International Journal of Engineering Research in Computer Science and Engineering (IJERCSE) Vol 3, Issue 6, June 2016

Review of Open Street Map Data Quality And its Assessment Tools

[1]Navneet Kaur, [2]Sukhjit Singh Sehra, [3]Sumeet Kaur Sehra
[1]PG Student, [2] [3] Assistant Professor
[1][2][3]Department of Computer Science & Engineering, Guru Nanak Dev Engineering College, Ludhiana, Punjab, India,
[1] navneet.kaur0312@gmail.com,[2]sukhjitsehra@gmail.com,[3]sumeetksehra@gmail.com

Abstract: **— Volunteered Geographic Information (VGI) has been becoming a significant source for understanding the surface of the Earth. VGI refers to the collection of huge spatial data with the help of people on voluntary basis. Open street map (OSM) project is one such example of VGI. Since contributors have varying levels of education, profession or skills, it has resulted into number of inconsistencies in OSM data. The aim of this study is to review the research that has been done for quality assessment of OSM data and the tools available for improving the quality of OSM data. It can be concluded that a number of quality assurance tools are available but many of them covers only specific parts of world and many evaluate errors that have little impact for most data uses. Also the densely populated areas are mapped better than sparsely populated areas as local editors being familiar with an area leads to more accurate OSM mapping.**

Index Terms— **Open street map, Open street map Data Quality, Quality Tools, Volunteered Geographic Information**

I. INTRODUCTION

Technology has been changing since last 10 years, which give rise to a new term "crowd sourcing". It is one of the most noteworthy developments in Web 2.0 which has controversial potential. It deals with how a large groups of people are performing their work which is difficult to automatize and costly to exploit [1], [2].

There are many companies which collect data without crowd sourcing and can provide such geographic information. The major companies among them are navteq and teleatlas. The data provided by them is very expensive, quickly outdated and covers only specific areas. A large investment is done by large companies for acquiring data from smaller companies e.g. Nokia acquired navteq in 2007, microsoft acquired the imagery and remote sensing company vexcel in 2006 [3].

a. Volunteered geographic information (vgi)

Good child has given a new term to this collection of spatial information by a group of people as "*volunteered geographic information*" [4]. Good child stated that vgi has so much potential in it for becoming a significant source for understanding the surface of the earth. The most important value of vgi is that it can tell about local activities in different geographical areas which go unnoticed by world media. Now-a-days the quality of this spatial information is the hot topic in the field of research [2].

b. Open street map (OSM)

A very powerful example of crowd sourcing is Open street map project. OSM project was started in 2004 by Steve Coast at University College London (UCL). OSM follows the similar model as followed by Wikipedia and it aims to create free and editable map licensed under new copyright conditions [1].

The main advantage of OSM is that it provides the free digital geographical information across the world. In most of the Western countries this information is provided by commercial agents or national agencies which is either too expensive or out of reach of individuals or organizations. In such cases, OSM is the cheapest and the only source for providing geographic information where there is an issue of national security [2], [4].

For configuring OSM information database, three elements are must: node, way, and relation. Node defines a specific point which is described by its latitude and longitude e.g., a water well. A way is an ordered interconnection of nodes between 2 and 2000 which defines a poly line for highways, railways, and building outlines and also Area is a closed way. A relation is used to define a relationship between nodes, ways or other data elements e.g., 'turn restriction' between two ways. A variety of information can be added by simply tagging above mentioned elements [5], [6].

c. User Contribution to OSM

At the time of writing, there are about 2,697,409 users of OSM [7]. Although the number of OSM users and their active contribution is very high, but it is to be noted that they all may not possess professional qualifications or background in collection or survey of geodata. There may also be case when users do not follow some specific guidelines while editing data and they produce inconsistent results. The quality of geodata collected by users also depends upon method or instrument used for its collection, e.g., PC, laptop, Smartphone or Global Positioning System (GPS) device. Also population density also has a role to play. The more densely populated areas are mapped better or are more complete than sparsely populated areas. However, most participants become the local experts due to possession of their local knowledge [8], [9].

d. Quality of Volunteered Geographic Information

The quality can be defined in a number of ways, according to Coote and Rackham, [10] it can be defined as fitness for purpose, that is how well a data is suited for particular problem.

In VGI, the contributors or cartographers have different professional backgrounds. The equipments or the methods used by them for producing geographic information are different. This leads to data with mixed quality resulted from combination of diverse professional or knowledge backgrounds, different technical constraints and also different views of the data. Hence, the analysis of VGI data quality became a focus in spatial data related research. There are three different methods for assessing the quality of spatial data: comparison with respect to reference data, semantic analysis, and intrinsic data analysis [11].

e. Elements of Spatial Data Quality

The OSM data is very huge containing large amount of data which relates to the people who collected it. They have produced this labelled data with different ways. Researchers have been working on the quality of OSM data and thus for evaluating the OSM data, different parameters are needed [3]. Oort [12] identifies the following aspects of spatial data quality:

1) Positional accuracy
It deals with how well coordinate position of a object relates to its actual location.

2) Lineage
Lineage is the history of a spatial dataset that how data is derived, from where it has come and how it is produced.

3) Completeness
It corresponds to how much information about object is missing or is present in excess.

4) Logical consistency
This aspect tells the topological relations and correctness which corresponds to internal consistency of database.

5) Semantic accuracy
This evaluates how the object is stored in database, is it meaningful, or how it is interpreted.

6) Attribute accuracy
The objects in OSM database are not only represented geometrically but also explained by their attributes, this aspect measures how accurate such values are.

7) Temporal quality
This evaluates the validity of changes done in database according to real world changes.

8) Purpose, usage and constraints
This tells how the data is stored according to its purpose and how easily it could be used by users.

f. Open street map Data Quality

The analysis of peer production content based upon urban and rural areas was done. It was found that in both the cases of Wikipedia and OSM, the data produced about rural areas is of lower quality than urban areas, and it is less likely to have been contributed by contributors whose main focus is on the local area, and is more likely to have been produced by automated software agents (i.e. bots). The research has also shown that local editors tend to provide better quality of work [13].

The study of the evolution of VGI was done by Neis et al. [8] in Germany using OSM as case study from 2007 to 2011. They specially consider the expansion of the total street network and the route network for navigation of cars. When the OSM database was compared with Tom Tom's commercial dataset based on the parameter of completeness, the results revealed that OSM has 27% more data for total street network and route information for pedestrians in Germany. While on the other hand, 9% of data with regard to car navigation is still missing in OSM database. The authors also did analysis regarding topology errors and the completeness of street name information. The results showed that the OSM data also have some flaws, but the trend reveals that both relative and absolute number of errors is decreasing.

In a study it was found that there is an active community of contributors in Germany, and the sum of features and the number of attributes in the Netherlands is the largest among all other countries in Europe. And also in Southern Europe, the activities are mostly restricted to urban areas only [14].

The OSM data values of Japan are also studied by Hayakawa et al. [5], which are approaching those of Europe and other developed countries. According to a survey of OSM Merit by active users values, Japan's data have many more orders of magnitude than other regions which means that much editing has been done by a few active users which has greatly enhanced the OSM data. As VGI situation is very unstable, the motivation of contributors is very much required. The lack of motivation of active users will have a great impact on updating and maintaining OSM data in Japan in near future.

g. Issues in OSM Data Quality

Following are the issues which are encountered with the quality of OSM data [13]:

1) ***Vandalism***
   As soon as creating account on OSM, anyone can start editing immediately. Thus the people with wrong intentions do malicious editing or add incorrect information. To deal with vandalism, OSM superusers tracks the changes and look for if any wrong changes took place. They can then undo the changes and users who are continuously vandalising the map can lose their editing rights.
2) ***Missing Details***
   There is no such perfect map in the world with which ours can be compared so it is difficult to find areas with missing data. The data present is often partial, e.g. a road is present without name or a building with missing basic details which are used for its categorization.
3) ***Routing Problems***
   One of the most important uses of Open street map is that it can be used for routing purposes. Tools like OSRM and MapQuest Open can give directions while driving based on of its data. This routing network can have problems related to disconnected roads, improper turn restrictions, and unmarked speed limits.

Quality Assurance of Open street map Data

II. RELATED WORK

Qian et al. [16] stated that geographic data cleaning such as finding duplicated points etc. must be done. For that, spatial and attribute information must be taken into consideration. The authors discussed various data cleaning methods for VGI data and experimentally used them in some possible application environments. Karam and Melchiori [17] discussed a way for assessing geospatial crowdsourced data which improves geoLinked Open Data. Their idea involves a scoring mechanism which improves the quality by organizing and ranking user generated corrections. The framework creates a special dataset which stores the data of volunteered corrections to geospatial data and also the problems detected like discrepancies and mismatching between data from different sources. This dataset is then used by volunteers or paid contributors, who use their local knowledge to solve the problem. For this work, the methodology proposed by authors in [18] is adopted which usually focuses on vocabulary reuse.

For checking the integrity of VGI data, Ali and Schmid [11] proposed two methods considering "hierarchical consistency and classification plausibility". These methods can do the validation of data at the time of contribution or later on. The correction can either be done manually or automatically. These methods are machine learning based and also checks the constraints over the data. Hashemi and Abbaspour [19] also proposed a framework For determining the probable inconsistencies in OSM data which uses the concept of spatial similarity in multi-representation along with taking into consideration three parameters: distance, direction and topological relationships between objects.

For dealing with spatial data quality issues like connected polygon features in Open street map, Ying et al. [20] developed a software tool which selects suitable sets of polygons from Open street map. In the testing of generalisation algorithms, the selected polygons can be used. This software helps the authors in identifying the connectivity issues in the Open street map database. This is required for generalisation in applications such as Location-Based Services where new polygon features are inserted and existing polygon features are updated. For improving the semantic quality and reducing the semantic heterogeneity of VGI data, Vandecasteele and Devillers [21] proposed a tag recommender system, called OSMantic, which can automatically suggest suitable tags to contributors during the editing process. Thus decreasing the semantic heterogeneity and improving semantic quality. This system has been evaluated by the contributors and there is high level of satisfaction among them. For improving quality management in User Generated Spatial Content, Brando et al. [22] also gives an approach built on formal specifications and on external reference data. Formal specifications helps in quality management during contribution in the following ways: they include integrity constraint to check consistency, to improve quality of VGI through external reference data, and to resolve concurrent copy of data. Ali et al. [23] proposed a novel approach for addressing phase of data quality in VGI: classification ambiguity and plausibility. This approach is developed with machine learning from VGI data only without using any reference data.

Helbich et al. [24] did the analysis of positional accuracy by comparing VGI and proprietary geospatial data while taking into consideration a well-mapped German city in OSM. The results showed significant clusters of high and low positional accuracy which were obtained by applying bi-dimensional regression analysis.

Ciepluch et al. [25] discussed that the quality assessment is currently done based on bugs reporting or by involving human interactions. If ground-truth dataset is accessible it can be compared with OSM for some specific region for analysing its quality. If ground-truth data is not available then alternative solution is to identify points of data within grid squares. It simplifies more complex spatial analysis problem. For improvement of OSM data quality, the authors have developed a suite of quality indicators. The quality indicators evaluate a wide range of characteristics. Some of the quality investigations in this work are easy to implement, fully automated and gives results quickly. While other quality assessments such as user profiling takes longer time.

Pourabdollah et al.[24] did the quality evaluation of OSM data of Great Britain. They checked ISO-19157 quality types and examined the dynamics of errors and features in OSM. They studied Great Britain's OSM database using 17 quality rules. This study was done for a period of 50 days i.e. from 28-10-2012 till 17-12-2012. The results showed overall increase as 7.5 per 1000 in bugs which were earlier 14 per 1000. The line intersections without junction bug grew by 6500(approximately) during that period which represents some particular area in community of OSM. The results also showed that rates of removal and addition of bugs are relatively balanced. In the sample taken by the authors, intersections without junction bugs are growing mostly in both absolute and relative terms whereas line spikes are quickly fixed.

Quality Assurance Tools

Quality Assurance tools helps in improving the quality of OSM data. They do so by providing a list of bugs in data which is fixed by mappers using editing tools of OSM. The bugs are detected- automatically based upon some rules or data analysis, or tools provide means to report them manually, or by the combination of these two. There are number of different tools based upon different ideas. OSM data has better quality than other commercial maps while dealing with new Ways. Different types of quality assurance tools are described below:

1) Bug reporting tools

These tools highlight the parts of data which are likely to be incorrect. They can help in visualisation of areas which require attention and spot and correct errors.

- Notes Notes is one of features of the Open street map website. The website has 'Add a note' button in the lower right corner where error reports can be added.
- Mapdust

  MapDust is operated by skobbler whose original focus was on navigation-relevant bugs but now serve as a general bug tool.

2) *Error detection tools*

Error Detecting Tools check for the potential errors or inaccuracies in OSM data. After that users check if the structures are really incorrect and correct them for improving quality of data. Some of the errors detecting tools are:

- ***Keep Right***t detects a number of errors automatically based on rules and shows them on map or inlisted form. It reports false positives and labels bugs as fixed. It can detect errors such as non-closed areas, dead-ended one-ways, deprecated tags, almost junctions, missing tags, motorways without ref, etc [27].
- ***Osmose*** stands for Open street map Oversight Search Engine, which detects issues in Open street map data. this is a tool for reporting issues regarding two parts [28]:
  - Frontend- displaying issues over slippy map
  - Backend- OSM data analysis and issues detection
- ***JOSM Validator*** checks the loaded data in editor, and shows the errors and warnings. On request it can also automatically fix them [27].
- ***OSM Inspector*** is web based tool provided by Geofabrik. There are many views on a map, each having several layers representing specific details or errors in OSM data. It is very easy to switch on and off between layers, and on mouse click it will give details about any feature and links will redirect to editor to fix the problems easily [29].
- ***Maproulette*** uses a gamified approach to fix OSM data bugs by breaking big problems into small tasks. There is need to grab a challenge for improving OSM data. At a time only one challenge is activated on Maproulette [30].

A comparison of some of the error detection tools is shown in Table I [27].

Table I
Comparison Between Some Of The Quality Detection Tools [27]

Tool	Coverage	Error Type	Fix Suggestion
Keep Right	World	Many(50+	No
Osmose	Some	Many(200	Yes
JOSM	Local	Many	Yes
OSM Inspector	World/Partial	Many	No
Maproulette	World/Partial	Many(20+	No

3) Monitoring tools

There are also a number of tools that helps in spotting erroneous changes and edits. The following is the list of some monitoring tools for quality assurance:

- ***Open street map Watch List (OWL)***
  OWL provides a service for monitoring, processing and storing changes in OSM data. The main feature of OWL is that small tiles are attached with changes instead of calculating the bounding box of changes in a changeset which are very useful for visualisation i.e. only relevant changes are shown,not all changes [31].
- ***Who Did It***
  WhoDidIt is an OSM Changeset Analyzer which analyzes which is changed in an area. For this, ”month” or ”half a year” is selected for ”age” parameter to see changes done in particular area.
- ***Live Edit Map Viewer J***
  LiveEditMapViewerJ is a Java Program where live changes being done on world map can be seen based on Planet.osm.diff.
- ***OSMZ miany***
  OSMZ miany is also an monitoring tool but it has more features than LiveEditMapViewerJ like loading changes from past, filtering options based upon bounding box and user, etc [27].

III. CONCLUSION

In this study, we have reviewed various quality aspects of spatial data, work that has already been done for improving quality of OSM data, and different tools for quality enhancement of OSM data set. It can be concluded that due to varying levels of knowledge or experience of contributors or due to absence of their professional qualifications or background, it has resulted in inconsistencies in OSM data. It may also because of different means are being used for collecting geodata by different contributors. This all can create flaws while doing analysis using OSM data. Since now there are so many tools available for improving OSM data quality e.g., ‘Keep Right’ application created by OSM community presents the positions of possible mistakes against a set of pre-defined rules for data conformance. It can also be concluded that the contributors from urban areas such as in European countries, US and UK are providing higher quality work as compared to other countries.

REFERENCES

[1] M. Haklay and P. Weber, “Open street map: User-generated street maps,” *Pervasive Computing, IEEE*, vol. 7, no. 4, pp. 12–18, 2008.

[2] M. Haklay, “How good is volunteered geographical information? a comparative study of Open street map and ordnance survey datasets,” *Environment and Planning B: Planning and Design*, vol. 37, no. 4, pp. 682–703, Aug. 2010.

[3] S. Sehra, J. Singh, and H. Rai, “A Systematic Study of Open street map Data Quality Assessment,” in *2014 11th International Conference on Information Technology: New Generations (ITNG)*, pp. 377– 381, Apr. 2014,.

[4] M. F. Goodchild, “Citizens as sensors: the world of volunteered geography,” *GeoJournal*, vol. 69, no. 4, pp. 211–221, Nov 2007.

[5] T. Hayakawa, Y. Imi, and T. Ito, “Analysis of Quality of Data in Open street map,” in *2012 IEEE 14th International Conference on Commerce and Enterprise Computing (CEC)*, pp. 131–134, Sep. 2012.

[6] “Elements - Open street map Wiki,” accessed on 5th June 2016. [Online]. Available: http://wiki.open street map.org/wiki/Elements

[7] "Open street map Statistics," accessed on 27th May 2016. [Online]. Available: http://www.open street map.org/stats/data stats.html

[8] P. Neis, D. Zielstra, and A. Zipf, "The street network evolution of crowdsourced maps: Open street map in Germany 2007–2011," *Future Internet*, vol. 4, no. 4, pp. 1–21, Dec 2011.

[9] O. Kounadi, "Assessing the quality of open street map data," *University College of London, Department of Civil, Environmental And Geomatic Engineering*, pp. 9–21, 2009.

[10] A. Coote and L. Rackham, "Neogeographic data quality is it an issue?" in *the AGI Conference*. ConsultingWhere Ltd, 2008.

[11] A. L. Ali and F. Schmid, "Data Quality Assurance for Volunteered Geographic Information," in *Geographic Information Science*, ser. Lecture Notes in Computer Science, M. Duckham, E. Pebesma, K. Stewart, and A. U. Frank, Eds. Springer International Publishing, no. 8728, pp. 126–141, Sep. 2014.

[12] v. P. Oort, "Spatial data quality: from description to application," Ph.D. dissertation, Wageningen University, 2006.

[13] *Not at Home on the Range: Peer Production and the Urban/Rural Divide*. California, USA: ACM, 2016. [Online]. Available: http://hdl.handle.net/1942/20228

[14] O. Roick, J. Hagenauer, and A. Zipf, "OSMatrixgrid-based analysis and visualization of Open street map," *Proceedings of the State of the Map Europe, Vienna, Austria*, vol. 1517, p. 1517, 2011.

[15] "Open street map Data Quality" accessed on 4th June 2016. [Online]. Available: http://osmlab.github.io/osm-data-quality/

[16] X. Qian, L. Di, D. Li, P. Li, L. Shi, and L. Cai, "Data cleaning approaches in Web2.0 VGI application," in *2009 17th International Conference on Geoinformatics*, pp. 1–4, Aug. 2009.

[17] R. Karam and M. Melchiori, "A crowd sourcing-based framework for im- proving geo-spatial open data," in *2013 IEEE International Conference on Systems, Man, and Cybernetics*. Institute of Electrical & Electronics Engineers (IEEE), Oct 2013.

[18] L. M. V. Blzquez, B. Villazn-Terrazas, V. Saquicela, A. d. Len, . Corcho, and A. Gmez-Prez, "GeoLinked data and inspire through an application case," in *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 446– 449, Jan. 2010.

[19] P. Hashemi and R. A. Abbaspour, *Assessment of Logical Consistency in Open street map Based on the Spatial Similarity Concept*. Springer International Publishing, 2015.

[20] F. Ying, P. Mooney, P. Corcoran, and A. Winstanley, "Polygon Processing on Open street map XML Data," in *Proceedings of the GIS Research UK 18th Annual Conference*, pp. 149–154, 2010.

[21] A. Vandecasteele and R. Devillers, "Improving volunteered geographic information quality using a tag recommender system: The case of Open street map," in *Lecture Notes in Geoinformation and Cartography*. Springer Science mathplus Business Media, pp. 59–80, 2015.

[22] C. Brando and B. Bucher, "Quality in user generated spatial content: A matter of specifications," in *Proceedings of the 13th AGILE international conference on geographic information science*. Springer Verlag: Guimar aes, Portugal, pp. 11–14, 2010.

[23] A. L. Ali, F. Schmid, R. Al-Salman, and T. Kauppinen, "Ambiguity and plausibility: managing classification quality in volunteered geographic information," in *Proceedings of the 22nd ACM SIGSPATIAL Interna- tional Conference on Advances in Geographic Information Systems*. ACM, pp. 143–152, 2014.

[24] M. Helbich, C. Amelunxen, P. Neis, and A. Zipf, "Comparative spatial analysis of positional accuracy of Open street māp and proprietary geodata," *Proceedings of GI Forum*, 2012.

[25] B. Ciepluch, P. Mooney, and A. Winstanley, "Building generic quality indicators for Open street map," 2011.

[26] A. Pourabdollah, J. Morley, S. Feldman, M. Jackson, and J. Campus, "Studying the Dynamic Patterns of Open street map Bugs in Great Britain," in *Proceedings of 16th AGILE International Conference on Geographic Information Science, Leuven, Belgium*, vol. 1417, 2013.

[27] “Quality assurance - Open street map Wiki,” accessed on 24th May 2016. [Online]. Available: http://wiki.open street map.org/wiki/Quality assurance

[28] “Osmose - Open street map Wiki,” accessed on 26th May 2016. [Online]. Available: http://wiki.open street map.org/wiki/Osmose

[29] “OSM Inspector - Open street map Wiki,” accessed on 26th May 2016. [Online]. Available: http://wiki.open street map.org/wiki/OSM Inspector

[30] “MapRoulette - Open street map Wiki,” accessed on 27th May 2016. [Online]. Available: http://wiki.open street map.org/wiki/MapRoulette

[31] “OWL – Open street map Wiki,” accessed on 27th May 2016. [Online]. Available: ht.tp://wiki.open street map.org/wiki/OWL