sk53-osm.blogspot.com

3 "R"s of Data Consumption

- Rendering
  - Cartography
  - Transformations of data (mkgmap, etc)
- Routing
- Rummaging
  - Search
  - Geolocation

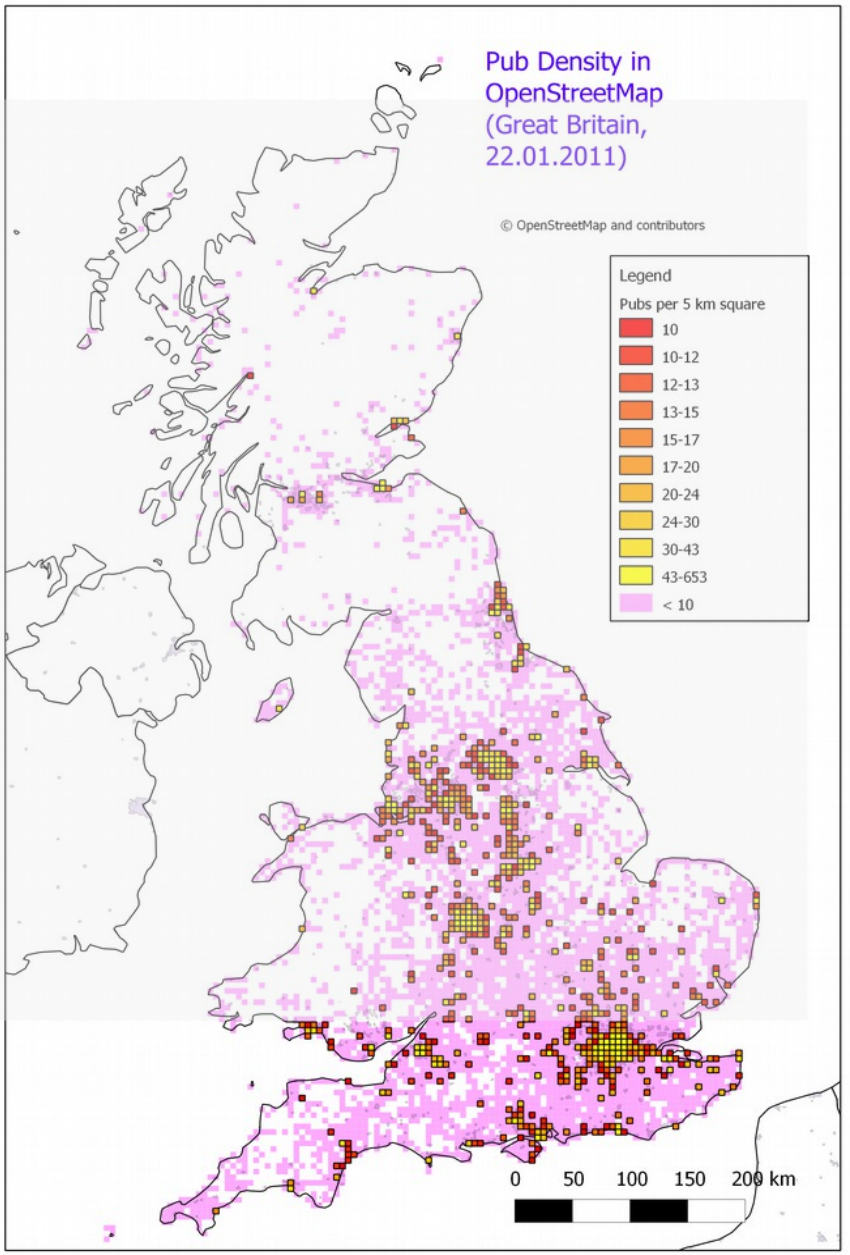
The OSM Trivium

Beyond the 3 "R"s:

- the 4 "A"s
  - Analysis
  - Amusement
  - Auguries (Simulation)
  - Art
- the OpenStreetMap Quadrivium ?

Simple analyses

- Landuse
- Biotopes / Habitats
  - Length of hedgerows per unit area
  - Wetland habitats
- Density of Facilities
  - Pubs / Shops / Banks etc.



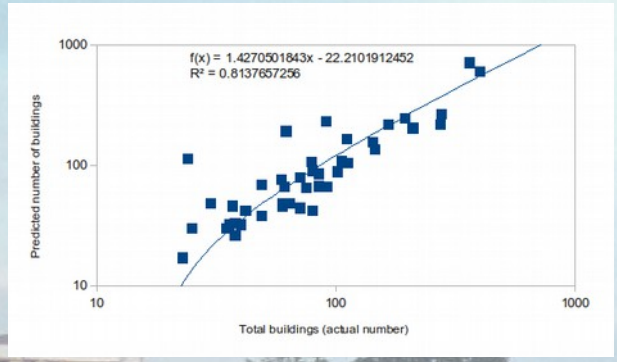
Some 'Interesting' Stats for GB

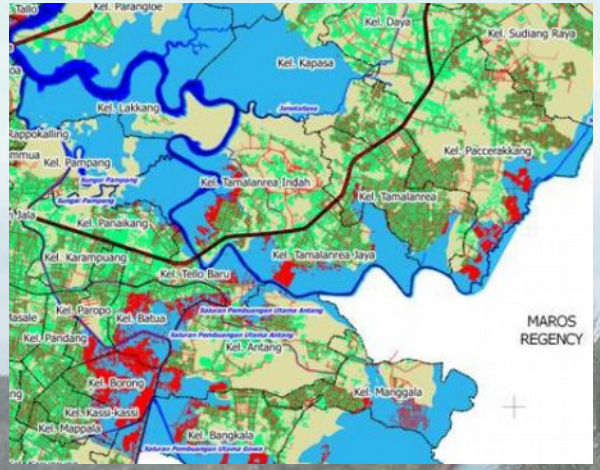
(with apologies to Ordnance Survey)

- Pylons: 58,487 (OSGB: 80,517)
- Post Boxes: 42,742 (93.728)
- Camp sites: 3,192 (8,908)
- Buildings: 1,890,835 (35,397,754)
- Bus Stops: 215,720 (354,099)
- Petrol Stations: 6,299 (7,702)
- Addresses: 27,341,262 (OSGB); 532,886
- Electricity Poles: 94,199 (183, 987)
- Road length: 522,627 km (407,532 km)
- 5 post boxes with Edward VIII cypher
- Only 110 War Memorials
- 847 Fire Hydrants
- 1,378 Real Ale pubs
  - 82 with Real Fires
- 4771 Cycle Parking
- 300 Wildlife Hides
- 5,552 Stiles
- 1,774 Canal Locks
- 2 Knitting Shops

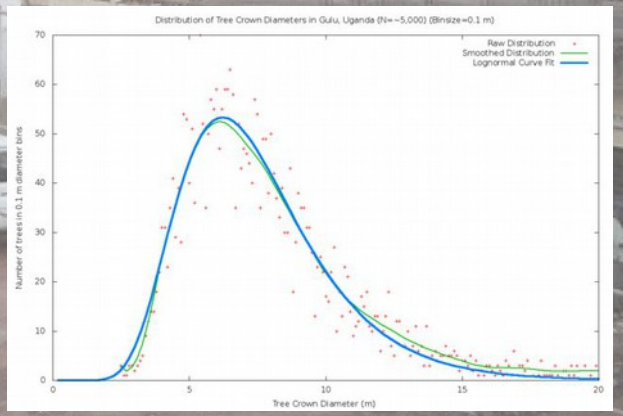
Ordnance Survey figures: /www.ordnancesurvey.co.uk/blog/2013/04/10-fascinating-facts-from-ordnance-survey/
OSM figures (April '13): /taginfo.openstreetmap.org.uk/

Humanitarian Analyses



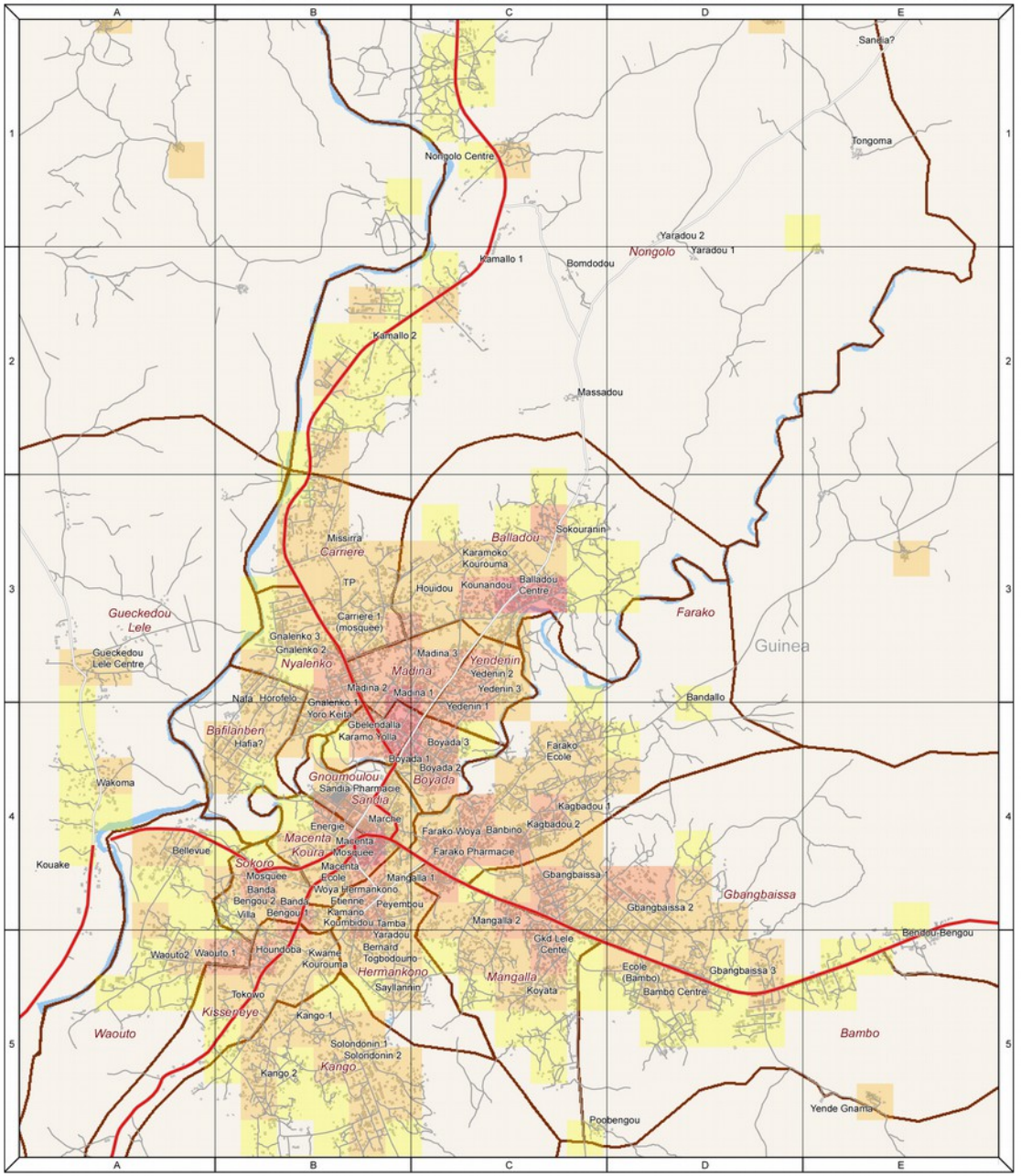


- Mapping trees for fire risk (Gulu, Uganda)
- Calculating population density (Ebola outbreak, West Africa)
- Flooding Risk (Sulawesi)



Further information: Gulu & Ebola— Diary Entries by Andrew Buck (http://www.openstreetmap.org/user/AndrewBuck/diary/21465 & http://www.openstreetmap.org/user/AndrewBuck/diary/17551);
Sulawesi: Maning Sambale (http://pt.slideshare.net/esambale/using-osm-qgis-and-inasafe-for-contingency-plan-in-indonesia)
Background Image: Gulu town centre https://www.flickr.com/photos/29971881@N06/3712921365 CC-BY-NC-SA

Gueckedou - Guinée - Densité de population



Légende

Densité (hab/carré)

- 0 - 80
- 80 - 200
- 200 - 500
- 500 - 800
- 800 - 1219

- Route principale
- Route secondaire
- Autre route
- Limite des quartiers
- Batiments
- Rivière

1 carré : 250 x 250m soit 62500 m² soit 0.0625 km²

N

Sources

Limite des quartiers : MSF-CH
Routes et bâtiments : OSM
Rivière : NaturalEarth

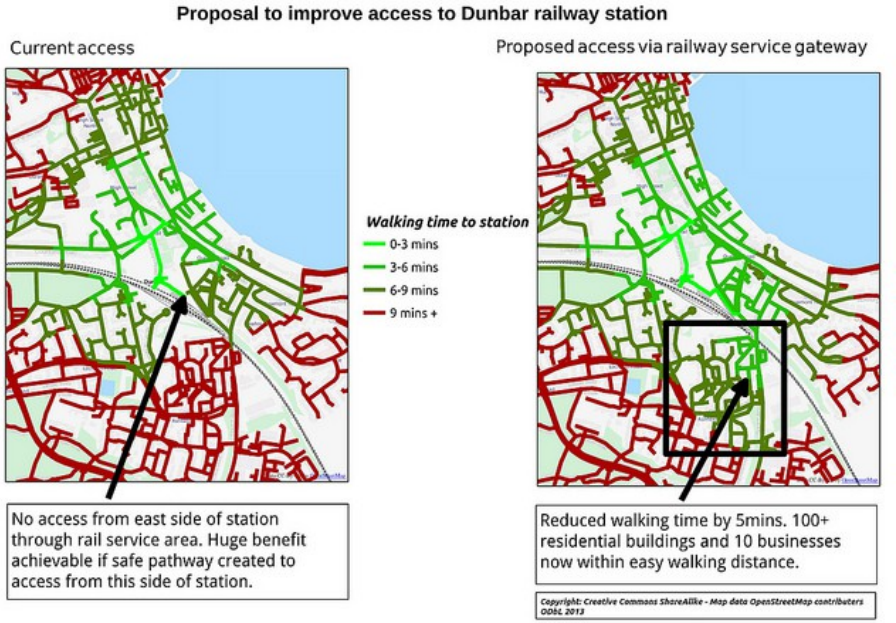
Nom du document : GIN_Gueckedou_Density
Date : 02/04/2014
Realisé par CartONG pour MSF-CH
Dimension pour l'impression : ISO A3

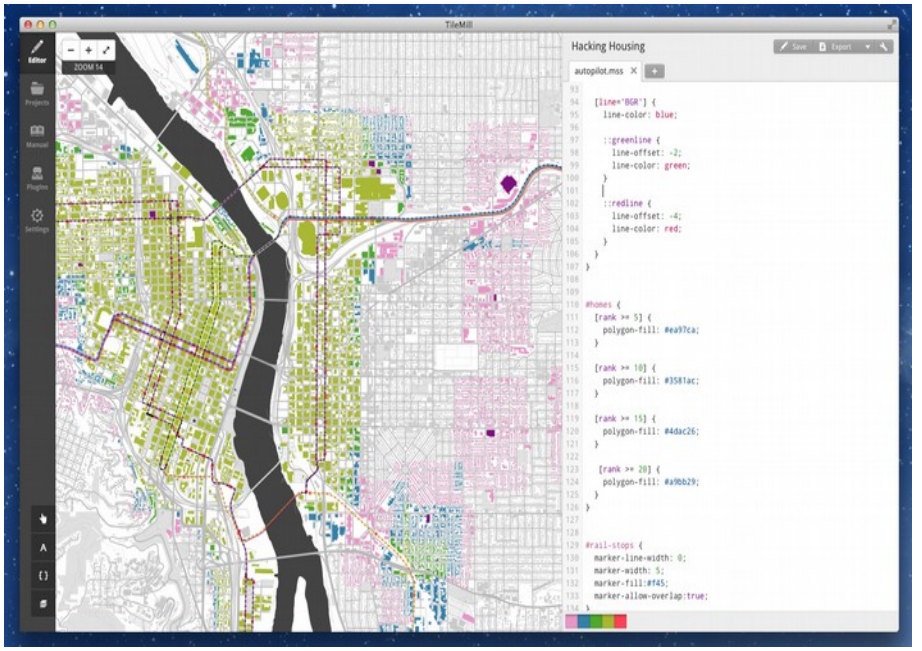
MEDECINS SANS FRONTIERES
ÄRZTE OHNE GRENZEN

0 1 2 Km

Routing-based Analyses

- Dunbar Railway Station
- Tsunami escape routes (Inasafe, Indonesia)
- Where to live in Portland
  (http://dealloc.me/2014/05/24/opendata-house-hunting/)

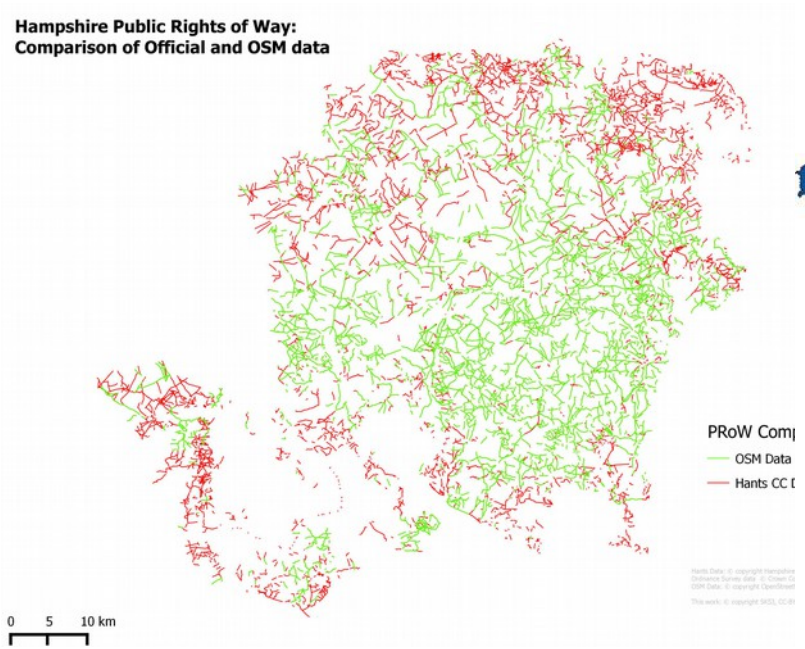


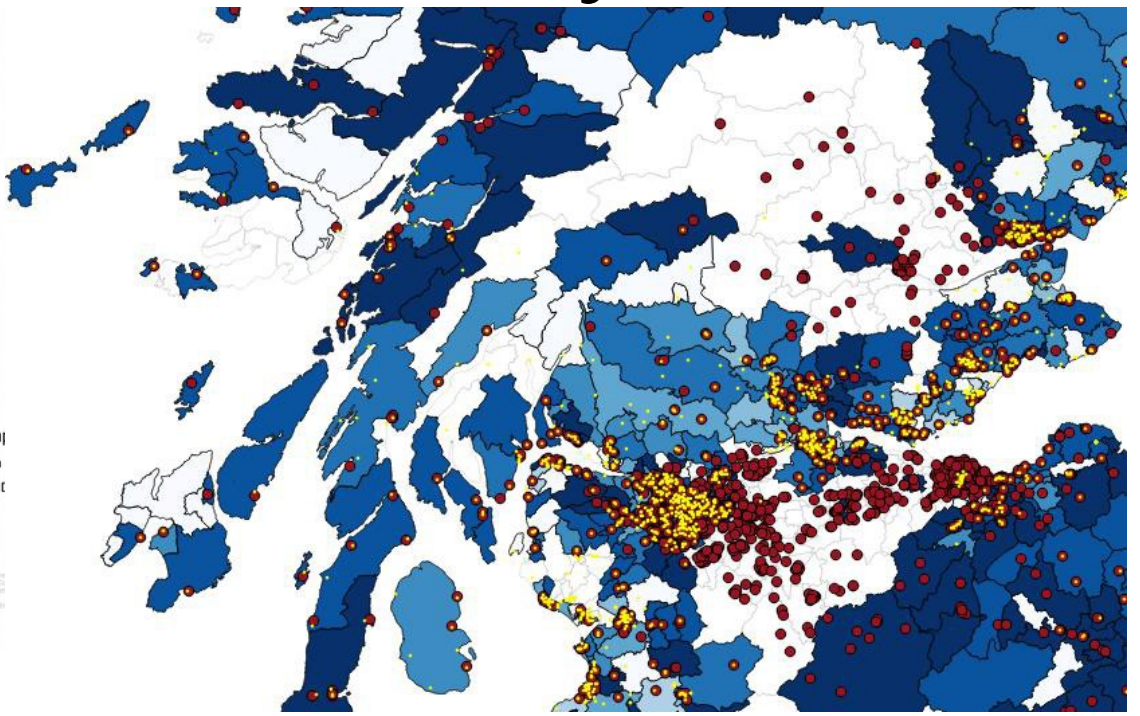


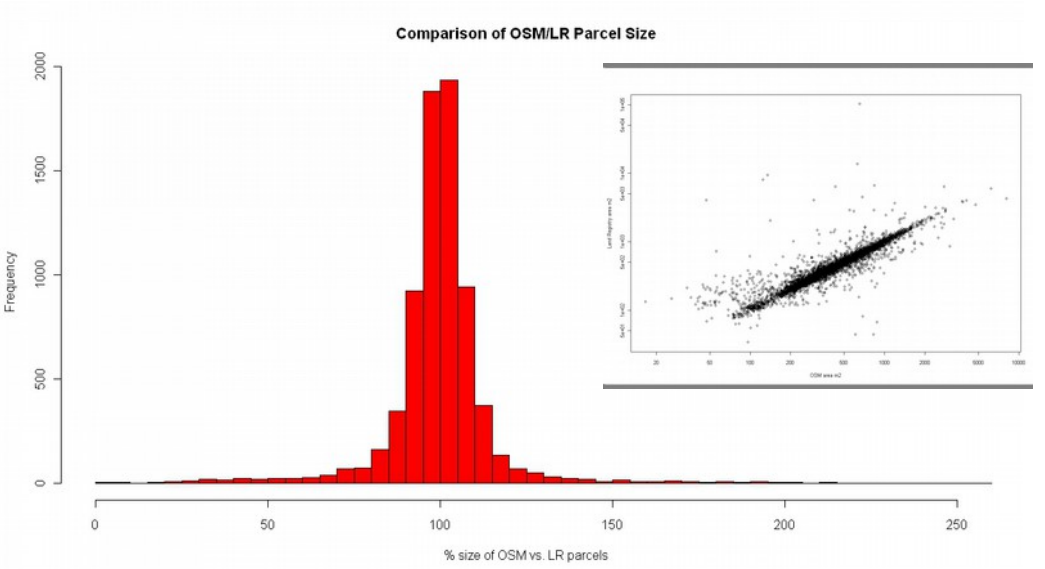
Introspective Analyses

- Estimating degree of coverage
- OSM Quality assessment
  - e.g.,at SotM-EU 2014 talks
    - Alfonso Crisci
    - João Porto de Albuquerque
    - Kirill Bondarenko

Introspective Analyses







What's easy about analysis

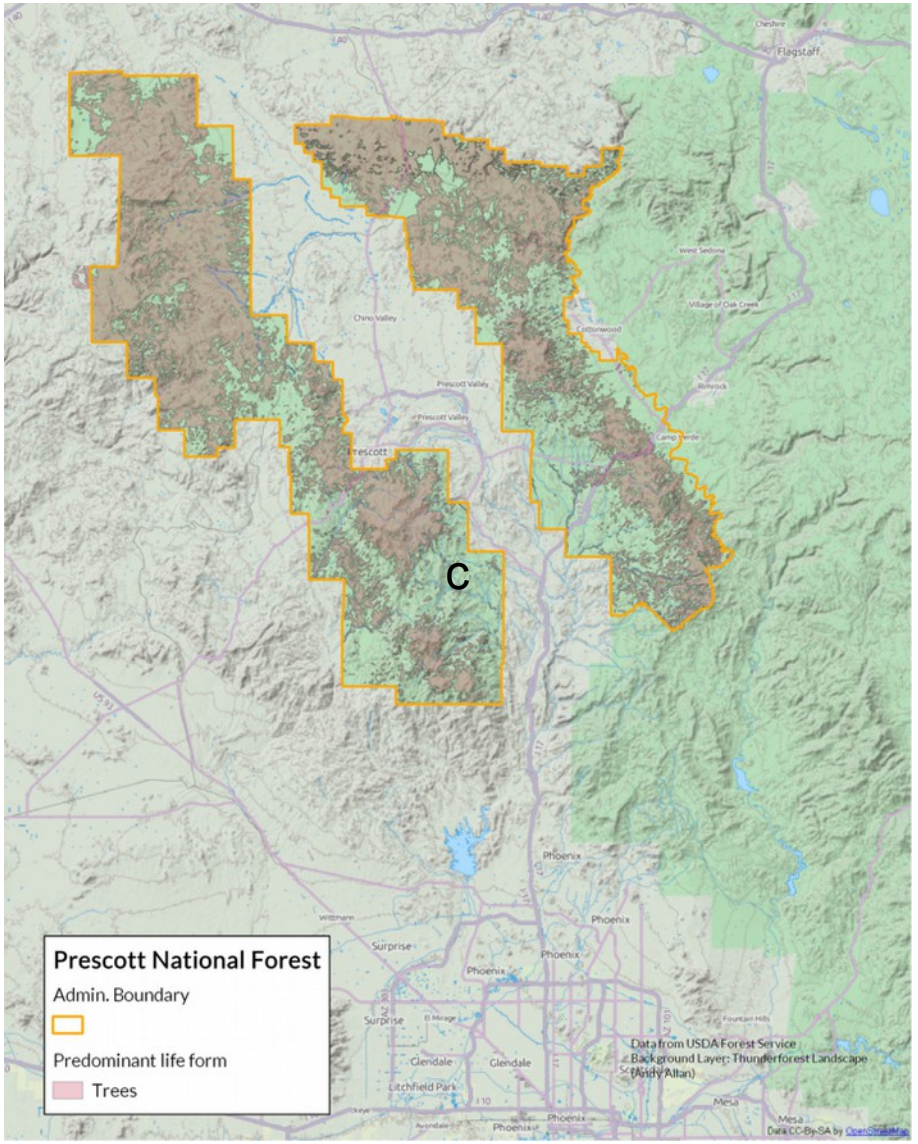
- Big data sets
- Currency: often more up-to-date than other sources
- Ease of transformation of tags
- Wide range of tools

Why is analysis hard

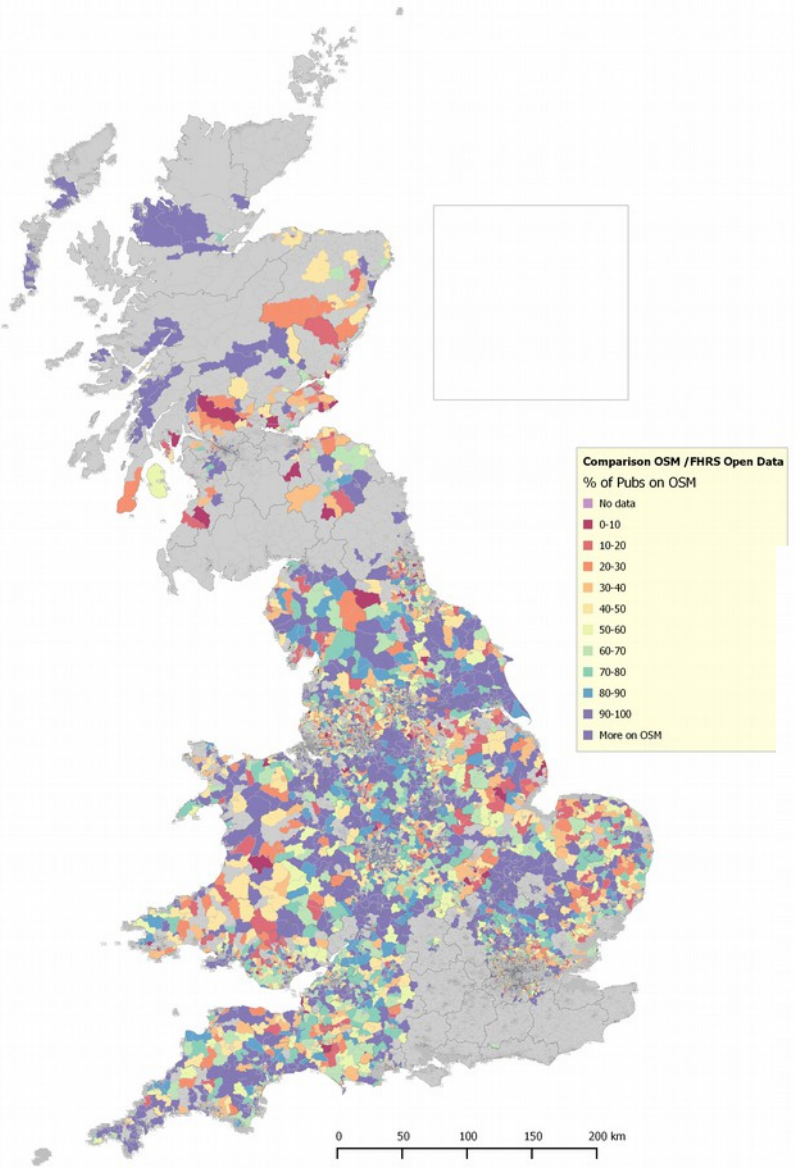
- Tagging semantics
- Incomplete data
- Inconsistent data
- Areas (polygons)
- Unifunctional data
- Generalisation
- Emergent data

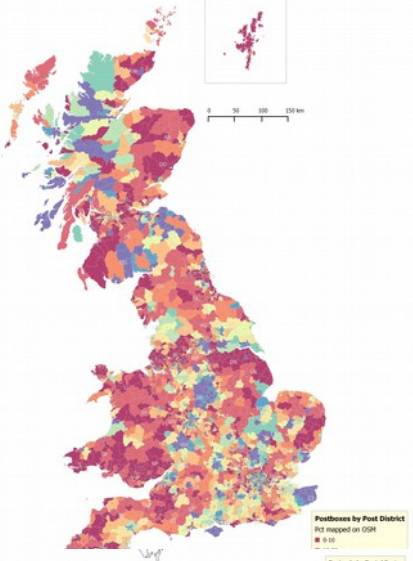
Why is analysis hard : tagging

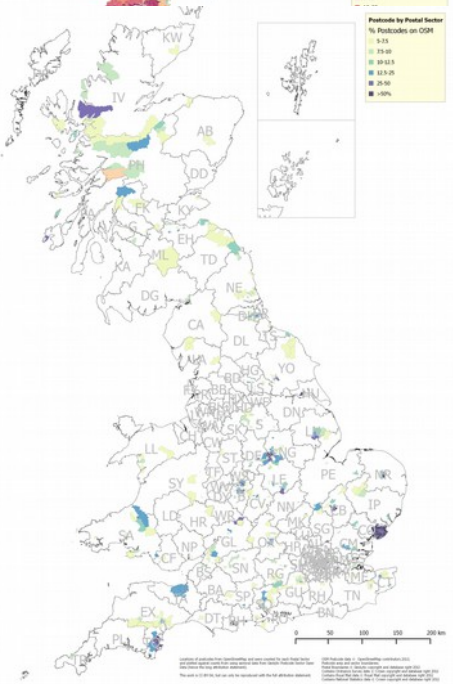
- Tagging for the renderer
- Semantic degradation of tags
  - Hamlet for industrial districts of cities in US
  - Forest for National Forests in US
  - Ski Areas as leisure=recreation_ground
- Over generalisation
- Under generalisation
- Long tail

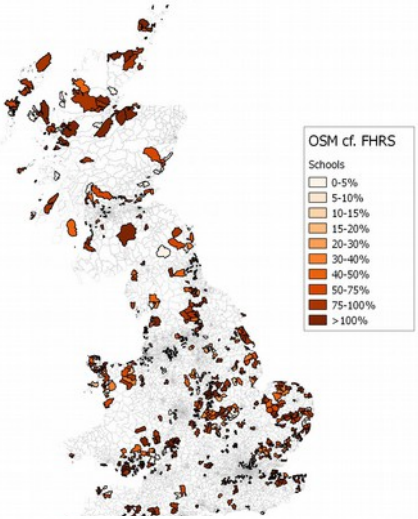


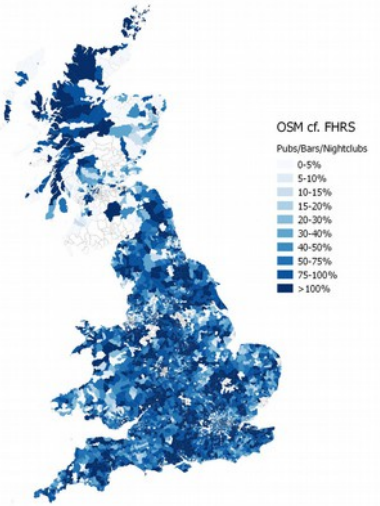
Incomplete Data











Locations of postboxes from OpenStreetMap and from Royal Mail data obtained in an FOI request in 2009 (see http://edwardbetts.com/postboxes/) were counted for each Postal district and plotted using districts aggregated from Geolytix Postcode Sector Open Data (hence the long attribution statement).

Inconsistent data

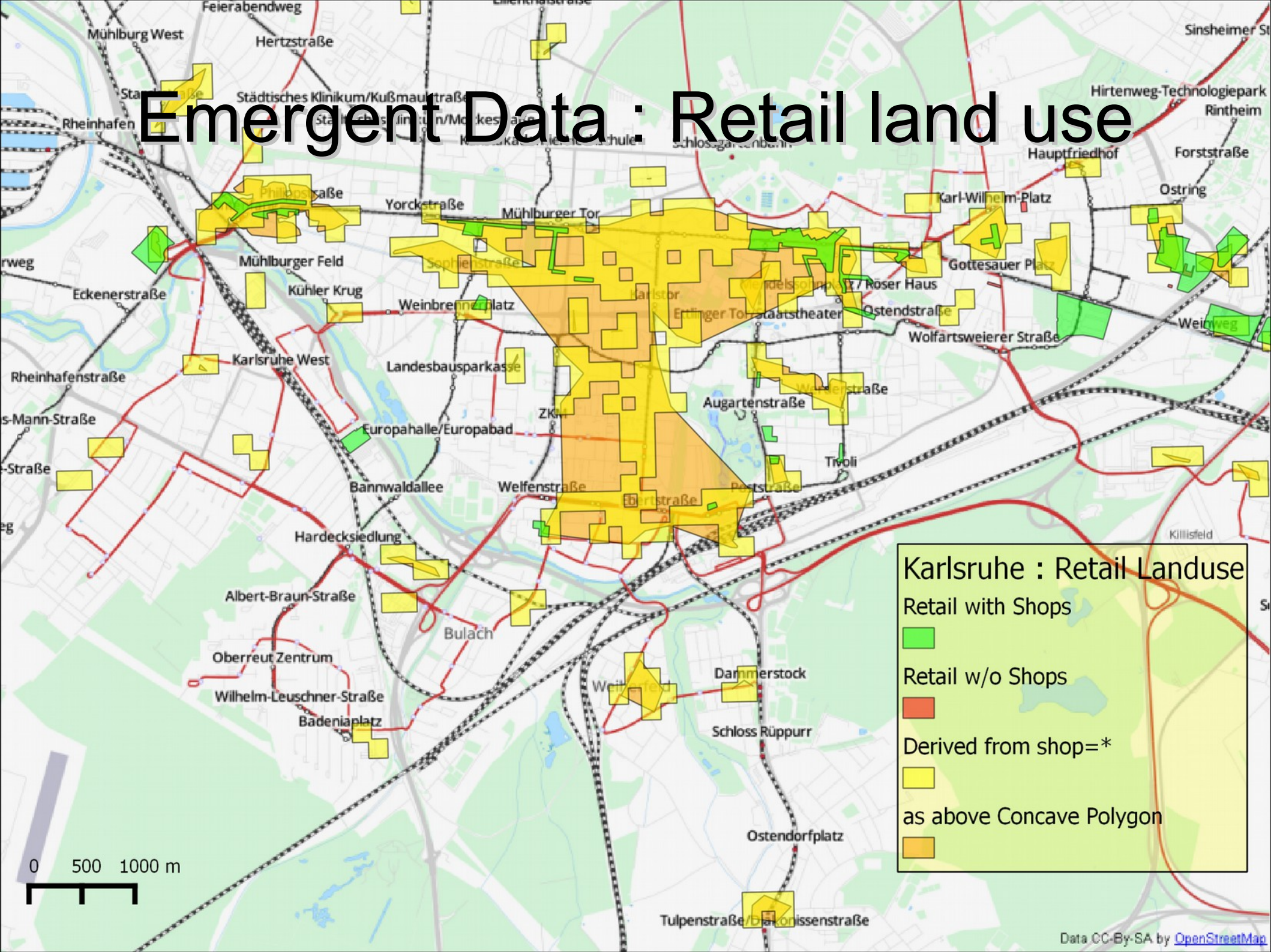
- Some examples
  - Sports facilities
  - Shop types
    - Games Workshop
    - Accessorize / Claires
- **NOT** multiple tags (synonymy) for same thing
  - Relatively easy to post-process

The Problem with Polygons

- No Area primitive in OSM
- Overlapping polygons
- OSM
  - Broken polygons
  - Intersecting polygons
  - osm2pgsql
- In QGIS
  - Render OK
  - Geometry Operations fail
- Essential tool: `cleangeometry` PostGIS function (SOGIS)

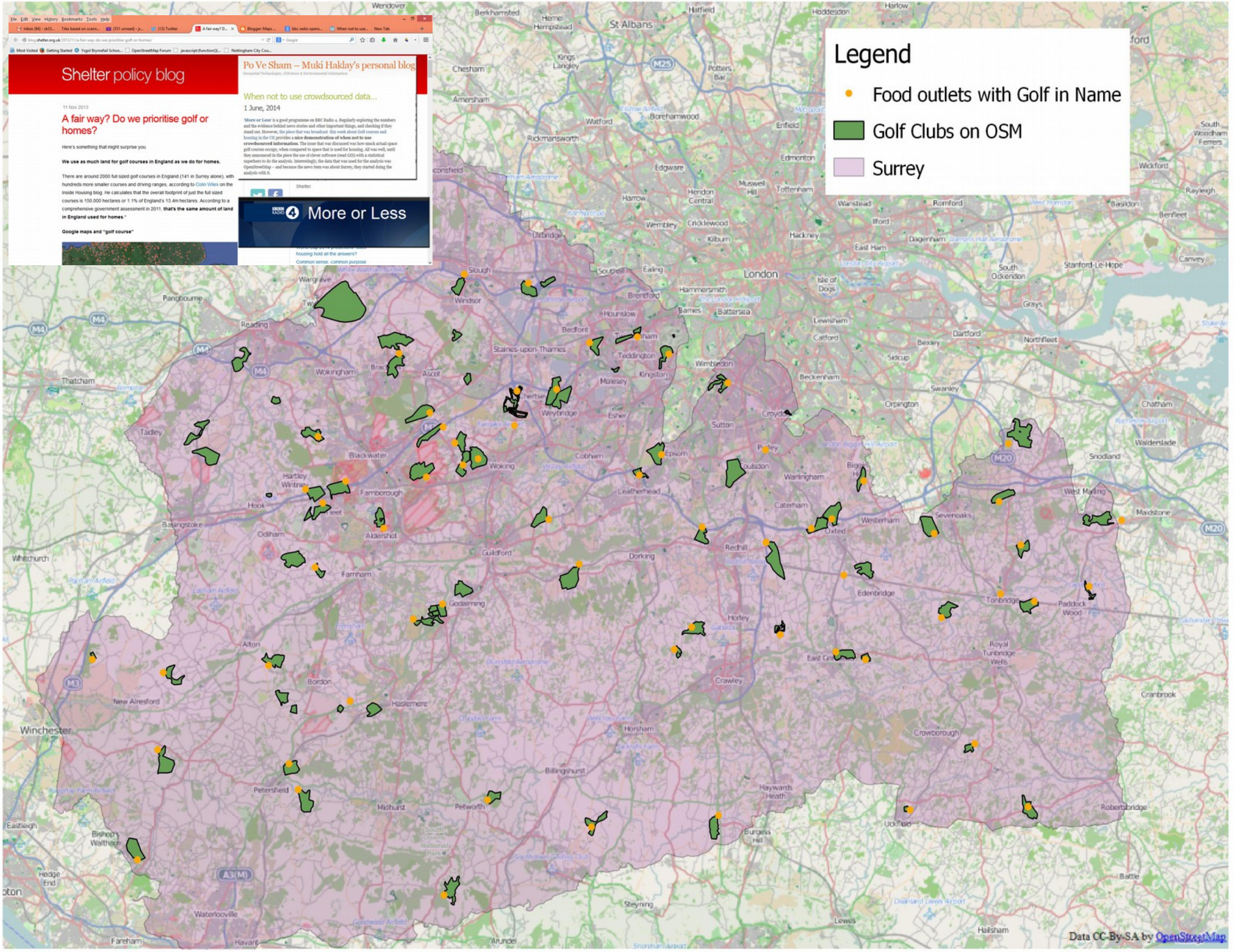
http://www.sogis1.so.ch/sogis/dl/postgis/cleanGeometry.sql

Emergent Data : Retail land use

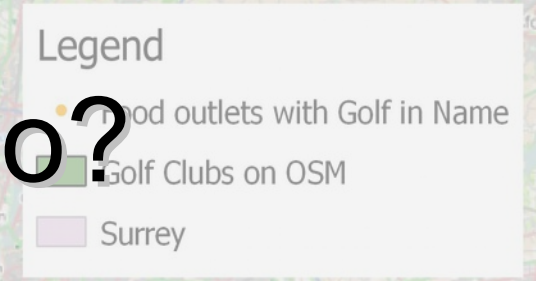


Emergent Data: Built-up areas

- No tag or collection of tags identifies built-up areas
  - Industrial can refer to quarries etc.
- Landuse is rarely comprehensively mapped
- Landuse imports usually inaccurate
  - CORINE
  - Georgia, New Jersey


Shelter policy blog
A fair way? Do we prioritise golf or homes?
Here's something that might surprise you.
We use as much land for golf courses in England as we do for homes.
There are around 2000 full sized golf courses in England (141 in Surrey alone), with hundreds more smaller courses and driving ranges, according to Colin Wiles on the inside Housing blog. He calculates that the overall footprint of just the full sized courses is 150,000 hectares or 1.1% of England's 13.4m hectares. According to a comprehensive government assessment in 2011, that's the same amount of land in England used for homes.
Google maps and "golf course"
Po Ve Sham – Muki Haklay's personal blog
When not to use crowdsourced data...
1 June, 2014
More or Less
housing hold all the answers?
Common sense, common purpose
Legend
Food outlets with Golf in Name
Golf Clubs on OSM
Surrey
Data CC-By-SA by OpenStreetMap

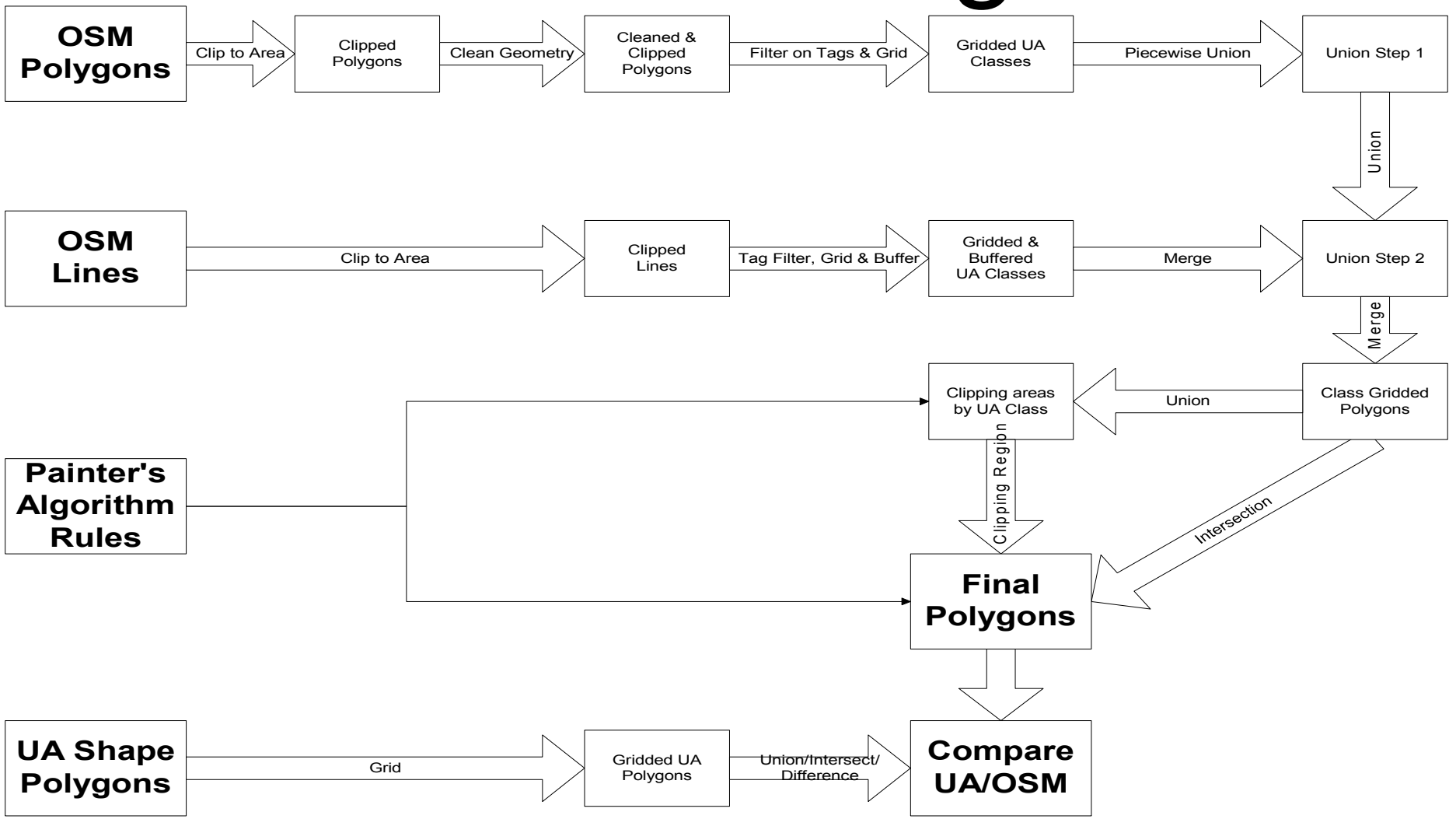
What can we do?



- Plenty of Omphaloskepsis
  - Contemplating OSM's navel
- Be aware of other use cases
  - Aka "DONT TAG FOR THE RENDERER...", router, search engine,
- Develop better techniques for generalisation
- **DON'T** degrade tag semantics
  - Rules on wiki often contribute to this
- Recognise that some things which are implicit in the data must be mapped explicitly

Supplementary Slides

PostGIS Polygon Processing



Long-tail of shop tags