

Review of Contributor's Categories and their Motivations for Open Street Map

^[1] Pawan Verma ^[2] Sukhjith Singh Sehra, ^[3] Kapil Sharma
^[1] PG Student, ^[2]^[3] Assistant Professor

^[1]^[2]^[3] Department of Computer Science & Engineering, Guru Nanak Dev Engineering College, Ludhiana, Punjab, India,

^[1] pawanverma1612@gmail.com, ^[2] sukhjitsehra@gmail.com, ^[3] kapilsharma701@gndec.ac.in

Abstract: — Open Street Map(OSM) is a global map data that is collected by the volunteers who are involved with OSM project while roaming all around the world. It is a geographic data which describes the feature of a real world object. This study investigates the various OSM data elements, OSM editing tools, motivation factors in contribution, data quality parameters and types of contributors in OSM project. As Web 2.0, Volunteered Geographic information(VGI) and Global Positioning System(GPS) devices are introduced, any registered user of OSM can create and edit the OSM data globally. Basically, there are three tools JOSM, Potlatch 2 and iD that are used to contribute to OSM. As the number of contributors are rising, data quality of geographic data is also improving. There are various motivation factors like knowledge about area where contributor lives which compels the contributors to contribute more and more about that area. Some classification rules are used to categorize the user based on the quantity and quality of data.

Index Terms—Contributor, Open Street Map, Volunteered Geographic information, Web 2.0

I. INTRODUCTION

OSM project is initiated in 2004 by Steve Coast with the idea of creating a map of whole world with the help of volunteer who collect the data across the world [1], [2]. Contributors are playing a great role in collecting geographic data for OSM. With the use of Web 2.0, VGI and GPS devices any registered user can easily contribute to OSM, so citizens are acting as a sensor for collecting geographic data [3], [4].

There were approximate 2 million registered user at the end of December 2011, but active users are only 24,000 which are only 5% of total users [5]. At the end of May 2016, there are approximate 2.7 million (26, 97,409) registered users [6], but registered users are not equal to the active users. There are basically three components of OSM that describes the conceptual model of the real world. These elements are Nodes, Ways and Relations [7]. A registered user of OSM can create and edit these elements. There are various OSM editors used for editing OSM data but top three editors are: iD, Potlatch 2, JOSM [2], [8]. Contributors have good knowledge about their surrounding, so most of the information provided by them is of useful quality [3]. When there are many numbers of contributors that are working on a specific spatial unit then quality of data is good. If contributor do not follow any quality assurance

program and there is not any central coordination between contributors then quality of spatial might be of low quality [9]. There are various elements which define the quality of spatial data. These elements are Lineage, Positional accuracy, Attribute accuracy, Logical consistency, Completeness, Semantic accuracy, Usage, Temporal quality, Variation in quality, Meta-quality, Resolution [10].

Motivation factors are also playing a great role in collecting useful and large amount of geographic data. like individual's knowledge about their surroundings [11] or individual's belief to provide free information to improve the world's geographic data [2] are some motivation criteria that helps in success of OSM project.

A. Web 2.0

In the early web there was only one-directional services in which user can view the sites. Web 2.0 is developed to provide two way communication in which user can view the data placed at remote site and he/she can also send the data to remote server. So user acts as a volunteer in collecting the data [12]. Wikipedia and OSM are the most common examples of web 2.0. There are small number of reviewers that verify the uploaded data by volunteers [3].

B. Crowd sourced Map Data

In crowd sourcing various people from any corner of the world work on a single project. The Crowd sourced Map data is a collection of geographic data collected by the

people all around the world. Each person contributes information to the map [13].

C. Volunteered Geographic Information

Volunteered geographic information (VGI) is using of tools for creating, assembling, and scattering geographic data provided voluntarily by individuals. VGI is a kind of user generated content. Examples of this phenomenon include OSM, Wiki Mapia and Google Map Maker [14]. Improvements in web based mapping, mobile networks, GPS devices, and wiki technologies have introduced significant changes in availability, volume and nature of geographic information. The critical and participatory approaches to GIS are extended which can be seen as VGI. Only general base map information is given by these sites for creating content regarding various features or events which exists but are not shown on the base map. Although VGI is creating the geographic data at large scale but it has major concerns in quality and reliability of data [15].

II. LITERATURE SURVEY

Good child [3] says that every person in the world acts as sensor. With the advent of Web 2.0, VGI and GPS devices, people are participating in collection of large data. Most of the Information provided by volunteers is of useful quality. but there is not proper mechanism to detect and remove errors and to ensure quality. Haklay and Weber [2] identified that in earlier year, cartographers were only responsible for creating the maps and display it on paper. But nowadays a common person with the little computer knowledge can create and edit the digital map. Various editing tools are provided by OSM to work on the editing of map.

Antoniou et al. [16] suggest that there is need to examine issues like the quality, credibility and fitness of geographic data because any user can insert and edit the data. "Open Street Map does not have any content restrictions on tags that can be assigned to Nodes, Ways or Areas. You can use any tags you like". A set of rules are created to identify the inaccuracies in OSM data. There are common editors such as Potlatch 2 and JOSM which provide high data quality and systematic way to edit data. Haklay et al. [9] says that the spatial data provided by the volunteers is of not good quality because they don't follow some quality assurance programs. There is not coordination among the creators of OSM data because they don't follow specific procedure to collect data. The quality of spatial data also depends on the number of contributors working on that data. Linus' Law is fully applicable to OSM. When there are 15 contributors in per square kilometre then good positional accuracy is received within 6 metres. The first 5 contributor in an area contribute most of the data. The number of people joined with the project receive special attention, it is called

Linus's law. Positional accuracy is used to evaluate Linus' Law on OSM. There are large number of contributors so there is need to evaluate Linus's Law for betterment of spatial data quality. Elwood et al. [15] says that from when the VGI is introduced there is extraordinary shift in the quality, quantity, characteristics and variety of geographic information. This shift has changed the way of creating, editing, distribution and use of geographic data. In VGI most of people acts as geographer. VGI data is very much helpful in answering the questions raised by researchers.

Neis and Zipf [5] says that the active number of contributor on OSM is very less than the total registered users. It is concluded that only 38% of the registered member has created only one changeset and only 5% of contributor are the active contributor to OSM data. Registered members are increasing in numbers continuously. Members are divided into three groups Junior mappers, Nonrecurring mappers and senior mappers based on the quantity of data provided to OSM. Most contributor only edit the OSM data within 50 km of place where they resides and they actively participate only for first 3 month. Arsanjani et al. [17] proposed that the categorization of a contributor can be done based on the quantity and quality of data provided to OSM by them. According to quantity there are five mappers: "Beginner", "Regular", "Intermediate", "Expert", and "Professional mappers". To categorize based on quantity, a number of conditional rules are created. There are various quality parameters to describe quality of data like "Positional accuracy", "Completeness", and "Semantic accuracy" etc. Neis et al. [18] says that contributors in urban areas are most active participant in VGI and provide quality data to OSM. An analysis is done on OSM data of 12 urban areas. It says that in urban areas there is good coverage of OSM with regards to completeness of data. European cities are covering most of data of OSM.

III. OSM DATA ELEMENTS

Data elements describe the physical properties of real world objects. It acts as a conceptual model for OSM data. Various tags are associated with each data items. There are three OSM data elements:

A. Node

Node is a specific point on a map which represent the features of that point. Each node has latitude and longitude associated with it. A way can also be represented by referring various nodes. For example, Amenity=atm represents a node as a atm.

B. Way

A polyline or a straight line which consists of 2 to 2000 nodes is called a way. Roads, rivers etc. are the types of way which are linear in nature. Boundaries of a object

can also be represented using ways such as buildings. When a way start and end at same node then that way is called closed way.

C. Relation

A relation is an ordered list of ways, nodes and other relations. These all are called member of a relation. So a relation is used as multi-purpose data element that contains two or more than two data elements like ways, nodes and other relations. Exmaple of relation are:

- 1) A route relation that contains the major parts of a road like cycle route, bike route or a bus route.
- 2) A turn restrictions that says you can not turn from one way to other way.

These three data elements have tag associate with them. Tag: Each data elements(nodes, ways or relations) in OSM is associated with some tags. The meaning of particular element can be identified by the tag associated with it. Each tag has two text fields: 'key' and 'value'. This text field can hold 255 characters. For example, building = house describe that the building is house of someone. There are many tags available on OSM that represent the common objects of the world [7].

1) Common Attributes: There are many attributes for nodes, ways and relations. But we need not to use all of them. The common attributes are:

TABLE I

I. COMMON ATTRIBUTES OF OSM TAG [7]

Name	Value	Description
id	integer	it is unique id associated with data elements to identify them uniquely.
user	string	it display the name of user who have modified the object recently.
uid	integer	it is unique user id who have recently modified the object.
timestamp	date and time format	it display the date and time of last modification
visible	true or false	when the object is removed from database. if visible = false then object can only be accessed by history call.
version	integer	the edit version of the element
changeset	integer	it is the unique id when the object is created or modified.

Changeset: When we are going to create or modified any of above data elements then each time a new changeset is created. Like other data elements changeset has tag associated with it.

Comment tag is considered as recommended tag while creating changeset. It shows the cause that why you are going to create changeset. The maximum number of edit on a changeset is 50000. There should not be edit request on changeset within one hour of recent edit [19]. We can use HTTP PUT and GET on changeset.

IV. OSM EDITING TOOLS

OSM community developed a set of editing tools to contribute OSM data. There are more than 27 OSM editing tools available nowadays for different platforms by different organisations [8]. Top three OSM editors are:

- 1) iD
- 2) Potlatch 2
- 3) JOSM

A. iD

iD is the open source OSM editor which is programmed using JavaScript language. It is available under the editing option on www.Open Street Map.org. It is a light weight editor and uses d3js as a layer for rendering. For managing OSM data it uses modular core which is very fast.

B. Potlatch 2

Potlatch 2 is advanced version of Potlatch editor series. Potlatch 2 is written in Action Script 3 using the open source Flex framework. It is also available under the editing option on www.Open Street Map.org. Various features of Potlatch 2 are:

- ❖ _ Undo/redo support
- ❖ _ OAuth support
- ❖ _ WYSIWYG rendering
- ❖ _ Vector Background layers

C. JOSM

JOSM stands for Java Open Street Map Editor (JOSM). More experienced users use JOSM because of the advanced functionality provided by it. Using JOSM, user can import, edit and tag offline OSM data and user can easily upload bulk amount of data using OSM application programming interface (API). JOSM supports resolution of data conflicts and supports linking of photos and audio with OSM data [2].

V. SPATIAL DATA QUALITY PARAMETERS

It is not necessary that the data provided by the volunteers is of good quality. Their knowledge about particular spatial point and quality assurance program provide great impact on quality of spatial data. Linus's law is fully applicable to OSM. Positional accuracy is used to evaluate Linus' Law on OSM [9]. Oort [10] gives the detailed view on data quality parameters. These parameters

decide whether the spatial data is of good quality or not. Contributor need to know all these quality parameters to understand spatial data efficiently.

Various quality parameters are:

- 1) Lineage
- 2) Positional accuracy
- 3) Attribute accuracy
- 4) Logical consistency
- 5) Completeness
- 6) Semantic accuracy
- 7) Usage, purpose, constraints
- 8) Temporal quality
- 9) Variation in quality
- 10) Meta-quality
- 11) Resolution

VI. MOTIVATIONS

Motivation can be easily understand from the experience while working with open source communities like Wikipedia and OSM. Motivation puts a great effect on contributor. It answers the question, Why people should actively contribute to OSM project? Coleman et al. [20] proposed various motivation factors which strongly affect the contribution made by contributor. These factors are:

- _ Altruism: Contributor contributes purely for benefit of others.
- _ Professional or Personal interest: If contributor is contributing in job or his personal interest in contribution.
- _ Intellectual Stimulation: While contributing improvement in knowledge, technical skills and experience.
- _ Enhancement and securing personal investment
- _ Social Reward: When there are social rewards for constructive contribution.
- _ Enhanced Personal Reputation: To make a reputed personal recognition across the world.
- _ Pride of Place: While adding group or community information from where contributor belongs is good for public relation.

There are some negative motivations like Mischief, Agenda and Malice or criminal intent [20]. Negative motivation can cause serious problems. There are various miscellaneous positive factors. Like when someone sees other contribution on map then he can easily get motivated or individuals knowledge about specific place makes him good contributor of data about that place or people contribute for their leisure. In free time they contribute most of the data [11].

VII. CATEGORIZING THE CONTRIBUTOR

Sometimes it is necessary to categorize the contributor based on certain criteria. By categorization one can easily check his rank, level and values. Coleman et al.

[20] proposed three types of individual in which a user can fall. These types are "Market-Driven", "Social Networks" and "Civic/Government". When a professional contribute for commercial organization it serve the purpose of "Market-Driven". When contribute contributes for open source communities like OSM or Wikipedia for welfare for human being. When someone acts as concerned citizen and works for government, a member of social group or animal rights group. Coleman et al. [20] identified the five types of contributors. Table II gives the detailed view on this. Arsanjani et al. [17] is able to categorize the contributor based on some quality parameters of spatial data that is collected by contributor. Quality parameters used to categorize contributor are "Positional accuracy", "Completeness" and "Semantic accuracy" etc. A set of conditional rules are applied based on the quality and quantity of data to categorize contributor. There are five types of contributors [17].

- 1) Beginner mappers: Beginner mappers have little knowledge about coordinate system and they do not contribute regularly. Data provided by them is not complete data.
- 2) Regular mappers: Regular mappers are those who contribute the data and they do not concern about semantic and positional accuracy of objects.

	Market-Driven (Example: GPS Navigation)	Social Networks (Example: Restaurant Reviews)	Civic/Governmental (Example: PPGIS)
Neophyte	No experience with GPS positioning, map reading or navigation	Reads and uses review information provided at a given Website without question.	Citizen views a GIS map in a town hall meeting around the siting of a power plant in the town
Interested Amateur	Owms a personal system, uses it extensively, and is aware of key strengths and limitations	Regular viewer; Occasional contributor to restaurant review Website.	Citizen fashions a map to present a counter claim in a town hall meeting around the siting of a power plant in the town
Expert Amateur	Familiar with the strengths and weaknesses of multiple systems, has owned more than one	Is familiar with a variety of different sites and can rank them in terms of reliability. Regularly contributes reviews.	Individual familiar with conditions in a given neighborhood and with the operation of the Web-based PPGIS system in use.
Expert Professional	Surveying or mapping professional specializing in GPS positioning	Paid Restaurant Reviewer for Local Newspaper	Practicing Urban Planner
Expert Authority	Noted specialist regularly consulted re: specific problems and/or new developments by other professionals.	Restaurant Reviewer for national media chain with extensive professional experience in hospitality and food services.	City Planner with extensive knowledge of developments in the area of interest.

Table II
Examples of contributors in each category along the spectrum [20]

- 3) Intermediate mappers: Whenever contributor has free time then they contribute most of the geographic data. They take care of positional accuracy and do not mind semantic accuracy.
- 4) Expert mappers: Expert mappers are those when they have local knowledge about location where they live and data provided them is best to his/her knowledge. They pay attention to both semantic and positional accuracy.
- 5) Professional mappers: professional mappers contribute most of the data and the data provided by them is of useful quality. They share the data best of their knowledge [17]. Neis [21] categorizes the contributor based on quantity of data provided by contributor. Changeset data is used for this purpose. Table III describes that what type of mapper you are based on changeset created by contributor.

Table III
What Type Of Mapper You Are? [21]

Type of Mapper	Changeset created
Hit-and-run Mapper	< 25
Newbie	< 100
Casual Mapper	< 500
Heavy Mapper	< 1000
Addicted Mapper	< 5000
Crazy Mapper	< 10000
Epic Mapper	< 30000
Bot	> 30000

Neis and Zipf [5] has also categorized the contributor based on nodes created by contributor. There are four types of mappers "Senior mappers", "Junior mapper", "Nonrecurring mappers" and "No Edits". Senior mappers are those who have created more than 1000 nodes. Members who have created nodes between 10 to 1000 are called "Junior mappers". Members who have created less than 10 nodes are referred as Nonrecurring Mappers. Last group is "No Edits", who have not created a single node. Members with "No Edits" are very large than other categories. So there are various criteria to categorize the contributor. Contributor should provide data of large quantity and of good quality if they want high rank among other contributors.

VIII. CONCLUSION

In this study, we have reviewed OSM data elements, OSM editing tools, motivation factors in contribution, data quality parameters and types of contributors in OSM project. From this study it is concluded that active numbers of mappers are very less than registered mappers. Only 5% mappers contribute the data regularly. If more people participate actively then they can create huge map data and of good quality. While creating the data, contributor should be fully aware of data element that he/she is going to create, because each OSM data element has its own characteristics. Contributor can easily create these data elements if contributor has good knowledge about OSM editing tools. Motivation is the another solution for contributor to contribute data constructively with the use of editing tools. Positive motivation is necessary for welfare for human being. While contributor are contributing OSM data from different places of world, each contributor need to be categorize based on quantity and quality of data provided by them. There are various criteria to categorize the contributor. By categorization, one can easily know his contribution rank and then get inspired to contribute more and more data to gain high rank and satisfaction.

REFERENCES

- [1] "About OSM," (accessed on 30 April 2016). [Online]. Available: https://en.wikipedia.org/wiki/Open_Street_Map
- [2] M. Haklay and P. Weber, "Open Street Map: Usergenerated street maps," *Pervasive Computing*, IEEE, vol. 7, no. 4, pp. 12–18, 2008.
- [3] M. F. Goodchild, "Citizens as sensors: web 2.0 and the volunteering of geographic information," *GeoFocus*, vol. 7, pp. 8–10, 2007.
- [4] —, "Citizens as sensors: the world of volunteered geography," *GeoJournal*, vol. 69, no. 4, pp. 211–221, 2007.
- [5] P. Neis and A. Zipf, "Analyzing the contributor activity of a volunteered geographic information project — the case of Open Street Map," *ISPRS International Journal of Geo-Information*, vol. 1, no. 3, pp. 146–165, jul 2012.
- [6] "OSM Statistics," (accessed on 27 May 2016). [Online]. Available: <http://osmstats.neis-one.org/>
- [7] "OSM Data Elements," (accessed on 30 April 2016). [Online]. Available: <http://wiki.OpenStreetMap.org/wiki/Elements>
- [8] "OSM Editors," (accessed on May 28 2016). [Online]. Available: <http://wiki.OpenStreetMap.org/wiki/Editors>

- [9] M. Haklay, S. Basiouka, V. Antoniou, and A. Ather, "How many volunteers does it take to map an area well? the validity of linus law to volunteered geographic information," *The Cartographic Journal*, vol. 47, no. 4, pp. 315–322, 2010.
- [10] v. P. Oort, "Spatial data quality: from description to application," Wageningen Universiteit, 2006.
- [11] N. R. Budhathoki, "Participants' motivations to contribute geographic information in an online community," Ph.D. dissertation, University of Illinois at Urbana-Champaign, 2010.
- [12] "Web 2.0," (accessed on 30 April 2016). [Online]. Available: https://en.wikipedia.org/wiki/Web_2.0
- [13] "Crowdsourced Map Data," (accessed on 30 April 2016). [Online]. Available: <http://www.imapbuilder.net/interactive-map-samples/crowdsourced-map-example.php>
- [14] "Volunteered Geographic Information," (accessed on 30 April 2016). [Online]. Available: https://en.wikipedia.org/wiki/Volunteered_geographic_information
- [15] S. Elwood, M. F. Goodchild, and D. Z. Sui, "Researching volunteered geographic information: Spatial data, geographic research, and new social practice," *Annals of the Association of American Geographers*, vol. 102, no. 3, pp. 571–590, may 2012.
- [16] V. Antoniou, M. Haklay, and J. Morley, "A step towards the improvement of spatial data quality of web 2.0 geospatial applications: the case of Open Street Map," in the 18th GISRUK Conference, London, UK, 2010, pp. 197–201.
- [17] J. J. Arsanjani, C. Barron, M. Bakillah, and M. Helbich, "Assessing the quality of Open Street Map contributors together with their contributions," in *Proceedings of the AGILE*, 2013.
- [18] P. Neis, D. Zielstra, and A. Zipf, "Comparison of volunteered geographic information data contributions and community development for selected world regions," *Future Internet*, vol. 5, no. 2, pp. 282–300, 2013.
- [19] "OSM Changeset," (accessed on 30 April 2016). [Online]. Available: http://wiki.Open_Street_Map.org/wiki/Changeset
- [20] D. J. Coleman, Y. Georgiadou, J. Labonte et al., "Volunteered geographic information: The nature and motivation of producers," *International Journal of Spatial Data Infrastructures Research*, vol. 4, no. 1, pp. 332– 358, 2009.
- [21] P. Neis, "Mapper Type," (accessed on 30 April 2016). [Online]. Available: <http://neis-one.org/2012/01/what-type-of-mapper-are-you/>