

SKETCHES OF GENERIC FRAMEWORK FOR QUALITY ASSESSMENT OF VOLUNTEERED GEOGRAPHICAL DATA

Błażej Ciepluch, Ricky Jacob, Adam C. Winstanley*

Geotechnology Group
Department of Computer Science
National University of Ireland
Maynooth, Co. Kildare. Ireland
Tel. (+353 1 708 3847) Fax (+353 1 708 3848)
bciepluch@cs.nuim.ie
(corresponding author)

Peter Mooney
Environmental Protection Agency
Richview, Clonskeagh
Dublin 14. Ireland

1. INTRODUCTION

The history of volunteered geographical information movement doesn't have such a long tradition. As was mentioned by [1] Kounadi (2009) in her thesis, a whole movement emerged when Google map API was hacked and the first crowd sourcing services like HousingMaps.com and Adrian Holovaty's Chicago Crime³ emerged. In general, the whole movement is based as marked by [2] Goodchild (2007) "human is able to act as an intelligent sensor, perhaps equipped with such simple aids as GPS or even the means of taking measurements of environmental variables". Amateur enthusiasts start to be cartographers but the difference is that they don't have much knowledge about the many aspects of map making. Some authors [3] Coleman et al. (2009) even argue that these contributors can be organized in groups from "Neophyte" to "Expert Authority". Everyone can contribute to the Neogeography field but not everyone contributes data of the same quality. The reason for that is a lack of practice and knowledge which can be improved by practice and experience in map making, amazing to see the citation from [4] Andrew et al. (2009) "How such technologies and tools evolve, is not only dependent on advances in technology itself but also the users of such technology". What is interesting when we look generally at an example of OpenStreetMap which is probably now the best and biggest service of volunteered geographical information system in use today but not the only one. If we talk about contributors from my own experience, I have found that knowledge about mapping practices and conventions is spread between people, from the most influential to complete beginners, by good advice and imitation. We, as researchers are interested how this evolution of quality in data may affect a final product map in a system like OpenStreetMap and we are curious if there will be some way for automatic assessment.

2. EARLY ATTEMPTS TO ASSESS OPENSTREETMAP DATASET

Our first connection with the measurement of data quality in OpenStreetMap was manual and labour intensive. We would like to answer one question, what is going on in Ireland how good are maps there. We have not any access

*Thanks to Environmental Protection Agency for funding this Research

to proprietary datasets so it was necessary to compare Openstreetmap dataset with Bing , Yahoo and Google maps by overlaying OSM lines on available tiles. We chose five cities of different size and representing different types of townland in the Republic of Ireland. Every town was carefully and manually checked, every set against OSM, to compare positional accuracy [5] Blazej et al. (2010) and we gave marks to towns depending on how many mistakes or inconsequences we found inside the road network. Our conclusion from that research was that Motorways are always CORRECTLY shaped in all systems. National Roads are also REPRESENTED ACCURATELY in all systems but sometimes the data is a little outdated, especially in the proprietary systems. When the city is well mapped like Waterford or Dublin, we have nearly all regional roads and estates PRESENT in Openstreetmap. In Drogheda THERE are still big gaps in the coverage of roads. Generally that early attempt was labour intensive and slow but gave us a general overview of how the situation looks in Ireland in the case of VGI data accuracy and reveals to us some indicators, which allow us to say that there is some activity in a particular area.

3. SEMI AUTOMATIC METHOD

We decided to develop an attempt to access our datasets loaded to PostGIS database in an automatic way and measure some simple quality indicators. Our work was influenced by some previous authors who had used similar methods which it is necessary to mention, [6] Muki(2010) and his pioneer research in the quality of the OpenStreetMap area. Part of our work was an algorithm for generated boxes 5km X 5km used in our project and was influenced by his method used to check OSM quality in London city centre. Similar to his method, we created boxes but we chose Ireland , Lithuania and Belgium as our test areas. We had at our disposition an Ordnance Survey Ireland dataset from 2008, donated by StratAG project. Our first measurement was checking Ireland, in which boxes of OSI data had better representation than OSM and vice versa. Similar checking was done by Zielstra et al. [7] also with boxes, but in Germany and Navteq dataset was involved. Our algorithm was developed in PHP and involved scanning through the chosen region which was in our local Postgis database. Then, after extracting and comparing data with another set,we saved the results in another table. That allowed us to produce maps of Ireland which represented different road classes and simple marks where OSI or OSM data were better represented. The conclusion from those experiments was that from Motorways down to secondary roads, there were no big differences between both of our datasets but lower grade roads, especially in rural areas, were better represented in proprietary dataset. After checking our country, Ireland, in that conventional way, we tried to think if there are some quality measurements which are not related to checking something against something else. Where we would be able to say a map has good quality, because these indicators inform us. Our first, from this type, was the number of points in an area box. All map features are built on them so they are simple and important indicators which can tell us what is going on in OSM dataset in a particular area. Simple markers are not only restricted to points. We checked, for example, how the distribution of pubs and shops looked and were not surprised to find that they are usually marked along roads and are under-represented in rural areas, where road networks are still not well mapped. Also the creation of maps, where major contributors influence on particular boxes was represented, gave us something like an overview of what is going on in the Mappers community.

4. CONCLUSION

After some results with these semi automatic ways of checking data quality without a dataset to compare it to, we started to be convinced that it will be possible to build a better algorithm. This would automatically take the data and check the more complex relations between features in an area. We are now in the process of creating this complex relationship. We believe, that based on some algorithm that we are trying to develop, it will be possible to build a

tool which will tell you how good the map is, under your mouse cursor, and why.

5. REFERENCES

- [1] Ourania Kounadi, “Assessing the quality of openstreetmap data,” Msc geographical information science, University College of London Department of Civil, Environmental And Geomatic Engineering, 2009.
- [2] Michael F. Goodchild, “Citizens as voluntary sensors: Spatial data infrastructure in the world of web 2.0,” *International Journal of Spatial Data Infrastructures Research*, vol. Vol.2, pp. 24–32, 2007.
- [3] David J Coleman, Yola Georgiadou, and Jeff Labonte, “Volunteered geographic information: The nature and motivation of producers,” *International Journal of Spatial Data Infrastructures Research*, vol. 4, pp. 332–358, 2009, University of New Brunswick, Canada dcoleman@unb.ca International Institute for Geo-Information Science and Earth Observation, The Netherlands georgiadou@itc.nl DMD Earth Sciences Sector Natural Resources Canada, Canada labonte@nrca.gc.ca.
- [4] Andrew Hudson-Smith, Michael Batty, and Andrew Crooks, “Mapping for the masses accessing web 2.0 through crowdsourcing,” *Social Science Computer Review*, vol. 27, no. 4, pp. 524–538, November 2009.
- [5] Błażej Ciepluch, Ricky Jacob, Peter Mooney, and Adam Winstanley, “Comparison of the accuracy of openstreetmap for ireland with google maps and bing maps,” in -, 2010, Accuracy 2010 Conference.
- [6] Muki Haklay, “How good is volunteered geographical information? a comparative study of openstreetmap and ordnance survey datasets,” *Environment and Planning B: Planning and Design*, 2010.
- [7] Dennis Zielstra and Alexander Zipf, “A comparative study of proprietary geodata and volunteered geographic information for germany,” 13th AGILE International Conference on Geographic Information Science 2010, 2010.