



NUI MAYNOOTH

Ollscoil na hÉireann Má Nuad

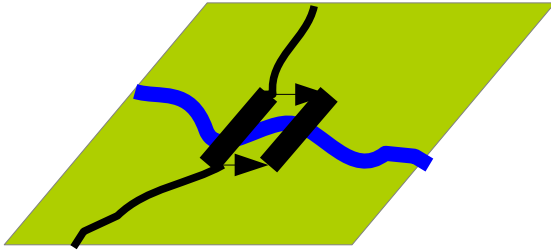
VGI-Specific Quality Analysis: The OpenStreetMap Example

by BŁAŻEJ CIEPŁUCH,
Peter Mooney, and
Adam C. Winstanley

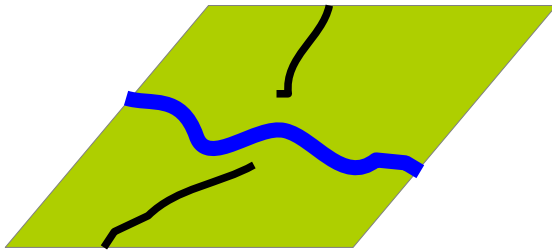


Environmental Protection Agency

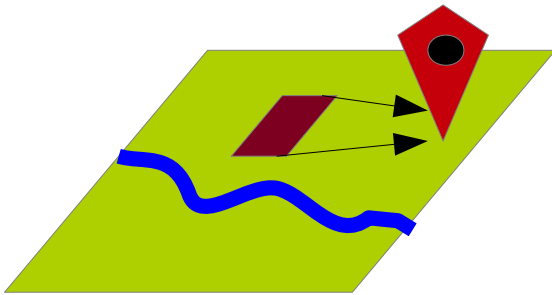
Quality in Spatial Data is expressed in several different ways



- Positional Accuracy (Is a feature in the correct place on map?)



- Completeness (Are all the features of a given class represented)



- Consistency (Representation Rules:
Are Post Offices represented as POI or as polygons?
How consistent is the attribute meta-data?)

Several services exist for checking these quality issues on OpenStreetMap databases



Feel free to put the modifications you would like to see on [OpenStreetMap](#) on the map.

Do NOT use licensed data like paper maps, Google Maps, etc. Use only your knowledge of the reality or public domain data.

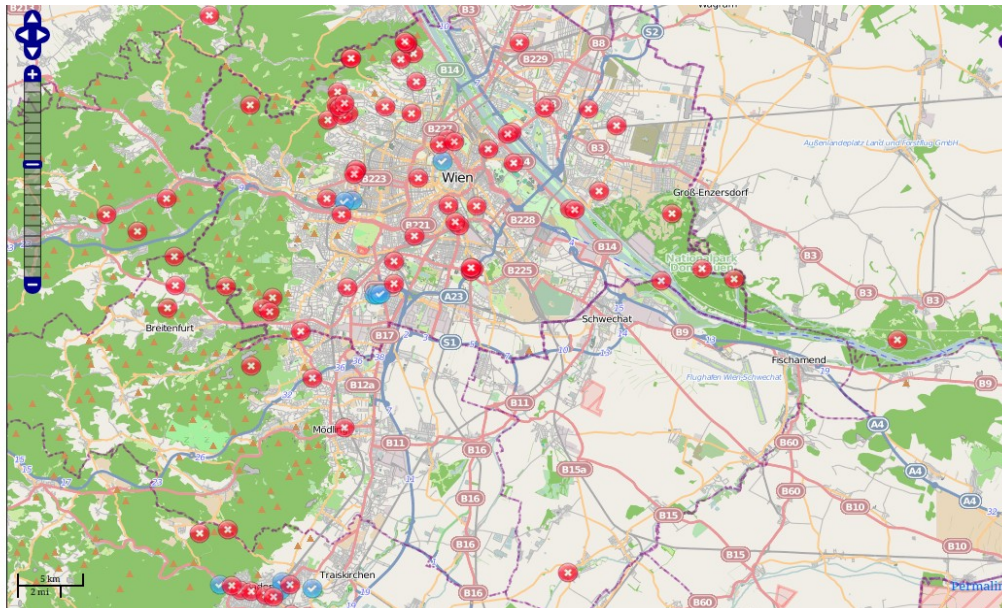
According to the OpenStreetMap license, the data that you add on the map will be licensed [CC BY-SA](#).

To add a bug, click at the desired spot in the map.

[More information](#)

current view

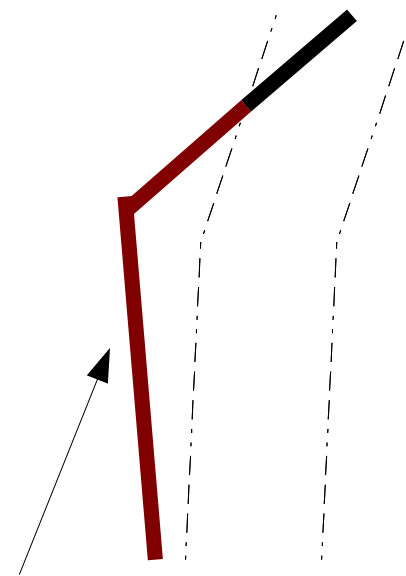
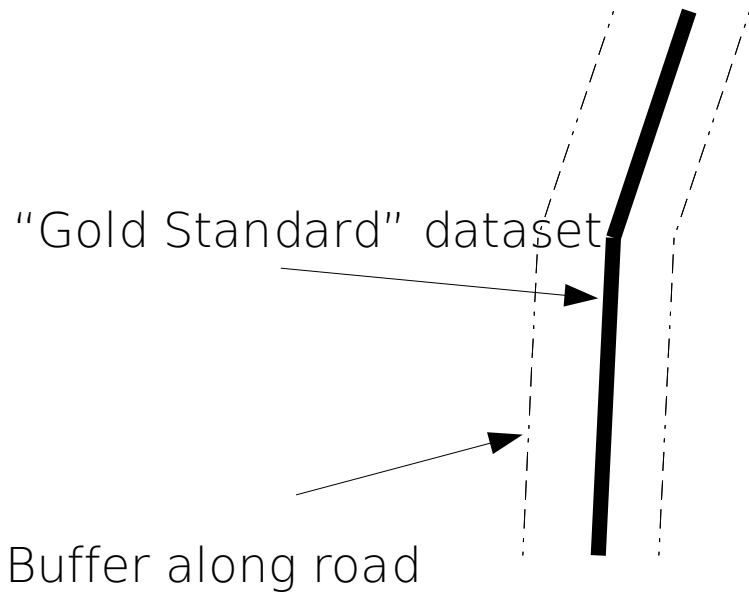
- [Permalink](#)
- [OpenStreetMap.org](#)
- [Geofabrik Map](#)
- [RSS feed](#)
- [GPX export](#)
- [GPX export \(open bugs\)](#)



- OpenStreetBugs
- Tag INFO
- QualityStreetMap
- MapDust
- Other – ad-hoc tools

In the literature – many of the geometric accuracy comparisons of OSM and other datasets focus on road and road network representation

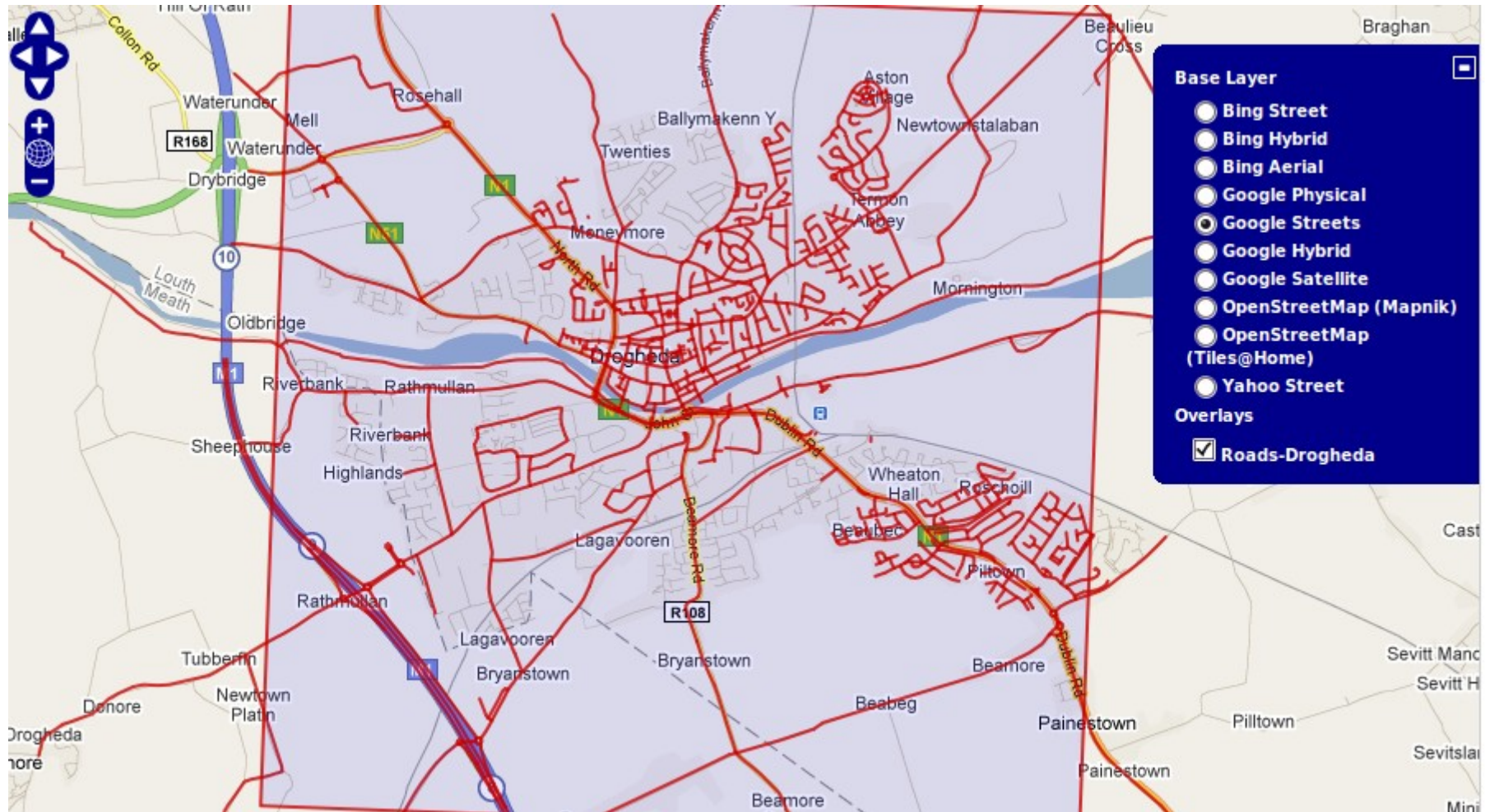
These methods use the classical buffer technique against a “gold standard” dataset



Compute how well the OSM road fits into the buffer

Haklay (2010)
Over et al (2011)
Ludwig et al, (2011)

In Ciepluch et al (2010) we carried out a visual comparison of Google, Bing, and OSM in Ireland



No Vector Data was available from Google or Bing

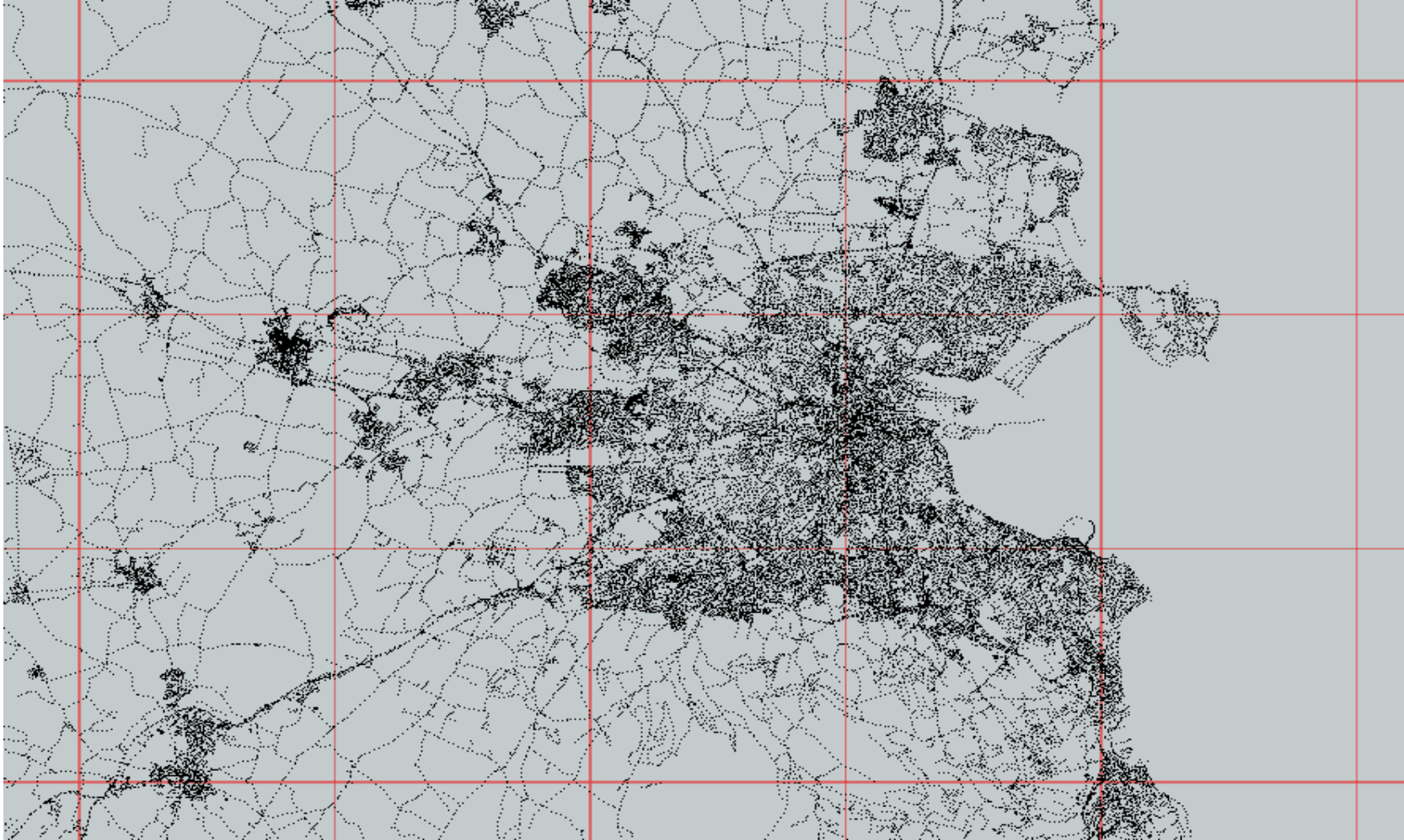
Visual comparison - and manual counting of roads
Extensive use of local knowledge.

Results show that all three map providers offered significant coverage of 5 major cities

	Bing	Google	OSM
Ennis	98%	99%	99%
Drogheda	99%	99%	87%
Maynooth	95%	97%	100%
Waterford	93%	100%	99%
Dublin	97%	98%	98%

Google and Bing for 2010 often did not have an update cycle fast enough to capture new roads or road upgrade/redesignation work. OSM capture this information. Of course the visual comparison method does not scale and is unsustainable

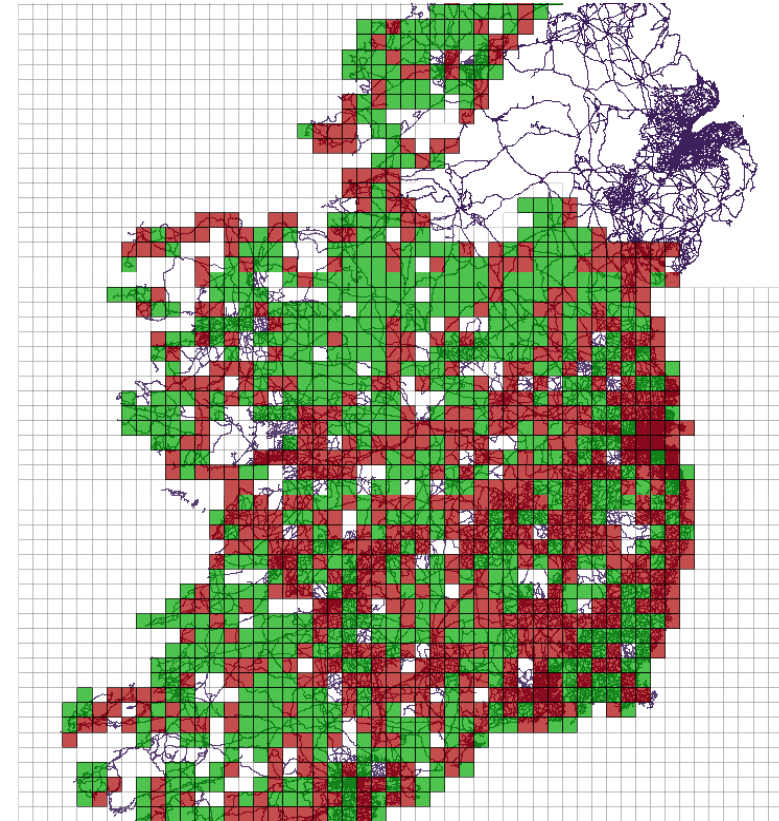
We performed our own “gold standard” comparison for OSM with Ordnance Survey Ireland (OSI) roads



We compared roads in OSM and OSI using a grid-based approach with varying grid-cell sizes

The greatest advantage of the OSI ("gold standard") was geographical coverage and completeness

```
input : Grid cell size, coordinates of bounding rectangle (in UTM) -  
      (NWEasting,NWNorthing) and (SEEasting,SENorthing)  
output: A spatial database table with polygons representing each  
      grid cell in UTM coordinates  
currX ← NWEasting;  
currY ← NWNorthing;  
GRID ← 5000;  
while currY ≥ SENorthing do  
  x1 = currX, y1 = currY;  
  while currX ≤ SEEasting do  
    Assign other vertices of cell;  
    x2 = x1 + GRID, y2 = y1;  
    x3 = x2, y3 = y2 - GRID;  
    x4 = x1, y4 = y1 - GRID;  
    Store Polygon in Database;  
    POLYGON(x1y1, x2y2, x3y3, x4y4, x1y1);  
    Move to next grid cell along easting;  
    x1 = x1 + GRID, currX = x1;  
  end  
  Move south currY = currY - GRID;  
  Reset easting currX = NWEasting;  
end
```



Secondary Roads - OSI vrs OSM



OSM

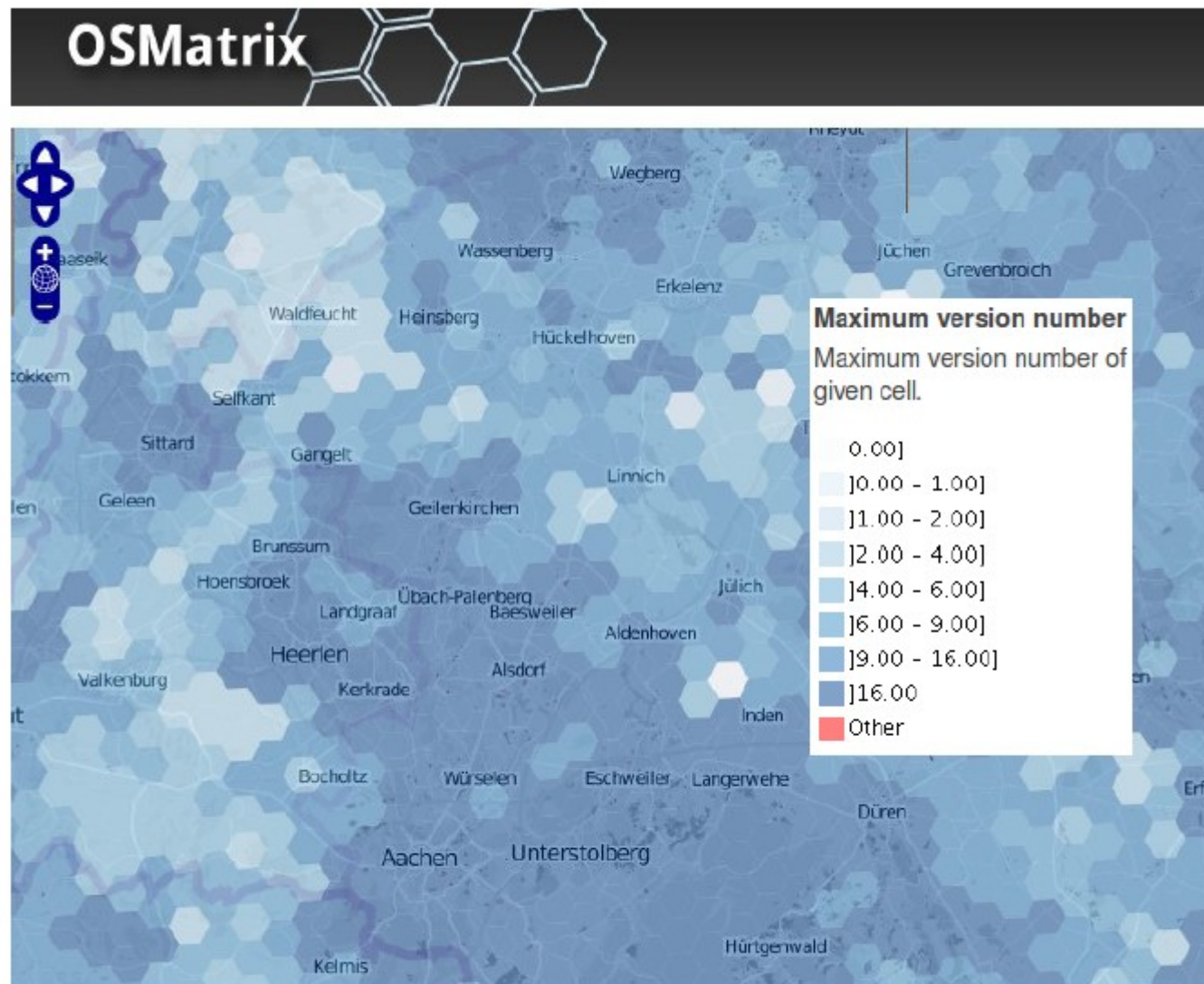


OSI

Colour of cell represents the dataset with greatest coverage

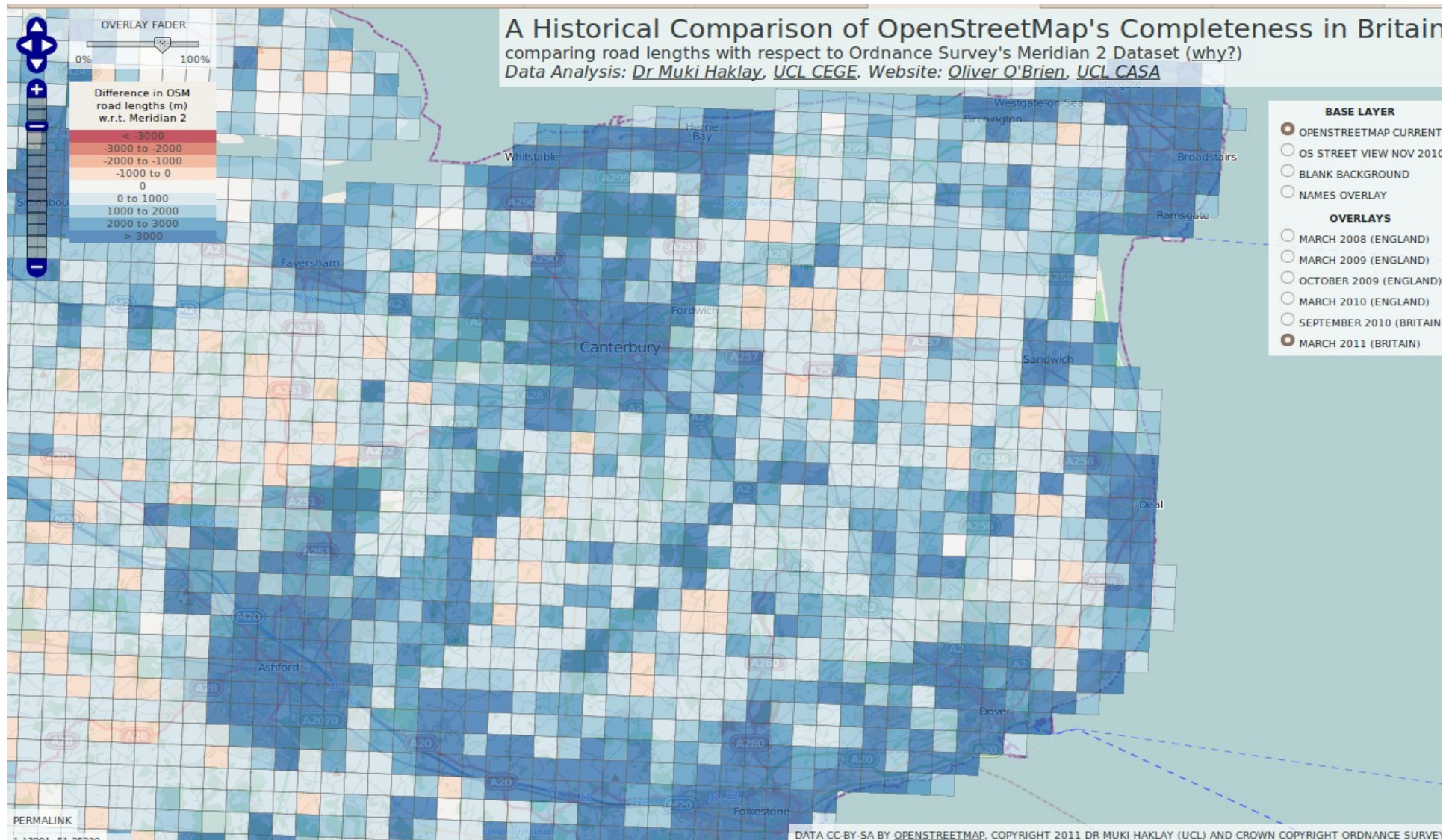
Similar Grid-based Methods are being implemented by other researchers.

- The OSMatrix by Rioch et al (2011)



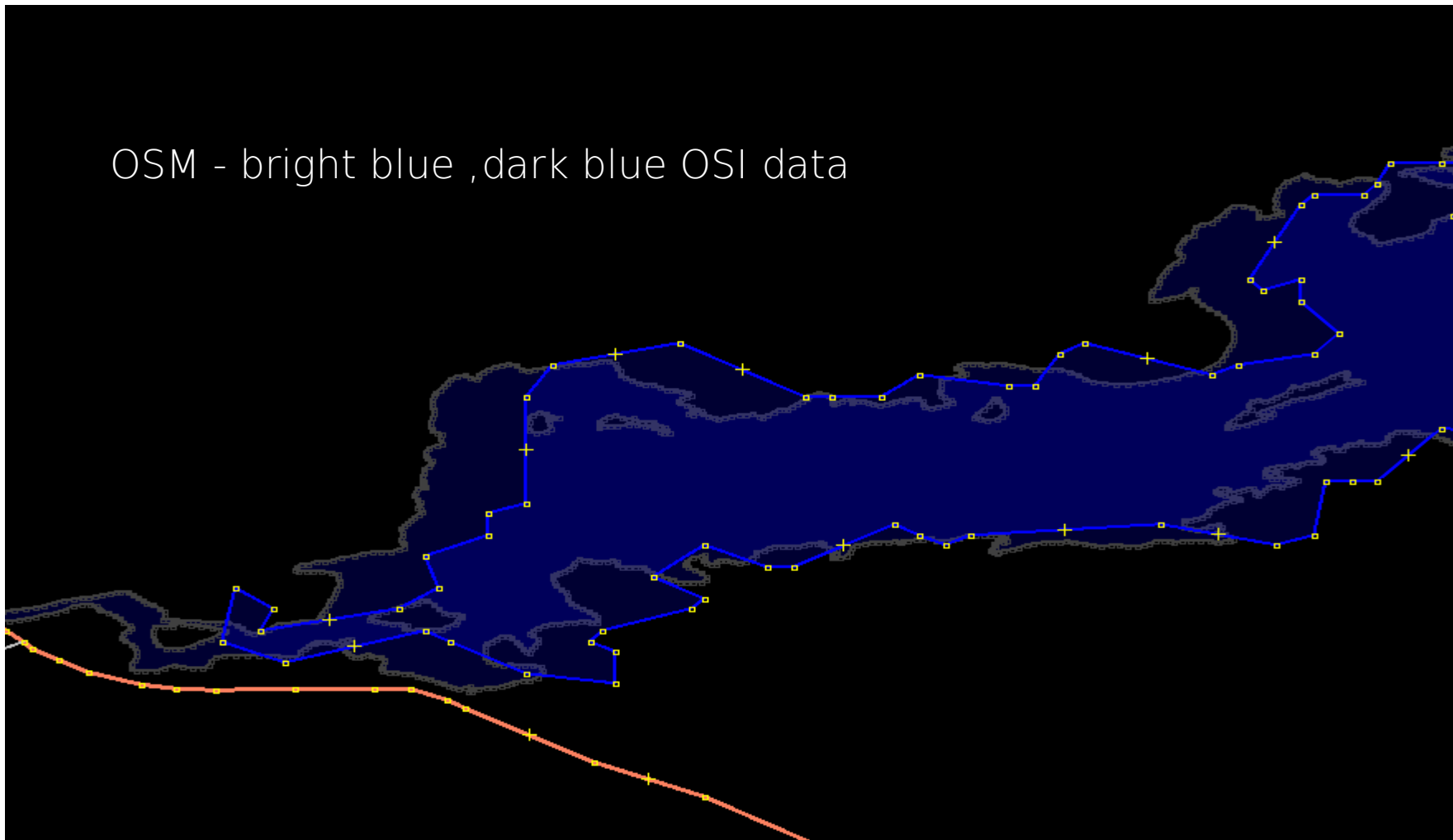
<http://koenigstuhl.geog.uni-heidelberg.de/osmatrix/#>

The work of Haklay et al (2011) has provides a very recent picture of OSM coverage and accuracy in the UK



- Comparison of OSM with Meridian in UK

We also carried out a study on comparison of OSI
Lakes and OSM Lakes



Mooney et al (2010 - ACM-GIS)

ISSUES: Coverage (very few lakes in OSM), Accuracy and representation a problem
Geographical Isolation of many lakes a possible reason for poor OSM representation

Difficult to access features are an issue in OSM Ireland

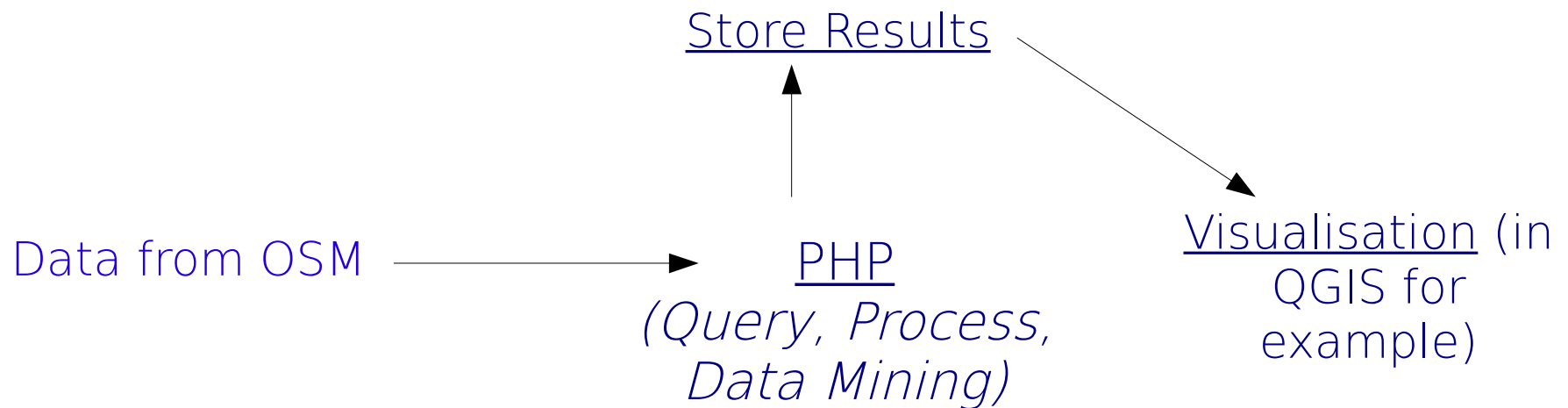


Dataset	Rail Lines	Disused Rail	Industrial Rail
OSM	1878	59	5
OSI	2092	1449	499

Table 3: Number of Kilometers of Railway lines in both OSM and OSI dataset

Our Research Question: Without using the classical methods of Quality – can we investigate OSM data “in isolation” - without comparison to other datasets

Our data workflow – from the previous work – could be re-engineered to provide a basis for analysis

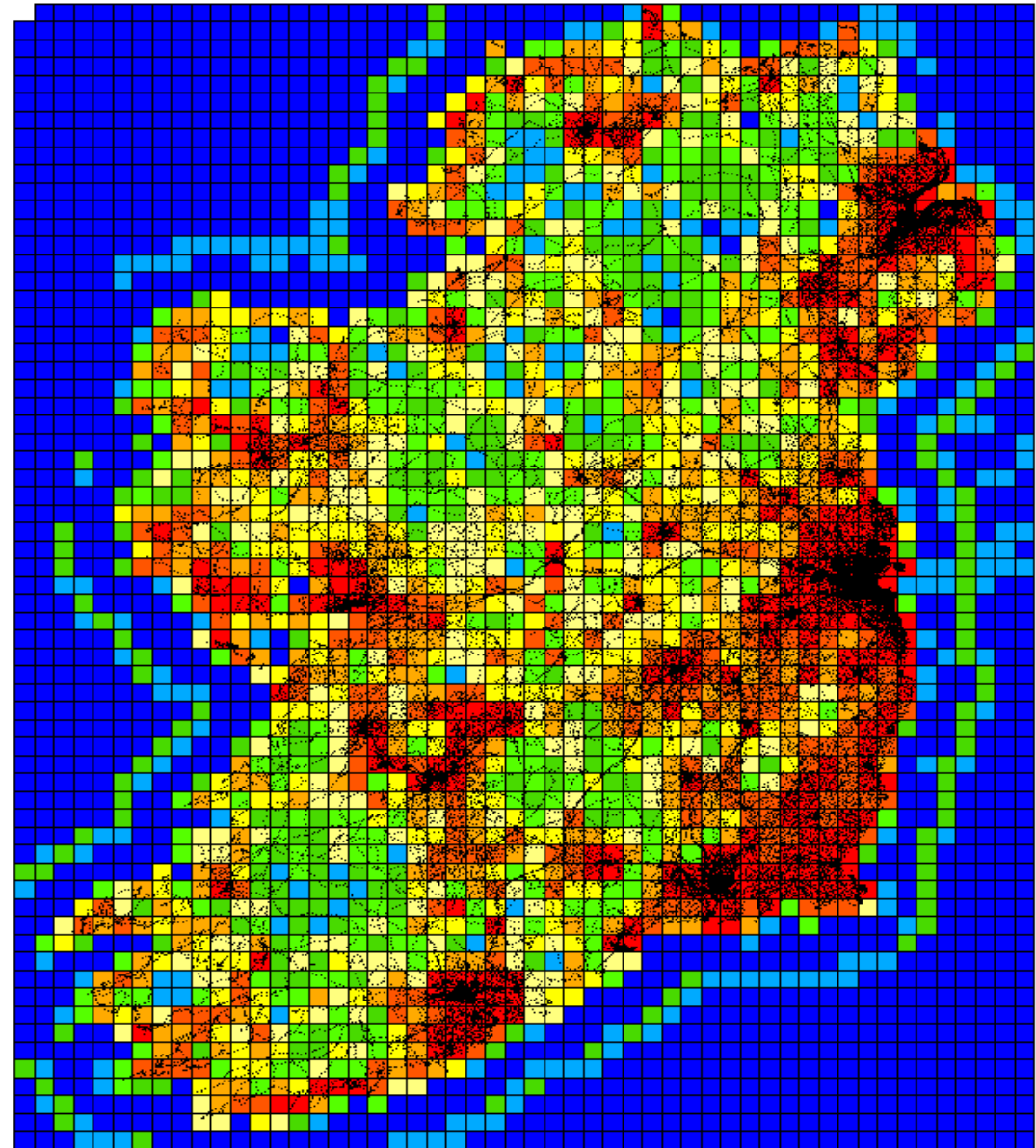


- Might not always be possible to obtain datasets to compare OSM to

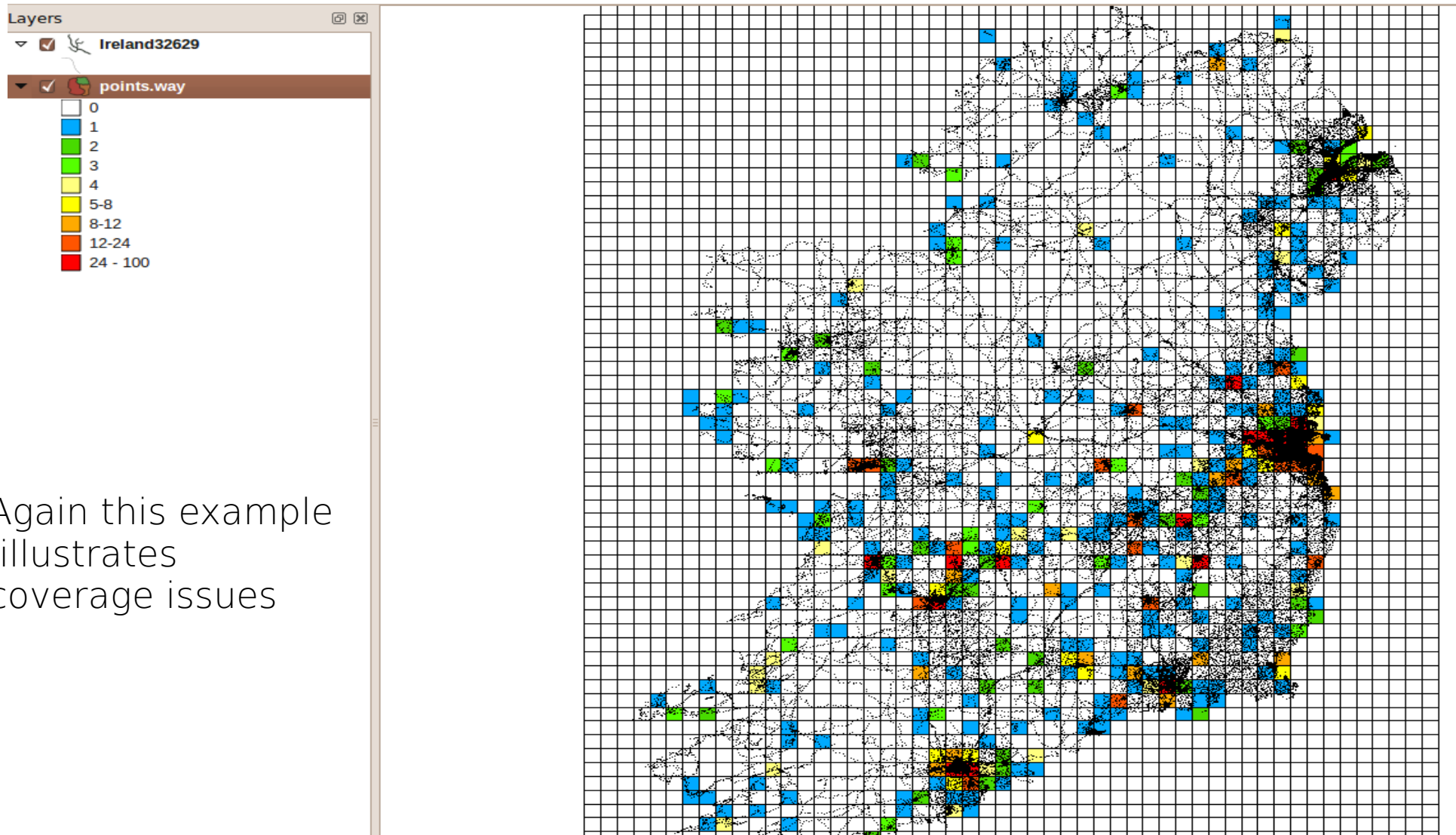
A very basic, but limited, measure of quality (and activity) is number of OSM nodes per cell



- Highlights coverage issues
- shows concentrations around cities/towns and major road arteries



Contributions of physical objects such as #Number of Shops – may indicate actual “on the ground” mapping activity

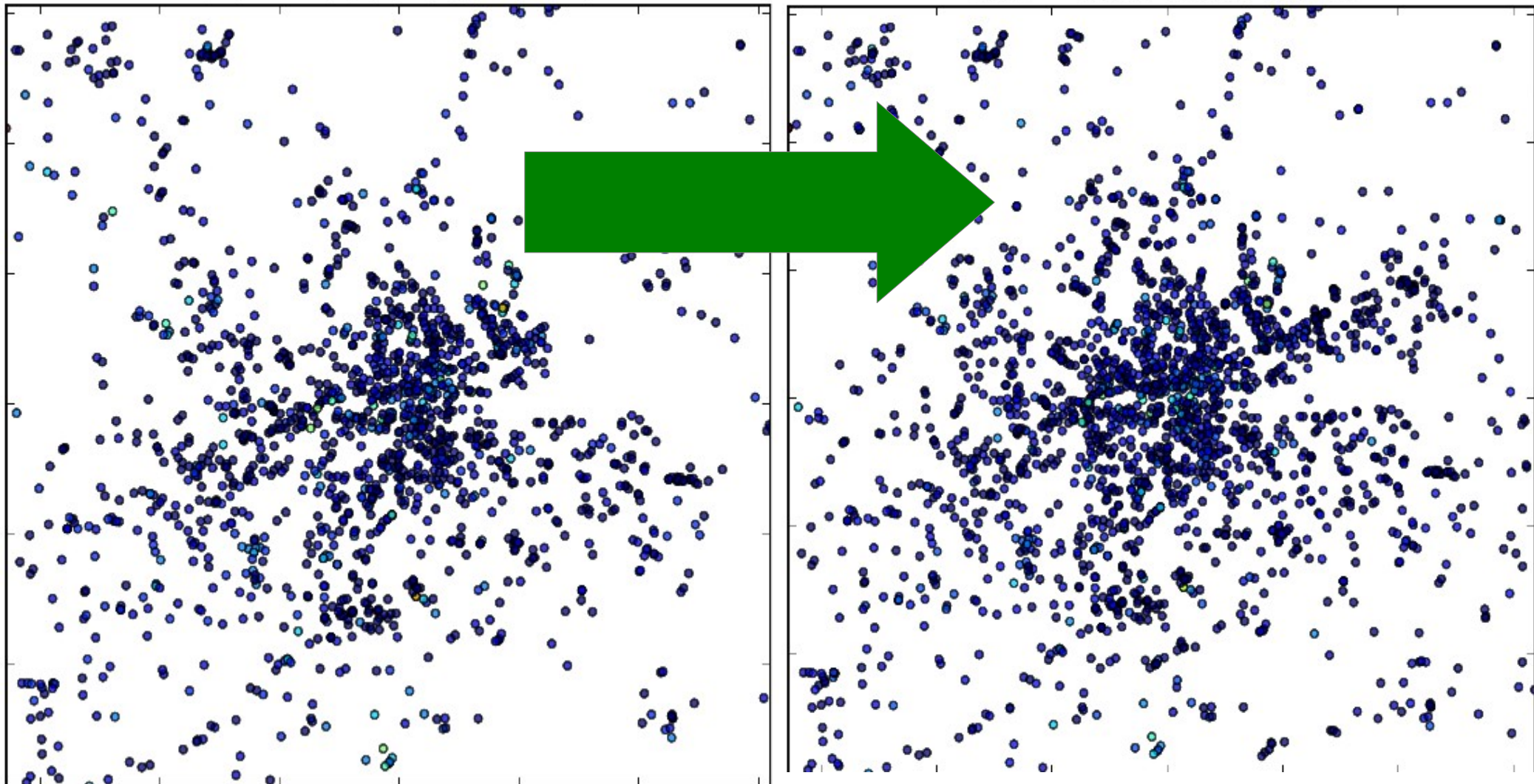


Again this example illustrates coverage issues

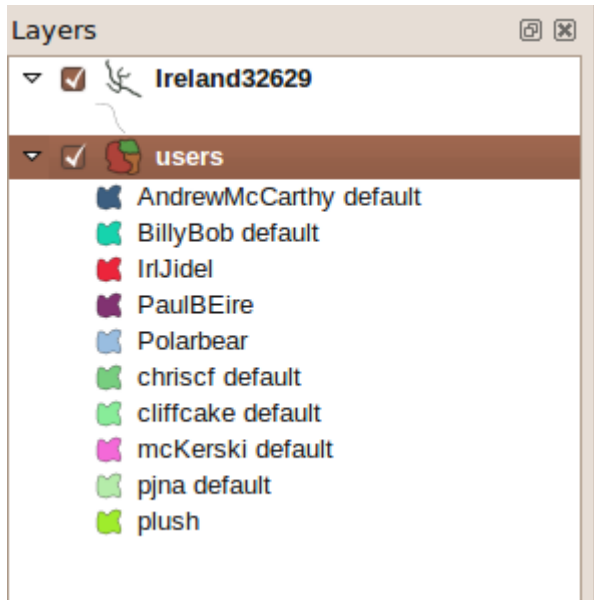
amenity=place_of_worship (London)

London City - place_of_worship
Total = 1590 (2009-10-01)

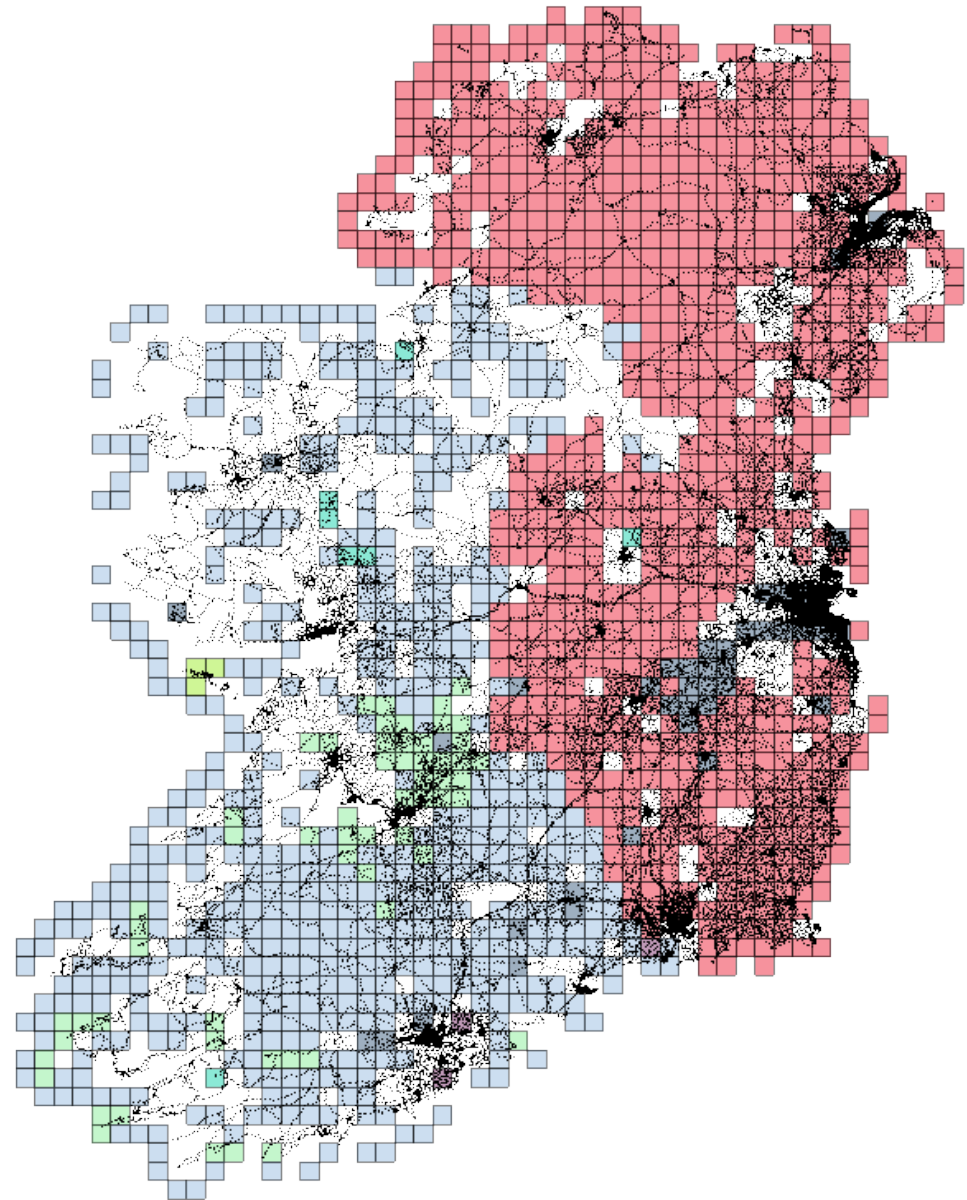
London City - place_of_worship
Total = 2189 (2011-10-01)



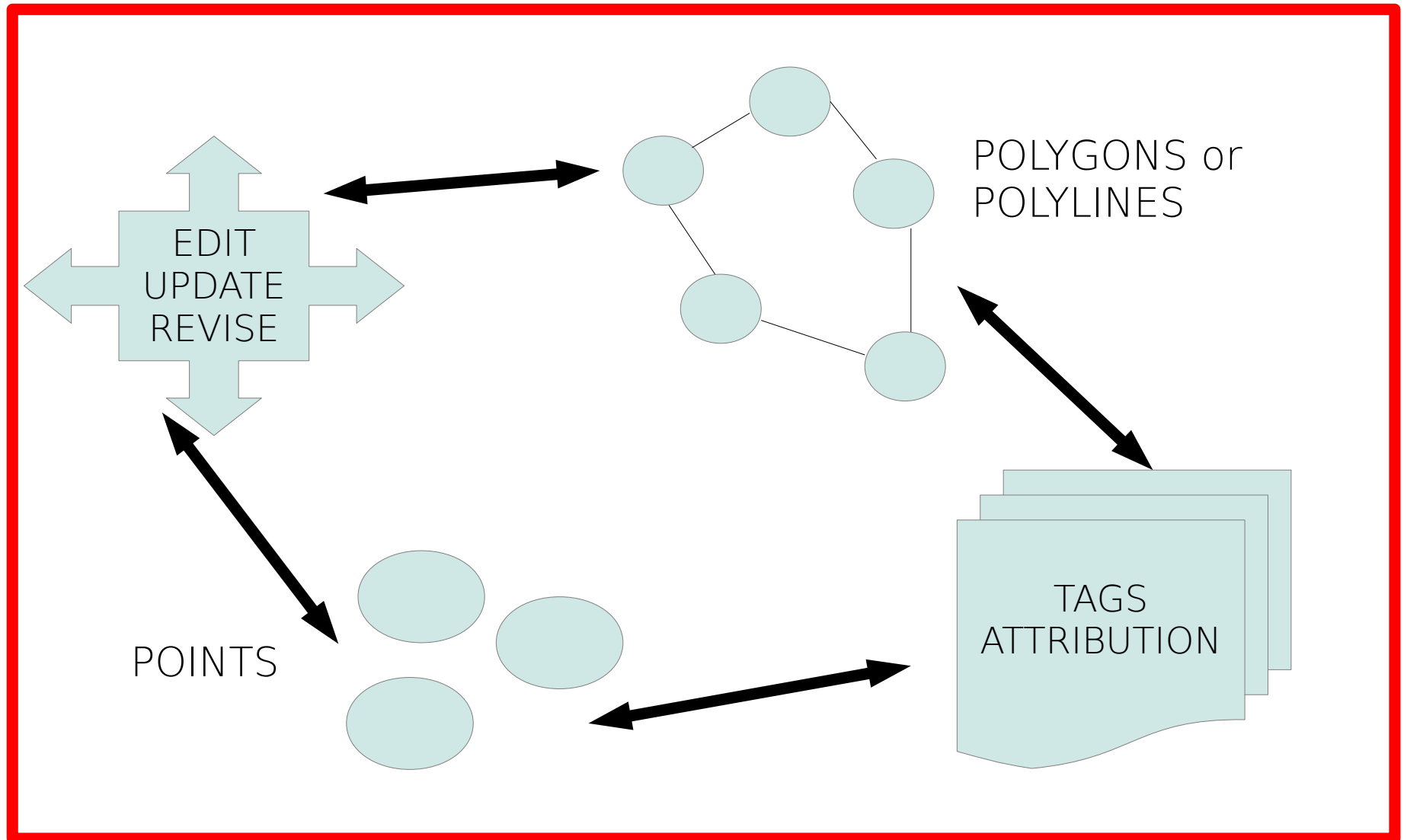
Dominant Users – who are they? And what influence do they have on the quality of OSM



Cell coloured according to the user with the highest number of contributions



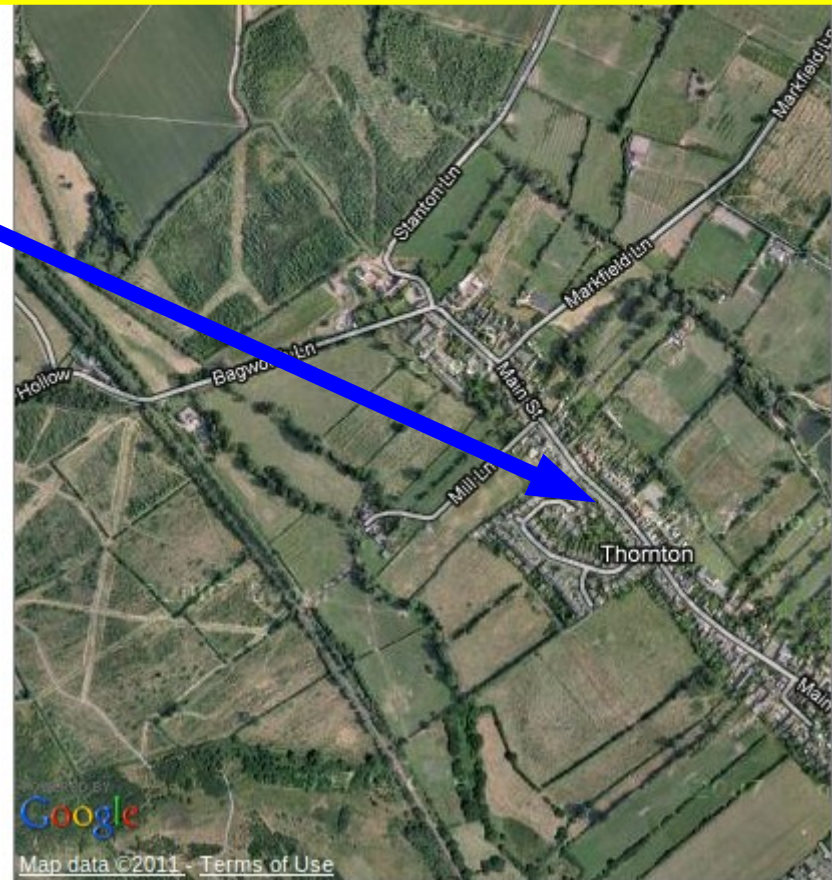
Could OSM “Mapping Structure” be used as a quality indicator?



Example: Automated Identification – Unmapped Villages

CHARACTERISTICS

- Have a POI Village Tag
- Have unclassified and tertiary roads at or near this POI
- No other OSM data within a radius R



<http://resultmap.neis-one.org/>

Unmapped places in OSM
(Villages in UK)

A more precise picture of OSM completeness - at the compound feature level

Example: Equestrian Centres



START

→ The feature is just a POI

→ Some polygons....

→ Polygons and tagging...

→ Compound polygon
with all necessary tags (+ others)
and polygons/polylines

CURRENT



Example: Equestrian Centres - logical steps towards building a complete feature representation

So what makes a "good" equestrian centre
(without comparison to "gold standard" ground-truth?)

Buildings - Stables maintenance building etc.

Greens for riding - open air or Indoor Arena

Parking

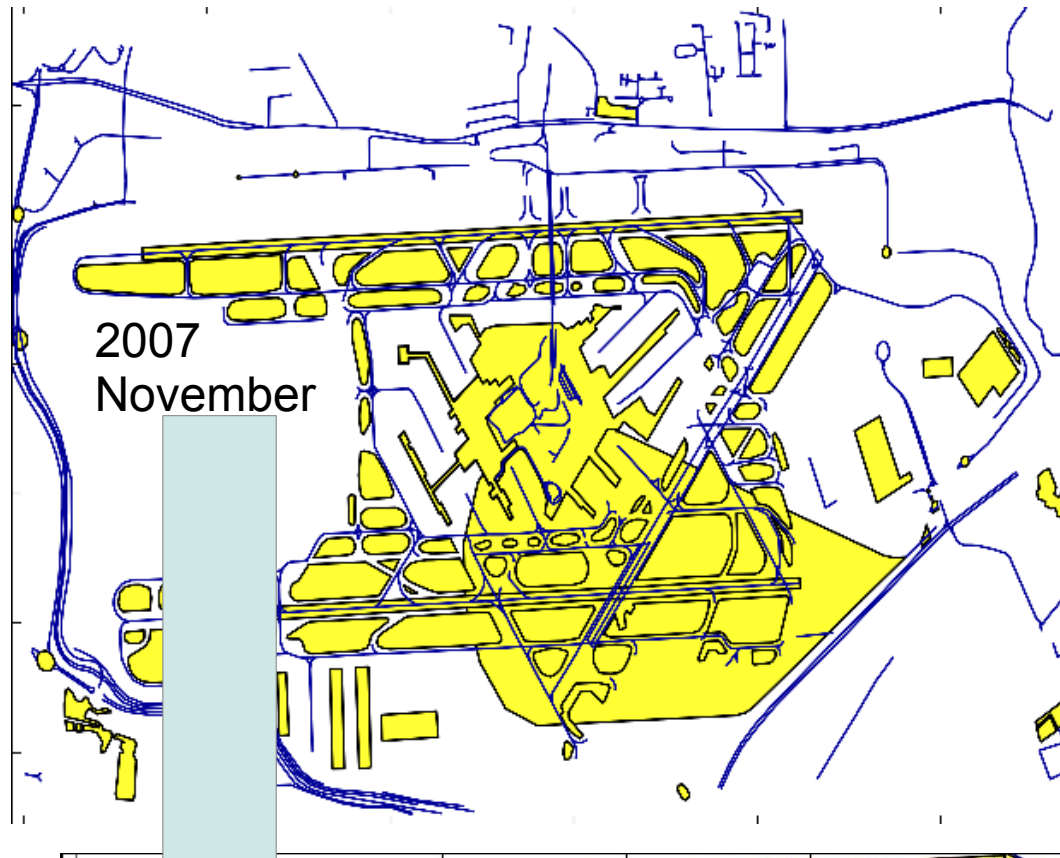
Access Roads

Agricultural Fences

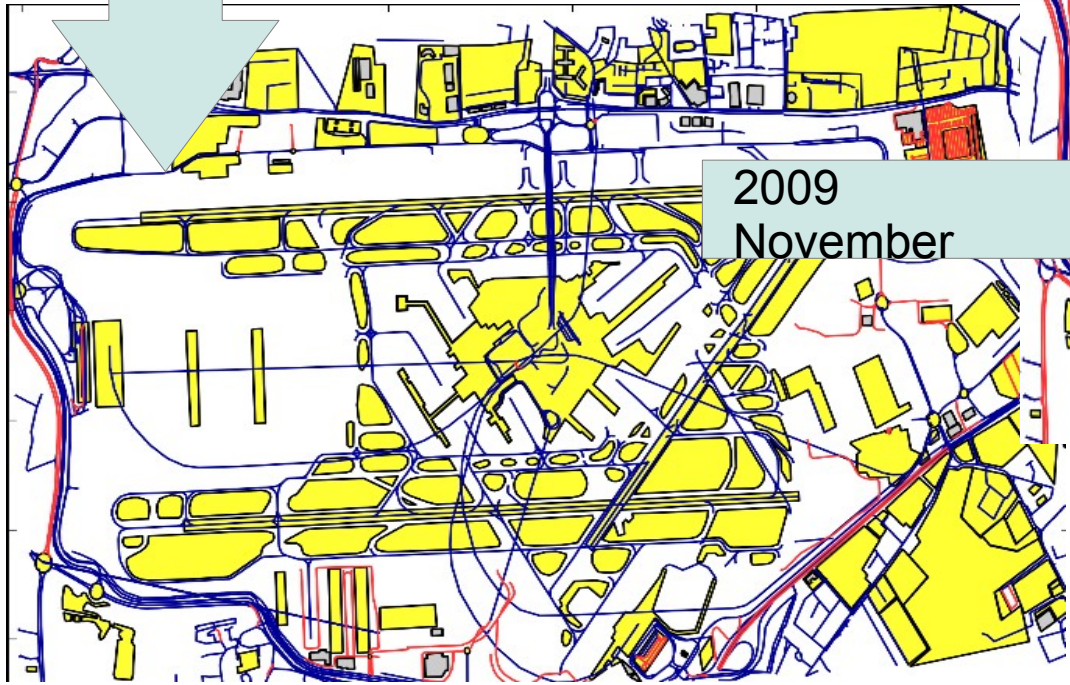


Heathrow Airport

2007
November



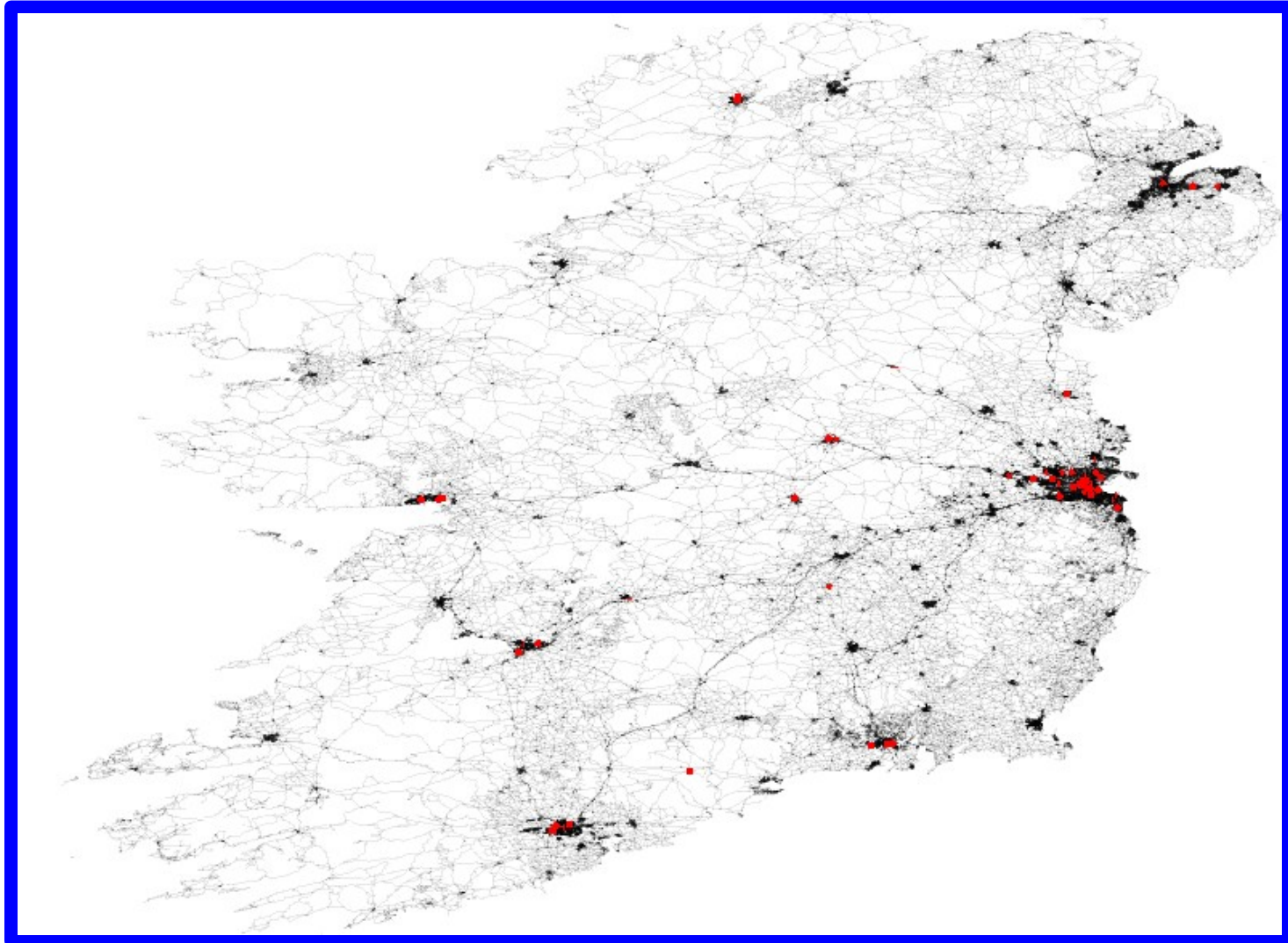
2009
November



November 2011

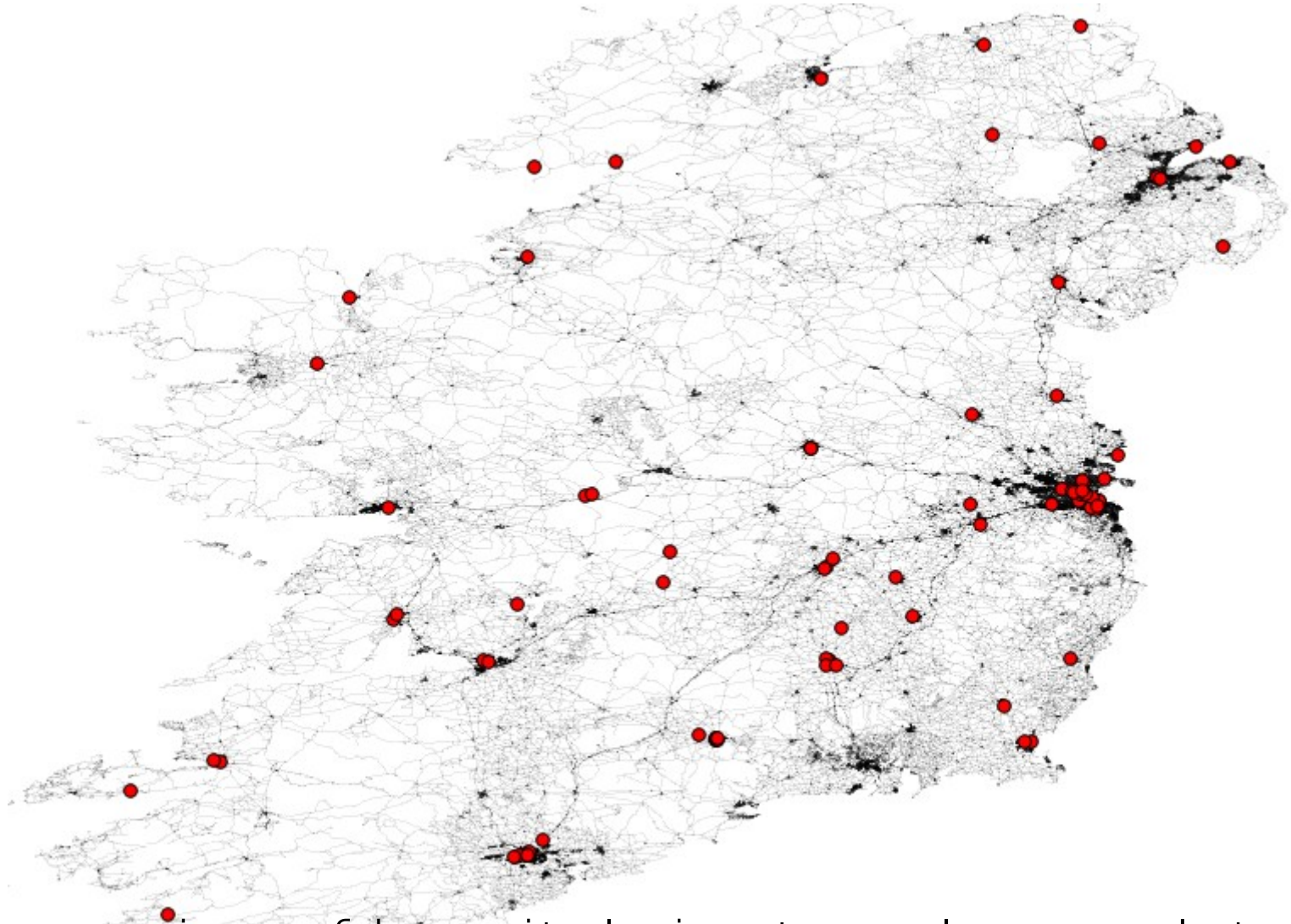


Example: 152 Hospitals in Ireland:
But only 68 are mapped as a polygons



Clustering of Hospital Polygons in Cities

However, 74 hospitals mapped only as a POINT feature (usually POI)



POINT mapping of hospitals is strongly correlated with location outside of large cities and towns

PRELIMINARY CONCLUSIONS

- Understanding the mapping structure of features in OSM requires data intensive computation and database searches – this is still a laborious task – semi automated
- *FOR LBS APPLICATIONS*: Features which are “better mapped” might be given preference in LBS applications – our investigations show that features that follow mapping rules are more complete and better quality
- *FUTURE*: Develop machine learning techniques to “learn” good mapping structure – for easier application in a more automated fashion.

Thanks and Questions?

Email: bciepluch@cs.nuim.ie