The Quality Assessment of OpenStreetMap

MANINDER PAL SINGH
Research Scholar(M.Tech I.T.)
GNDEC, Ludhiana, Punjab
ermaninderpalsingh85@gmail.com

SUKHJIT SINGH SEHRA
Assistant Professor
GNDEC, Ludhiana, Punjab
sukhjitsehra@gmail.com

HARDEEP SING RAI
Professor
GNDEC, Ludhiana, Punjab
hardeep.rai@gmail.com

JAITEG SINGH
Associate Professor
CIET, Punjab
jaitegkhaira@yahoo.co.in

***Abstract*—Just a few year ago mapping was primarily used in cars for navigation but now it enables everything from lucrative location lased services to game changing autonomous driving in era of internet of things. Revolutionary computer technology given us method of sharing and computing data contributed by users called Crowdsourcing. This method give rise to user generated geographical information system such as OpenStreetMap (OSM). The problem with OSM is because of loose coordination and no top-down quality assurance processes and little knowledge about contributors their skills and devices used for mapping. So there is big challenge to the quality of OSM. There is a challenge to the quality of OSM. This paper focuses on the quality of assessment of OSM data as compare to Ordnance Survey data, we have assessed the logical consistency of both datasets and it has been concluded as crowdsourced data has got more no features mapped as compare to Ordnance survey data and comparable quality. But with lots of logical errors, It may not suitable for navigation purposes.**

Keywords- ***Crowdsourcing, OpenStreetMap, Map Data Inconsistencies, Quality Assessment.***

1. INTRODUCTION

The past recent years have proved that vast accessibility and easy handiness of geographical technologies and spatial data is increased which is available on public map websites provided by Google, Microsoft, My Maps, Wikimapia , Flickr including geo-data that is truly free and open workable projects like OpenStreetMap. In crowdsourced map data, many volunteers over the world collect map data from various sources and make it available for users. In addition there are various websites developed under the crowdsourced revolution are most of them supports open source development. The merit of this type of website is that they can update preinstalled map database and provide access for new developing world that is incomplete yet. One such platform is OpenStreetMap (OSM) [1].

If OpenStreetMap provides free data of world, but still there is price attached to it. The price is the data quality, we are now living in the world of internet of things (IoT). The internet of things going to depend heavily onto map to show them onto map using their geographical location. One of rare example of an interactive map made from Nokia map tiles is thingful search engine [12]. It is a search engine for the Internet of Things, providing a unique geographical index of connected objects around the world, including energy, radiation, weather, and air quality devices as well as seismographs, iBeacons, ships, aircraft and even animal trackers. When private map providers are putting huge money to this era. We must make crowdsourced OpenStreetMap future ready so that it can also be used for IoT kind of projects. So to do it, the maps should maintain high quality standard. But problem with OSM is loose coordination, no top-down quality assurance processes and little knowledge about contributors their skills and devices used for mapping. So there is big challenge to the quality of OSM. It is required that the OSM data must be compared and assessed for its quality with ground data.

This paper is an endeavor to compare the OSM data with Ordnance Survey data for data quality parameters. The next section introduces OSM and its data structure. Then section 2 discusses about OpenStreetMap project. Section 3 & 4 elaborates the quality parameters used and data acquisition respectively. Section 5 and 6 discusses about methodology used and results. Last section discusses conclusion.

1.1 OPENSTREETMAP

OSM is one of the best known sources of Volunteered Geographic Information (VGI) [13]. Most of the OSM data is gathered by amateurs and non-specialist contributors. Basically OSM is free editable mapsheet of the world developed by volunteers all along the world by collecting and contributing geographic data. Every individual contribute to the map by simply registering on the OSM website.OSM worksheets has many favorable aspects like anyone can easily obtain vector data freely, users can improvise and up to date the map which is very useful when the regions are mapped poorly. OSM use three types of objects - points (nodes), lines (ways), and relations. Lines contain large amount of points, and the line convey the attributes that define what it represents. Polygons are represented as lines, which finishes where it begins in order to generate a polygon. These three objects are main constitute for generating OSM geodata. This data is collected by using various handheld devices which are GPS enabled [6]. For adding their collected geodata on OSM website the user must register on it. Before uploading that data it converts it into a digital format using offline / online editing tools like JOSM and Potlacth. By using OSMwiki page beginner level

user got instructions 'How to collect and upload their data in OSM website'. User can initiate by gathering GPS measurements. The android mobile GIS program named as OSMTracker and iOS based mytracks application will help to collect the GPS trails. Moving around nearby area, the data is gathered as GPS coordinates. After collecting all the data it will be ready to upload in a computer. That data will be in GPX format.GPX format is a standard XML format used to exchanging Geodata between various applications and is used to upload GPS traces on OSM website. The OSM track application can directly upload the traces to OSM, If user has added his OSM login credentials. These GPS traces are then digitized by using Potlacth online editing tools [6].

TABLE- 1-OPENSTREETMAP STATS REPORT RUN AT 2015-05-13 [9]

Sr.No	Facts	Stats.
1	Number of users	2094739
2	Number of uploaded GPS points	4638931546
3	Number of nodes	2854574891
4	Number of ways	286843946
5	Number of relations	3349533

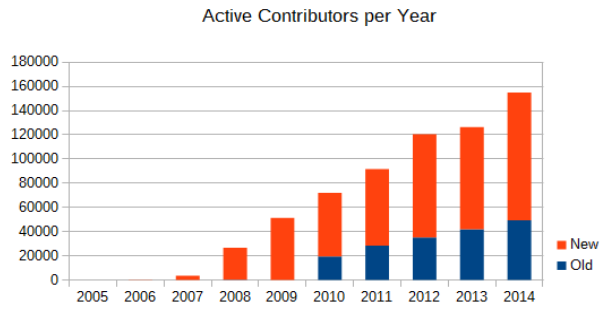


Figure 1. Active contributors per year [9]

Because OSM is open source project, that's why it gains huge popularity among peoples. Many new users are joining this community. Table 1 shows the huge interest gaining towards OSM. It gives statistics of active new and old contributors per year shown in Fig. 1. Now the users can be beginner, novice and expert too. What type of information they are adding, is a major issue. Is that information (map-data) is fit for further use or not?

2. QUALITY PARAMETER FOR ASSESSMENT

In year 2002, quality facets of geographical data enclosed in the International Organization for Standards (ISO) codes 19114 (Quality evaluation procedures) and 19113 (Quality principles) below the supervision of Technical Committee [10]. In their appraise of these standards recognized the following facets of quality: completeness, logical consistency, positional accuracy, temporal accuracy, thematic accuracy, purpose, usage and lineage [11]. It carries together a variety of quality standards and definitions are more comprehensive and identify the following facets:

I. **Lineage** – This facet of quality is about the historical record of the dataset, how it was acquired.
II. **Positional accuracy** – This facet of quality assess how good is the coordinate value of an object in the database related to the realism on the land.
III. **Attribute accuracy** – Objects in a geospatial database are depicted not only by their geometrical shape but also by supplementary attributes; this facet assesses how truthful these values are?
IV. **Logical consistency** – This is a facet of the internal consistency of the geospatial dataset, in terms of topological rightness and the relationships that are enclosed in the database.
V. **Completeness** – This is a facet of the lack of data, i.e. an evaluation of how many objects are predictable to be found in the database but are lost as well as an evaluation of overload data that should not be included.
VI. **Semantic accuracy** – This facet links the way in which the object is captured and represented in the database to its meaning and the mode in which it should be translates.
VII. **Usage, purpose and constraints** – This is a fitness-for-purpose affirmation that should help prospective users in deciding how the spatial data should be used.
VIII. **Temporal quality** – This is a facet of the legitimacy of changes in the database in relation to real-world changes and also the tempo of updates.

In this paper we focused on the Logical Consistency parameter

3. DATA ACQUISATION

The two datasets of same region is taken for testing purpose. Region is Greater London; datasets are OpenStreetMap (OSM) and Ordnance Survey (OS). The total length of OS roads is 14838882 m whereas OSM has 25173590 m. OSM is crowd-sourced map data which can be obtained from Overpass API, Planet OSM, Geofabrik downloads and Metro extracts whereas the Ordnance Survey is digital data provided by Government geo-spatial entities. For analysis the data preprocessing is necessary also dataset must be in .shp file format. The OS provide dataset in tiles. That tile contain various other cities, we need to clip our desired region. Clipping is done with the reference of administrative boundary data provided by Ordnance Survey using Quantum Geographic Information System (QGIS) [7]. "Fig. 2" shows the shape files of both dataset.

a. Greater London (Ordnance Survey)
b. Greater London (OpenStreetMap)

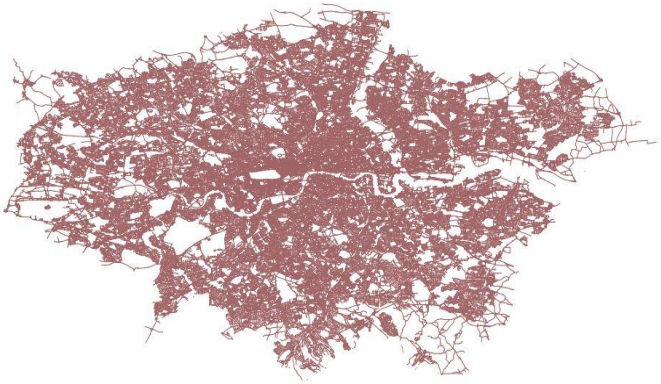
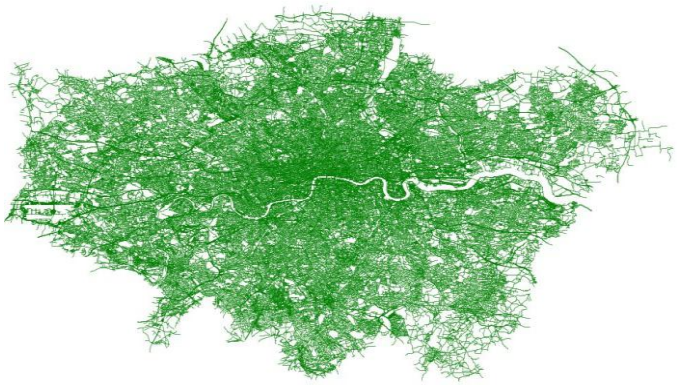


Figure 2 a. Greater London (Ordnance Survey)



b. Greater London (OpenStreetMap)

TABLE 2- ORDANANCE SURVEY (OS) TOTAL FEATURES- 165201

Sr.No.	Roads	No. of Attributes Feature
1	Primary Roads	7273
2	Primary road collapsed dual carriageway	2429
3	A roads	20377
4	A road collapsed dual carriageway	2272
5	B road	7175
6	B road collapsed dual carriageway	238
7	Local Street	102742
8	Minor Roads	20762
9	Minor road collapsed dual carriageway	528
10	Private road publically accessible	1405

TABLE 3- OPENSTREETMAP (OSM) TOTAL FEATURES- 191475

Sr.No.	Roads	No. of Attribute Feature
1	Primary Roads	9914
2	Secondary Road	2788
3	Trunk	5931
4	Residential Road	62515
5	Footway	40715
6	Service Road	34563
7	Cycleway	4165
8	Unclassified Roads	10108
9	Tertiary	7254
10	Motorway	615
11	Bridleway	702
12	Living_Street	148
13	Path	4330
14	Pedestrian	2197
15	Steps	3694
16	Track	1898
17	Unsurfaced	10

It is clear from table 2 and 3 that OpenStreetMap (OSM) has more feature as compare to Ordnance Survey (OS) also much more clear classification of Roads.

4. METHODOLGY

In this paper the logical consistency of the geo-spatial data is evaluated. By logical consistency we mean the inconsistent spatial data like line intersection, incorrect polygons closing, line duplicity or gaping in lines. It focuses on the structural integrity of the datasets [4].

The Logical errors which occur during digitizing process of the data are also called Semantic errors. We analyzed node mismatches, Undershoots error, Overshoot error, overlap segments, illegal geometries and coincident segments. These logical errors are shown in "Fig. 3".

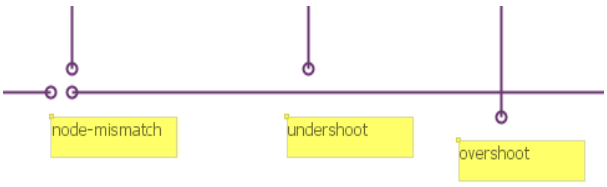


Figure 3- Various Logical Errors

- **Node mismatches**-nodes closed to each other but not exactly the same

- **Undershoot Error**- Such topological errors the end of a line falls short and is disconnected near the end of the neighboring line.
- **Overshoot error**- Such topological errors where nodes extend too far.

Map must be topologically correct else it is not useful in terms of navigation. To examine geospatial data for topological consistency shape file format (.shp) files are required by the algorithm used for analysis. The algorithm applied for the uncovering topological errors is a plugin from OpenJump. OpenJump is an open source Geographic Information System (GIS) uses Java programming language. OpenJump is part of the Java Topology Suite (JTS), a Java API that implements robust geometric algorithms and provides an absolute model for specifying 2-D linear geometry [5].

Micro-segments are the closely mapped points which hinder the GIS analyst to get precise map data. We tested both datasets for the same as shown in "Fig. 4" and "Fig. 5". Few micro-segments are found during analysis. Network Topology cleaning algorithm detects the node mismatch, overshoot and undershoot error within a distance threshold 3.0 m between the nodes. Both datasets are evaluated for the same as shown in "Fig. 6" and "Fig. 7". Coverage cleaning algorithm detects the undersized defects like overlaps and gaps. Here matching of each segment of polygon is done with neighboring polygons. It ensures that angle between both segments is less than the user defined tolerance also minimum distance between segments is less than user defined threshold [8]. Both datasets are evaluated for the same as shown in "Fig. 8" and "Fig. 9". The quality assurance algorithm detects illegal geometries such as shapeless designs and unclosed circles. The coincident segments are also detected by this algorithm. The coincident segments are those when one polygon has same tolerance as second polygon.

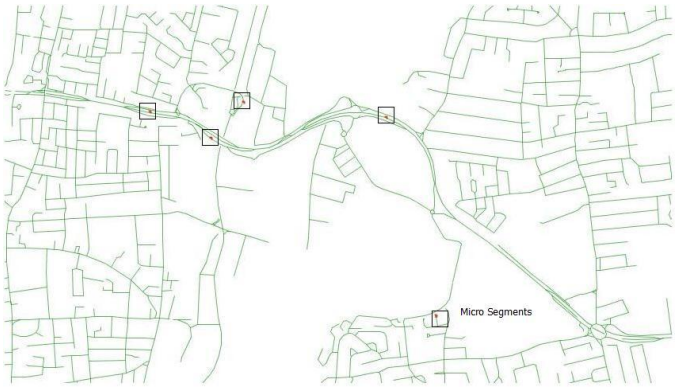


Figure 4. Closely mapped Segments / Micro Segments Detected in Ordnance Survey

Figure 5. Closely mapped Segments / Micro Segments Detected in.OpenStreetMap

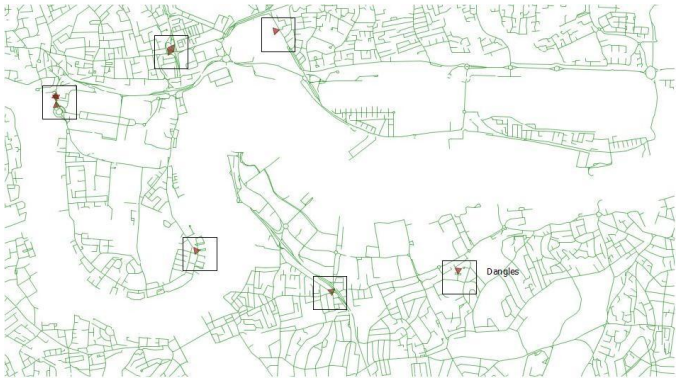


Figure 6. Dangles (Undershoot, Overshoot,Node Mismatch) in Ordnance Survey

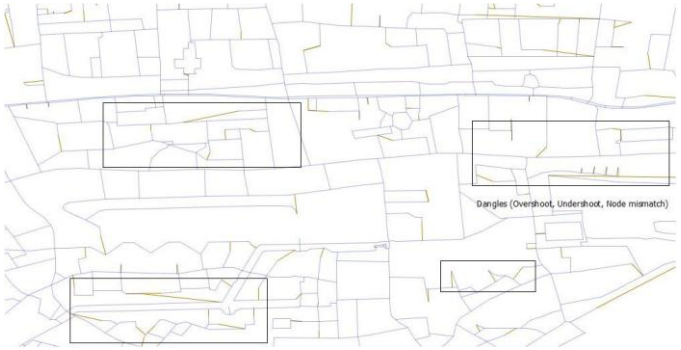


Figure 7. Dangles (Undershoot, Overshoot,Node Mismatch) in OpenStreetMap.

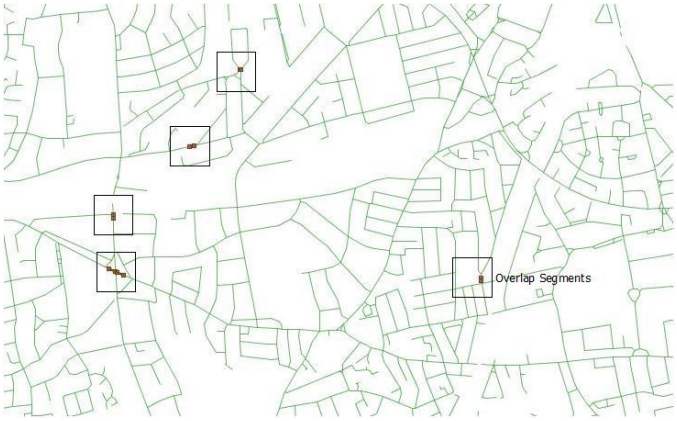


Figure 8. Coverage overlaps Segments in Ordnance Survey

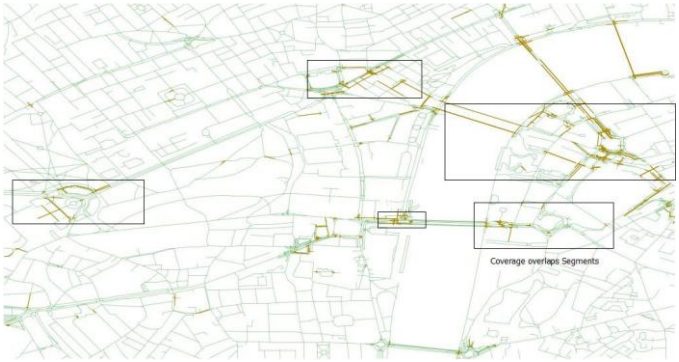


Figure 9. Coverage overlaps Segments in OpenStreetMap

5. RESULTS

The investigation of both datasets OS and OSM is done using OpenJump topological plug-in. The total data information in OSM and OS data are discussed in table 2, 3 respectively. Both dataset were processed by various algorithms for finding the logical errors discussed above, which produce following results describe in table 4.

TABLE 4. VALUE OF BOTH DATASET ON VARIOUS PARAMETERS.

Parameters	OSM dataset(Features)	OS dataset(Features)
Closely mapped segments removed.	0	47
Dangles (Undershoot, Overshoot, Node Mismatch)	59249	101
Coverage overlaps Segments.	14663	476794
Illegal Geometry	1291	19
Coincident segments.	87	244

6. CONCLUSION

OpenStreetMap is created by diverse users of different background. This crowd sourced computation system uses loose coordination and works without top down quality assurance background. This paper is an effort to bring forward the comparative analysis of OpenStreetMap (OSM) with Ordnance Survey (OS). OSM is developed with minimum cost and effort involved. Whereas the OS is governmental agency put in huge amount to map there region. The above facts describes that OSM dataset has 1.157 times more features as compare to OS dataset. Nearly 10,000 km more roads mapped in OSM dataset than OS data. Despite having more dangles than OS.

After analysing for error in both datasets this found that the crowd contribution has yields significant benefit to society by creating maps like OSM , which offers free map data of world with no cost and provides good quality map data too.

7. REFERNCE

[1] A .Ather, "A Quality Analysis of OpenStreetMap". Msc. Dissertation, 2009.

[2] P. Neis, and A. Zipf . "Analyzing the contributor activity of a volunteered geographic information project—The case of openstreetmap". ISPRS International Journal of Geo-Information 1.2 20121.pp. 46-165, 2012.

[3] P.Neis, M. Goetz and A.Zipf, "Towards automatic vandalism detection in openstreetmap" Geoinformatics Research Group, Department of Geography, University of Heidelberg, Germany. pp. 315-319, 2012.

[4] T. Ubeda and M. Egenhofer."Topological error correcting in GIS. In Advances in Spatial Databases", Michel Scholl and Agnès Voisard (Eds.). Lecture Notes in Computer Science, vol. 1262.Springer Berlin Heidelberg 2012, pp. 281–297, 2012.

[5] S. Steiniger and A.J.S. Hunter .Open JUMP HoRAE–A free GIS and toolbox for home-range analysis. Wildlife Society Bulletin, vol. 36, issue-3, pp. 600-608, 2012.

[6] http://www.OpenStreetMap.org.(2014)

[7] Quantum, G. Development Team (QGIS). Quantum GIS Geographic Information System, Version 1.7.0. Open Source Geospatial Foundation Project. Vancouver. British Columbia, Canada. 2011.

[8] M. Michaud, Personal Website: Topology Library. 2014 http://geo.michaelm.free.fr (accessed on 18 April, 2015)

[9] http://www.openstreetmap.org/stats/data_stats.html OpenStreetMapStatistics. (2014). (accessed on 9 May, 2015)

[10] W.Kresse and K.Fadaie, ISO Standards for Geographic Information, Springer , 2004, XI, p. 323, 2004.

[11] V. Oort, Spatial data quality: from description to application, PhD Thesis, Wageningen: Wageningen Universiteit 2006.

[12] http://www.thinkful.com (accessed on 20 May 2015)

[13] M.F. Goodchild, "Citizens as voluntary sensors: spatial data infrastructure in the world of web 2.0 ". International Journal of Spatial Data Infrastructures Research, vol. 2, pp. 24-32, 2007.

Author's Profile

1 Maninder Pal Singh
Research Scholar (M.Tech I.T.)
100377482738
GNDEC Ludhiana.

2 Sukhjit Singh Sehra is working as an Assistant Professor at Guru Nanak Dev Engineering College, Ludhiana, India. He is an OpenStreetMap contributor. His areas of research are assessment of spatial databases, spatial data mining and high performance computing.

3 Hardeep Singh Rai is a Professor at Guru Nanak Dev Engineering College, Ludhiana, India. His areas of research are OpenStreetmap, BRL-CAD, BIM./Structural Design, GIS, CAD, BIM and FOSS.

4 Jaiteg Singh is working with Chitkara University as an Associate Professor in the School of Computer Science. His area of research includes data mining, data quality and ETL testing in datawarehouses