Quality Evaluations on Canadian OpenStreetMap Data

Matthew Tenney[1]

1 McGill University, Montreal, QC, Matthew.Tenney@mail.mcgill.ca

Abstract

The OpenStreetMap (OSM) project represents one of the more popular volunteered geographic information (VGI) projects in the world. This contributed spatial data is used to create global maps that are free to use and edit by anyone. The research presented here, focuses on a systematic quality analysis for Canadian OpenStreetMap data and explores relationships of variation in spatial accuracy to social equity. Accuracy measures of OSM road features are determined through a comparison with an authoritative national dataset of the Canadian road network (Government of Canada 2011). These quality measures are then analyzed against the 2011 National Household Survey (Statistics Canada 2011), in search for social-economic factors like income, education-level, and population that can give insight on OSM quality and the social justice of VGI representation.

Background and Relevance

There have been several studies into the quality of OpenStreetMap and this research adds to these empirical results. The primary quality criterion is positional accuracy for linear features, which has been tested in (Helbich et al. 2012; Koukoletsos, Haklay, and Ellul 2011; Haklay 2010; Ciepluch et al. 2010; Zielstra and Zipf 2010b, 2010a; Neis, Zielstra, and Zipf 2013). Results from these studies have shown interesting patterns in OSM, including a relatively high-level of accuracy. However, with the exception of (Neis, Zielstra, and Zipf 2013) all of these studies were carried out in European countries. Aspects of completeness (Haklay and Ellul 2010; Hochmair, Zielstra, and Neis 2013; Koukoletsos, Haklay, and Ellul 2012; Razniewski and Nutt 2013; Zielstra, Hochmair, and Neis 2013; Wiki 2013) are another major indicator of the data quality available in OSM. Finally, we test attributional accuracy by testing road naming conventions with the reference dataset. Assessing user contributions and motivations in participating on the OSM project can help develop deeper understandings of the VGI production process (Grira, Bédard, and Roche 2010; Hardy, Frew, and Goodchild 2012; Lin 2011; Haklay and Budhathoki 2010; Hars and Qu 2002; Brabham 2012; Budhathoki 2010; Chuang et al. 2013; Mooney and Corcoran 2012, 2011; Mooney 2013). Furthermore, having reliable quality measures for OSM data will better guide any potential applications on a fitness for use basis (Ivánová et al. 2013; Feick and Roche 2013; Humayun and Schwering 2013; Mooney, Corcoran, and Winstanley 2010; Mooney, Corcoran, and Ciepluch 2012)

Methods and Data

Our study area includes over 5,000 census tracts across populated areas in Canada. We clip OSM and reference road datasets according to each tract boundary and calculate our quality measures. We follow the methods detailed in several published studies including (Jilani,

Corcoran, and Bertolotto 2013; Koukoletsos, Haklay, and Ellul 2012). In order to test positional accuracy we follow the methods used in other studies and detailed in (Goodchild and Hunter 1997; Koukoletsos, Haklay, and Ellul 2011; Haklay 2010). Completeness is also calculated from the definition offered in (Koukoletsos, Haklay, and Ellul 2011, 2012; Haklay and Ellul 2010). Road names are compared using a semantic-similarity and character-to-character comparison similar to those offered in (Ballatore, Bertolotto, and Wilson 2013; Janowicz et al. 2008; Girres and Touya 2010; Kardos, Moore, and Benwell 2003). Using the established quality measures we then test for any correlations in variance between income, education-level, and population.

Results

The results for this study are still in their preliminary stages. This is in part due to the ad-hoc nature of the methodology and because of omitted factors that should be considered before any significant findings are be reported. Negative correlations were found in all of the quality measures to our tested social equity data. A strong positive relationship was shown in error to census tract area size. There is evidence of a strong negative relationship between road name accuracy and positional accuracy, especially in rural areas with high-levels of data imports. The nature of our preliminary results indicate more attention should be paid to the role of imports to OSM data in terms of timelessness and systematic error propagation.

Conclusions

Although there is little confidence in the results from this pilot study – some aspects of our analysis show mixed findings when compared to previous work. There seems to be a similar relationship between rural and urban areas in OSM quality as was found in Europe (Corcoran and Mooney 2013). This corresponds to Canada's dense urban population patterns. Data imports into the OSM data seem to have both positive and negative impacts on the overall quality of the OSM map. Completeness and attributional accuracy increase with the data imports from authoritative datasets (Zielstra, Hochmair, and Neis 2013). However, positional accuracy and the up-to-datedness of the data is seriously hindered with these mass imports because of systematic errors in the outdated original data.

While there was indication of weak negative correlations between the tested social-equity criterion and OSM data quality, the tests used to this point are not necessarily the most appropriate for the task. Local indicators of spatial association (Anselin 1995) will be implemented to further understand any spatial relationships that might exist. Also, need to analyze the user-contribution, motivation, and development of OSM data in Canada is required to better understand variations of user-contribution patterns. This will also aide in understanding the applicability of using census tract boundaries and social-economic information to gauge social justice for OSM data representation. Understanding 'who' is being represented through the 'local-knowledge' of this collaborative mapping project can guide further research and evaluations on the nature of VGI and help establishing a fitness of use for a variety of applications using OSM.

References

Anselin, L. 1995. Local Indicators of Spatial Association—LISA. *Geographical Analysis* 27 (2):93–115.

Ballatore, A., M. Bertolotto, and D. C. Wilson. 2013. Geographic knowledge extraction and semantic similarity in OpenStreetMap. *Knowledge and Information Systems* 37 (1):61–81.

Brabham, D. C. 2012. Motivations for Participation in a Crowdsourcing Application to Improve Public Engagement in Transit Planning. *Journal of Applied Communication Research* 40 (3):307–328.

Budhathoki, N. R. 2010. Participants‘ motivations to contribute geographic information in an online community. http://hdl.handle.net/2142/16956.

Chuang, T.-R., D.-P. Deng, C.-C. Hsu, and R. Lemmens. 2013. The One and Many Maps: Participatory and Temporal Diversities in OpenStreetMap. In *Proceedings of the Second ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information*, GEOCROWD ’13., 79–86. New York, NY, USA: ACM http://doi.acm.org/10.1145/2534732.2534737 (last accessed 21 December 2013).

Ciepluch, B., R. Jacob, P. Mooney, and A. Winstanley. 2010. Comparison of the accuracy of OpenStreetMap for Ireland with Google Maps and Bing Maps. In *Proceedings of the Ninth International Symposium on Spatial Accuracy Assessment in Natural Resuorces and Enviromental Sciences 20-23rd July 2010*, 337. http://eprints.nuim.ie/2476/ (last accessed 21 December 2013).

Corcoran, P., and P. Mooney. 2013. Characterising the metric and topological evolution of OpenStreetMap network representations. *The European Physical Journal Special Topics* 215 (1):109–122.

Feick, R., and S. Roche. 2013. Understanding the Value of VGI. In *Crowdsourcing Geographic Knowledge*, eds. D. Sui, S. Elwood, and M. Goodchild, 15–29. Dordrecht: Springer Netherlands http://link.springer.com/chapter/10.1007/978-94-007-4587-2_2 (last accessed 18 December 2013).

Girres, J.-F., and G. Touya. 2010. Quality Assessment of the French OpenStreetMap Dataset: Quality Assessment of the French OpenStreetMap Dataset. *Transactions in GIS* 14 (4):435–459.

Goodchild, M. F., and G. J. Hunter. 1997. A simple positional accuracy measure for linear features. *International Journal of Geographical Information Science* 11 (3):299–306.

Government of Canada, N. R. C. 2011. GeoBase - National Road Network. http://www.geobase.ca/geobase/en/data/nrn/description.html (last accessed 23 December 2013).

Grira, J., Y. Bédard, and S. Roche. 2010. Spatial data uncertainty in the VGI world: Going from consumer to producer. *Geomatica* 64 (1):61–71.

Haklay, M. 2010. How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and planning. B, Planning & design* 37 (4):682.

Haklay, M., and N. Budhathoki. 2010. OpenStreetMap–Overview and Motivational Factors. *communication au Horizon Infrastructure Challenge Theme Day, Université de Nottinghan*. http://www.ideals.illinois.edu/handle/2142/16461 (last accessed 18 December 2013).

Haklay, M., and C. Ellul. 2010. Completeness in volunteered geographical information – the evolution of OpenStreetMap coverage in England (2008-2009). *Journal of Spatial Information Science*. http://discovery.ucl.ac.uk/150397/ (last accessed 18 December 2013).

Hardy, D., J. Frew, and M. F. Goodchild. 2012. Volunteered geographic information production as a spatial process. *International Journal of Geographical Information Science* 26 (7):1191–1212.

Hars, A., and S. Qu. 2002. Working for Free? Motivations for Participating in Open-Source Projects. *International Journal of Electronic Commerce* 6 (3):25.

Helbich, M., C. Amelunxen, P. Neis, and A. Zipf. 2012. Comparative Spatial Analysis of Positional Accuracy of OpenStreetMap and Proprietary Geodata. *J. Strobl et al., Angewandte Geoinformatik* :24–33.

Hochmair, H. H., D. Zielstra, and P. Neis. 2013. Assessing the completeness of bicycle trails and designated lane features in OpenStreetMap for the United States and Europe. In *Proceedings of the Ninety-second Annual Meeting of the Transportation Research Board*. http://flrec.ifas.ufl.edu/hochmair/pubs/hochmair_zielstra_neis_TRB2013.pdf (last accessed 23 December 2013).

Humayun, M. I., and A. Schwering. 2013. Selecting a Representation for Spatial Vagueness: A Decision Making Approach. In *Geographic Information Science at the Heart of Europe*, Lecture Notes in Geoinformation and Cartography., eds. D. Vandenbroucke, B. Bucher, and J. Crompvoets, 115–131. Cham: Springer International Publishing http://link.springer.com/chapter/10.1007/978-3-319-00615-4_7 (last accessed 18 December 2013).

Ivánová, I., J. Morales, R. A. de By, T. S. Beshe, and M. A. Gebresilassie. 2013. Searching for spatial data resources by fitness for use. :15–28.

Janowicz, K., M. Raubal, A. Schwering, and W. Kuhn. 2008. Semantic similarity measurement and geospatial applications. *Transactions in GIS* 12 (6):651–659.

Jilani, M., P. Corcoran, and M. Bertolotto. 2013. Automated Quality Improvement of Road Network in OpenStreetMap. http://flrec.ifas.ufl.edu/geomatics/agile2013/papers/jilani_ACTIVITY_AGILE_2013.pdf (last accessed 17 December 2013).

Kardos, J. D., A. Moore, and G. L. Benwell. 2003. Visualising uncertainty in spatially-referenced attribute data using hierarchical spatial data structures. In *Proceedings*

of the 7th International Conference on GeoComputation, UK. http://www.geocomputation.org/2003/Papers/Kardos_Paper.pdf (last accessed 19 December 2013).

Koukoletsos, T., M. Haklay, and C. Ellul. 2011. An automated method to assess Data Completeness and Positional Accuracy of OpenStreetMap. 20–22. http://www.geocomputation.org/2011/papers/koukoletsos.pdf (last accessed 18 December 2013).

Koukoletsos, T., M. Haklay, and C. Ellul. 2012. Assessing Data Completeness of VGI through an Automated Matching Procedure for Linear Data. *Transactions in GIS* 16 (4):477–498.

Lin, Y.-W. 2011. A qualitative enquiry into OpenStreetMap making. *New Review of Hypermedia and Multimedia* 17 (1):53–71.

Mooney, P. Understanding the activity of contributors to VGI projects. How, why, where, and when do they contribute geographic information? http://icaci.org/files/documents/ICC_proceedings/ICC2013/_extendedAbstract/448_proceeding.pdf (last accessed 18 December 2013).

Mooney, P., and P. Corcoran. 2011. Annotating spatial features in OpenStreetMap. *Proceedings of GISRUK 2011*. http://www.cs.nuim.ie/~pmooney/websitePapers/GISRUK2011_Peterv3-FinalSubmitted-MonFeb28th2011.pdf (last accessed 18 December 2013).

———. 2012. How social is OpenStreetMap. In *The 15th AGILE International Conference on Geographic Information Science. Avignon*. http://www.cs.nuim.ie/~pmooney/websitePapers/PETER_AGILE_2012Corrected-FINAL-MARCH13TH2012.pdf (last accessed 18 December 2013).

Mooney, P., P. Corcoran, and B. Ciepluch. 2012. The potential for using volunteered geographic information in pervasive health computing applications. *Journal of Ambient Intelligence and Humanized Computing* :1–15.

Mooney, P., P. Corcoran, and A. C. Winstanley. 2010. Towards Quality Metrics for OpenStreetMap. In *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 514–517. New York, NY, USA: ACM http://doi.acm.org/10.1145/1869790.1869875 (last accessed 18 December 2013).

Neis, P., D. Zielstra, and A. Zipf. 2013. Comparison of Volunteered Geographic Information Data Contributions and Community Development for Selected World Regions. *Future Internet* 5 (2):282–300.

Razniewski, S., and W. Nutt. 2013. Assessing the Completeness of Geographical Data. 228–237. Berlin, Heidelberg: Springer-Verlag http://dx.doi.org/10.1007/978-3-642-39467-6_21 (last accessed 17 December 2013).

Statistics Canada. 2011. National Household Survey. *Statistics Canada*. http://www.statcan.gc.ca/survey-enquete/household-menages/5178-eng.htm (last accessed 22 December 2013).

Wiki, O. 2013. Completeness - OpenStreetMap Wiki. http://wiki.openstreetmap.org/wiki/Completeness (last accessed 22 December 2013).

Zielstra, D., H. H. Hochmair, and P. Neis. 2013. Assessing the Effect of Data Imports on the Completeness of OpenStreetMap – A United States Case Study. *Transactions in GIS* 17 (3):315–334.

Zielstra, D., and A. Zipf. 2010a. A comparative study of proprietary geodata and volunteered geographic information for Germany. In *13th AGILE International Conference on Geographic Information Science*. http://agile2010.dsi.uminho.pt/pen/shortpapers_pdf/142_doc.pdf (last accessed 18 December 2013).

———. 2010b. OpenStreetMap data quality research in Germany.