Quality Assessment of Volunteered Geographic Information for Educational Planning

By

Hafiz Muhammad Muzaffar

(NUST201362184MSCEE62513F)

A thesis submitted in partial fulfillment of the requirements for degree of Master of Science in Remote Sensing and GIS

Institute of Geographical Information Systems
School of Civil and Environmental Engineering
National University of Sciences & Technology
Islamabad, Pakistan

July, 2017

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS/MPhil thesis written by **Mr Hafiz Muhammad Muzaffar (Registration No. NUST201362184MSCEE62513F), of Session 2013 (Institute of Geographical Information systems)** has been vetted by undersigned, found complete in all respects as per NUST Statutes/Regulation, is free of plagiarism, errors, and mistakes and is accepted as partial fulfillment for award of MS/MPhil degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in the said thesis.

Signature: ___________________

Name of Supervisor: Dr. Ali Tahir

Date: 10-07-2017

Signature (HOD):_____________

Date: 10-07-2017

Signature (Dean/Principal): ___________

Date: ____________

ACADEMIC THESIS: DECLARATION OF AUTHORSHIP

I, Hafiz Muhammad Muzaffar declare that this thesis and the work presented in it are my own and have been generated by me as the result of my own original research.

Quality Assessment of Volunteered Geographic Information for Educational Planning

I confirm that:

1. This work was done wholly by me in candidature for an MS research degree at the National University of Sciences and Technology, Islamabad.
2. Wherever I have consulted the published work of others, it has been clearly attributed.
3. Wherever I have quoted from the work of others, the source has been always cited. With the exception of such quotations, this thesis is entirely my own.
4. I have acknowledged all main sources of help.
5. Where the work of thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.
6. None of this work has been published before submission. This work is not plagiarized under the HEC plagiarism policy.

Signed: ……………

Dated: ……………..

DEDICATION

"To my loving mother & father who taught me how to read and write"

"Also to my Thesis Supervisor who helped me throughout my thesis. May Allah Bless them All."

ACKNOWLEDGEMENTS

All praises to Allah to whom all knowledge belongs and who is knower of unknown. All thanks to Allah who blessed me with guidance and knowledge to achieve this. His boundless mercy and blessings are foremost and everything. My all gratitude belongs to Him and only to Him.

My thanks to my family who believed in me and prayed for my success especially, my mother who kept me motivated in my journey to completion. I acknowledge all the sacrifices they made in terms of time, finances and my unavailability for them. My very special gratitude for my mentor Dr. Ali Tahir who not only supervised me for the work but also be kind enough to extend his support in every problem I came across. His farsightedness and visionary judgment made this day possible. Also it was only due to his personal efforts that I was able to submit a book chapter in an international book. May Allah reward him in return. My thanks to Miss Quratulain, Miss Khunsa Fatima and Mr. Munir Ahmed for guiding me and helping me out in fulfilling the thesis requirements. Their respect, courtesy and availability are highly acknowledged. May Allah bless them for their kindness.

I extend my special thanks to Dr Gavin McArdle, National University of Ireland & Michela Bertolotto , University College Dublin who co-authored a book chapter with me and made themselves available in completion and critique of my work. I am also thankful to all those whom I haven't mentioned here but they were part of their altruistic support directly or indirectly. I acknowledge all of their sincere efforts.

Hafiz Muhammad Muzaffar

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Volunteered Geographic Information (VGI) is the term used to describe the process of collecting spatial data using a network of volunteers. The approach collects spatial data to build maps which are often freely accessible. OpenStreetMap (OSM) and Wikimapia are examples of VGI. VGI has gained popularity within geospatial community due to its free of charge availability. In this study the quality of OSM data required for educational planning has been tested for Islamabad, Pakistan. OSM data has been assessed across three quality parameters i.e., data completeness, positional accuracy and attribute accuracy. For data completeness and attribute accuracy, OSM data has been compared with Federal Board of Intermediate and Secondary Education (FBISE) and Higher Education Commission (HEC) data while it has been compared with Survey of Pakistan (SoP) data for positional accuracy. It has been found that for positional accuracy OSM data has on average 5 meters inaccuracy, for data completeness it has 13% coverage, and for attribute accuracy it has 64% match. For positional accuracy OSM data satisfies the bench mark of at max 50 meters inaccuracy but for data completeness and attribute accuracy it does not satisfies the benchmarks of 100% coverage and at least 70% match respectively. It has been concluded that OSM data alone cannot be used for educational planning for Islamabad, Pakistan.

Chapter 1

INTRODUCTION

Crowdsourcing of data involves volunteers collecting, processing and sharing data through specialized digital platforms or via ad-hoc methods. There are many examples of crowdsourcing projects. Wikipedia (encyclopedia), OpenSignal (cell tower information), Open Food Facts (food information) and RootMetrics (cell carrier quality metrics) are examples of crowdsourced repositories covering multiple domains. Goodchild (2007) introduced the term Volunteered Geographic Information (VGI) to describe the crowdsourcing of map and spatial data. OpenStreetMap (OSM) is perhaps the best known and most successful VGI project. The aim of the ongoing project is to map the world and to provide free access to the editable map and underlying data. The project builds on the local knowledge of volunteers to create detailed maps.

As crowdsourced data are often contributed by non-expert volunteers, questions regarding its quality have been raised. In the spatial domain many studies assess the quality of OSM data (Ciepłuch et al., 2011; Haklay 2010; Girres & Touya 2010; Ludwig et al., 2011; Ballatore & Zipf, 2015; See et al., 2016). These studies assess OSM data for geometric accuracy, overlap, attribute accuracy and completeness. The quality assessment generally involves comparing the volunteered spatial data to official data sources such as data collected by national mapping agencies or mapping companies. Such studies assume that the professionals who have trained in map generation and cartography have produced an accurate ground truth. Although, the studies indicate that the quality of OSM is acceptable with accurate positional precision for road networks, the results also indicate that there are spatial and regional differences in relation to the quality of OSM data.

With the exception of route planning, studies assessing the correctness of volunteered spatial data are typically domain independent and focus on assessing the overall quality of OSM without drawing on specific spatial analysis tasks. This can hide the true value of VGI data for certain domains. For example, questions about the usefulness and quality of VGI for urban planning, infrastructure planning and service delivery have not been widely assessed. While such analysis is a subset of the overall quality assessment measurements, there may be subsets of the data where quality is better or worse than that seen in the overall OSM dataset.

In this thesis we investigate this by specifically examining the quality of OSM data for tasks related to educational planning. We use the city of Islamabad in Pakistan as a case study. We take the classic approach of comparing the spatial, semantic and thematic data of educational facilities in Pakistan to data collected by the National mapping agency in Pakistan (SoP). We augment this with other ground truth data which we collected as part of the study. We assess the volunteered geographic data for completeness, attribute accuracy and positional accuracy and report the results as they apply for educational planning in Pakistan.

1.1. OSM As VGI

VGI is the harnessing of tools to create, assemble, and disseminate geographic data provided voluntarily by individuals (Goodchild, 2007). The major goal of OSM is to create a free editable map of the world. The creation and growth of OSM has been motivated by restrictions on use or availability of map information across much of the world, and the advent of inexpensive portable satellite navigation devices. OSM is considered a prominent example of volunteered geographic information.

Created by Steve Coast in the UK in 2004, it was inspired by the success of Wikipedia and the predominance of proprietary map data in the UK and elsewhere. Since then, it has grown to over 2 million registered users, who can collect data using manual survey, GPS devices, aerial photography and other free sources. This crowdsourced data is then made available under the Open Database Licence. The site is supported by the OpenStreetMap Foundation, a non-profit organisation registered in England and Wales.

1.1.1. Structure of OSM

OSM uses a topological data structure, with four core elements (also known as data primitives):

- Nodes are points with a geographic position, stored as coordinates (pairs of latitude and a longitude) according to WGS 84. Outside of their usage in ways, they are used to represent map features without a size, such as points of interest or mountain peaks.
- Ways are ordered lists of nodes, representing a polyline, or possibly a polygon if they form a closed loop. They are used both for representing linear features such as streets and rivers, and areas, like forests, parks, parking areas and lakes.
- Relations are ordered lists of nodes, ways and relations (together called "members"), where each member can optionally have a "role" (a string).

Relations are used for representing the relationship of existing nodes and ways. Examples include turn restrictions on roads, routes that span several existing ways (for instance, a long-distance motorway), and areas with holes.

- Tags are key-value pairs (both arbitrary strings). They are used to store metadata about the map objects (such as their type, their name and their physical properties). Tags are not free-standing, but are always attached to an object: to a node, a way or a relation. A recommended ontology of map features (the meaning of tags) is maintained on a wiki.

1.1.2. OSM data sources

The main copy of the OSM data is stored in OSM's main database. The main database is a PostgreSQL database, which has one table for each data primitive, with individual objects stored as rows. All edits happen in this database, and all other formats are created from it.

There are a number of tools which facilitate to download OSM data. These include *Geofabrik*[1], *Gisgraphy*[2], and *Thinkgeo*[3]. We selected *Geofabrik* as this tool is more user-friendly. This is due to the fact that *Geofabrik* is updated on daily basis as opposed to *Gisgraphy* and *Thinkgeo*. Furthermore, *Geofabrik* provides data in multiple formats (pbf and shp) while *Gisgraphy* only provides pbf support. *Thinkgeo*, on the other hand is not completely free.

[1] http://download.geofabrik.de/

[2] http://www.gisgraphy.com/download/index.htm

[3] https://thinkgeo.com/

1.1.3. OSM size for Pakistan Vs. Other Countries

The OSM data for each continent, country and even for cities of some countries can be downloaded from *Geofabrik*. The size of the data for each country varies and it doesn't depend on the area of the country. Even the countries with less area have more size of OSM data then countries with more area. Figure 1.1 shows the size of OSM data in MBs of some of the countries of the world.

1.1.4. OSM role in Pakistan

There have been a number of disasters in Pakistan for which (Humanitarian OSM Team) HOT and OpenStreetMappers in general have responded by helping to produce/improve maps with the potential to provide useful maps to aid agencies. Disasters include

- 2013 Pakistan Earthquake
- 2010 07 Pakistan Floods

1.2. Statement of the problem

OSM is a well know VGI today. The project officially started in 2004 from England. The initial purpose of the project include just the road mapping but later on other spatial elements were added as per the need like POI, buildings and boundaries etc. As the data in OSM can be added / edited by any internet user, quality of OSM is a big question these days. Researchers from around the world are trying to assess the quality of OSM on regional basis. There are less studies exist where the quality of VGI has been assessed for a socio-economic phenomenon. There also exists only one research where the quality of OSM for Pakistan has been assessed. Hence there is full potential that the

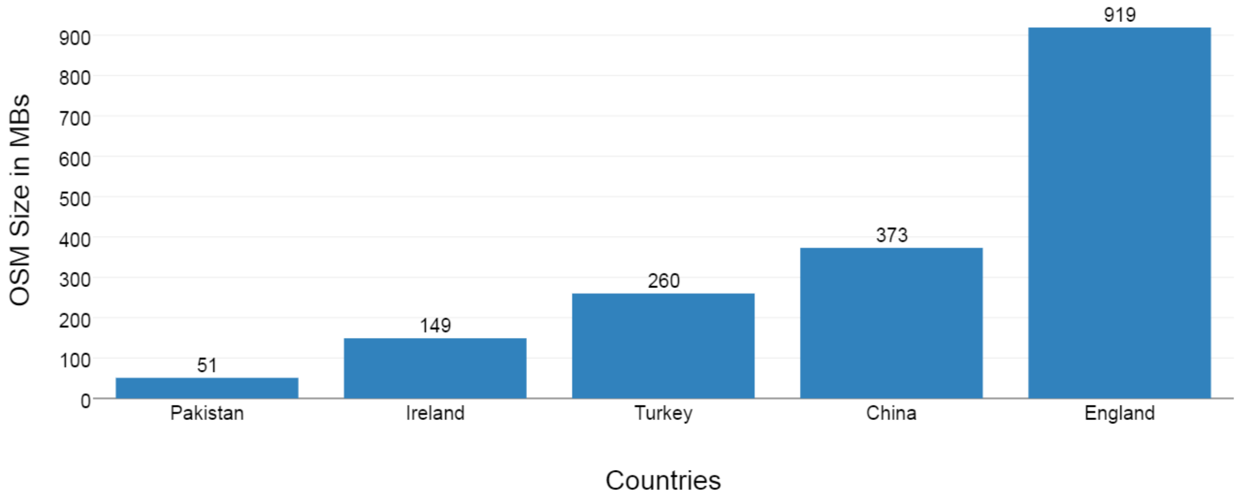

Figure 1.1. OSM data size of various countries.

quality of OSM data for Pakistan needs to be evaluated so that we have a bench mark to decide how better or worse it is to use OSM data for various purposes.

1.3. Scope of the study

It is very costly and hectic process to get data for any purposes especially in under developing countries like Pakistan. The reasons for this can be enormous. On the other hand there is plenty of data available online but there quality is objectionable. This research will determine the quality of OSM in Islamabad. In Pakistan, following fields could benefit from it.

a. Pakistan National Spatial Data Infrastructure (NSDI) is in process of developing which will link several government departments as well as Non-Government Organizations (NGOs) and Universities. If we assess the quality of OSM, it can be a part of this NSDI considering the fact that it has plenty of data which can be used as a base.

b. Education sector in Pakistan has been declining for many years and we need to do planning to boost the literacy rate. Planning needs the current infrastructure dataset. If we assess the quality of OSM for datasets related to educational planning we may be able to suggest to use this dataset for educational planning

c. There are many organizations in Pakistan which are working on various social projects and on the way they collect and store related data. Once we evaluat the quality of OSM, we can suggest them to use this dataset and we can encourage them to help improve the weaker part of this dataset.

Hence there is a need to assess the quality of OSM for Pakistan as this would result in an increased awareness toward the usage of open source data in the geospatial community.

1.4. Literature review

In this research we are assessing the quality of OSM, a VGI, for educational planning. So, we have divided literature review in two parts. The first part would cover the usage of VGI for planning purposes and the second part cover the quality assessment of OSM in various parts of the world.

1.4.1. VGI in Planning

Modern planning has multifaceted dimensions (Hall, 1996). Educational planning can be described as an exercise to establish an education system in order to accomplish the goals of policy makers. While educational planning is a broad subject, for the research presented in this thesis, educational planning concerns the setup of individual educational institutions in an area of interest by exploiting the power of spatial data. Public and private sectors need to establish educational institutions in areas where populations do not have easy access to the existing educational institutions. The main purpose behind establishing a new institution is to facilitate people and improve literacy rate. Okpala (2001) underpinned the importance of spatial information for practical and sustainable human settlement planning. Thus, planning to establish educational institution necessitates various spatial and non-spatial datasets. In Pakistan most of the spatial datasets held with public sector organizations are not shareable or accessible due to certain institutional and technical barriers like a lack of data policy, a lack of coordination, the high cost of data, the non-existence of data standards and infrastructure, non-availability of metadata as well as fear of exposure of data quality as identified by Ali & Ahmad, (2013). Moreover public sector data is not updated regularly due to updating cycle policies enacted in the country. This situation demands the use of free and accessible datasets like VGI in planning. There is immense potential of using VGI in planning (Elwood et al., 2012; Campagna et al., 2015) and its use in planning has been researched in several domain areas such as transportation, disaster management, environment, health, flood management and, flood response.

In order to support transportation planners for better decision making, a novel community-driven wiki based route analysis tool has been introduced (Masli et al., 2013). A large number of projects based on the potential use of VGI in disaster management have also been conducted. Poser & Dransch, (2010) elaborated on the opportunities and challenges associated with the use of VGI for disaster management especially in the context of response and recovery phases. The research was demonstrated with the help of a case study to rapidly estimate the flood damage using data collected from effected populations. Horita & Albuquerque, (2013) proposed a framework to integrate VGI and authoritative data for Spatial Decision Support System (SDSS) to support decision making in disaster management and to assist government organizations by providing mechanisms and information for better disaster management.

Similarly Hung et al., (2016) devised a method to estimate the credibility of VGI for disaster response. Datasets collected from two flood events in 2011 and 2013 from Brisbane, Australia and binary logistic regression were used to test the method. The result suggests the use of method for effective and rapid response as well as decision-making and coordination in emergency management. Degrossi et al., (2016) proposed a conceptual model that combines OSM, social media data and authoritative data in order to assess the quality of VGI. For the purpose of environmental monitoring Connors et al., (2012) conducted a case study in the Western United States to monitor a forest disease. The study makes use of a conceptual model to critically examine the OakMapper.org website that collects the spatial information pertaining to forest disease from volunteers. Authors argued that the proposed conceptual model can support environmental monitoring and help the public, researchers and policymakers collaborate to produce knowledge. Goranson et al., (2013) explore the value of VGI for public health research along with the potential challenges and ethical problems associated with the tools used for collecting VGI.

It is evident from the above discussion that VGI has been used for planning and decision making in a variety of domains and applications, however VGI for educational planning is not evident from the literature. The same methods and models as endorsed in the above discussion can be utilized to benefit educational planning.

1.4.2. Quality Assessment of OSM

OSM is a much cited example of Volunteered Geographic Information (Mooney et al., 2010). OSM is a project which aims to create an editable and non-proprietary map of the world. It is achieving this through the collaboration of millions of volunteers who map their local area and share their results via an editable map along with the underlying spatial data. Using volunteers to produce spatial data about their locality can produce very fine grained data which builds on local knowledge. However, given that volunteers are typically untrained in cartographic principles and spatial data collection methods, questions regarding the quality of data output via OSM and VGI have been raised. In order to measure the quality of a VGI dataset the parameters against which to check the quality of the dataset need to be identified. A first perspective about the quality of VGI has been established by Haklay (2010) where the quality of OSM data is discussed. Haklay (2010) adopted a way to check the quality of OSM data by comparing it with authoritative data. Authoritative data is considered to be accurate and as per ground reality as it is gathered in a systematic manner by professionals and passed through a quality test in a standardized way. Haklay (2010) took the data from Ordnance Survey (OS), an authoritative mapping organization in UK, and compared it with freely available OSM dataset. He checked the positional accuracy of roads considering motorways (A and B category roads) first by using a buffer method. He created a buffer against OS roads dataset of a specific distance and checked the length of OSM roads within the buffer. He concluded that on average, OSM data is accurate to about *6 meters*. The study proposed to continue research on other aspects of VGI data quality.

Several other researchers tested the quality of OSM data against the authoritative data in other countries. For example Ludwig et al., (2011) compared the OSM dataset against Navteq data in Germany. The authors matched the OSM roads with Navteq roads dataset in an automated way. They further concluded that at the country level there exist high level of similarity in the datasets while at detailed level there exist a difference in OSM and Navteq which further varies from town to town. Similarly, Ciepłuch et al., (2011) compared the quality of OSM data against authoritative data of Ordnance Survey Ireland (OSI). They calculated the length of all line features including roads, railways, rivers and trails etc. of OSM and the OSI dataset for Ireland. The same observations as Haklay (2010) were reported. De Leeuw et al., (2011) tried to investigate the role of local community in mapping the road network in Western Kenya. The authors used high resolution satellite imagery and got both the professionals and laymen to classify the roads with and without local territory knowledge. They observed that people with local area knowledge whether they are professional or not classified the roads with 92% accuracy. On the other hand, professionals and laymen who do not have local area knowledge classified roads data with 67.7% and 42.9% accuracy respectively. Hence, they concluded that data prepared by local community would be of higher quality other than that captured by general community.

Keßler et al., (2013) developed a concept to measure the quality of VGI data without authoritative data. They worked on the social part of VGI and assessed the number of people contributing in developing the VGI of a specific area and how often the data has been updated. They concluded that a high number of contributors and a high rate of altering the dataset meant the quality of data is good. However an increased number of corrections and revisions on a feature meant the quality of the data is lower. Similarly, D'Antonio et al., (2014) worked on the social aspect of data and tried to infer the quality of VGI data using VGI edit history. They presented a model to calculate the VGI user reputation and VGI feature trustworthiness as a standard to measure the VGI data quality. They concluded that data produced by a highly reputed user has a high quality.

Barron et al., (2014) presented a framework named iOSMAnalyzer which used the all history files of OSM data to evaluate the quality of OSM data in 25 different aspects. In order to evaluate the quality of OSM data based on the full history, iOSMAnalyzer divides its results in following use cases: General Information, Routing and Navigation, Address Search, POI Search, Map Application, User Information and Behavior.

Camboim et al., (2015) evaluated the quality of an OSM dataset in a heterogeneous area in Brazil. In the absence of a valid authoritative spatial dataset, the authors tried to compare the OSM data with a demographic dataset. They used the population count, average household income, Human Development Index (HDI) and other demographic datasets to evaluate the quality of OSM dataset in urban and rural areas. The results showed that the places where data is complete and is updated more frequently are usually the high density urban areas while the places where data do not exist are areas with a low population.

1.5. Why Educational Planning

As we know, VGI is associated with certain accuracy and quality issues as reported by Elwood et al., (2012) and others. Thus to effectively use VGI for planning purposes, it is essential to assess its quality against certain aspects like completeness, logical consistency, positional accuracy, temporal accuracy, thematic accuracy, purpose, usage and lineage (Kresse & Fadaie, 2004), (Oort, 2006) before its incorporation in the planning process. The same is also required for educational planning in the context of city of Islamabad. Educational planning is selected as a domain area because in Pakistan the education sector is declining due to poor management and monitoring (Memon, 2007). Thus it is essential to use educational planning as a special case to boost the education sector and hence the economy of the country.

1.6. Quality Assessment Parameters

As mentioned in literature, there are a number of parameters against which the quality of VGI data can be assessed. For example

- Positional Accuracy. This is probably the most obvious aspect of quality and evaluates how well the coordinate value of an object in the database related to the reality on the ground.
- Attribute Accuracy. As objects in a geographical database are represented not only by their geometrical shape but also by additional attributes, this measure evaluates how correct these values are.
- Completeness. This is a measure of the lack of data; that is, an assessment of how many objects are expected to be found in the database but are missing, as well as an assessment of the excess data that should not be included. In other words, how comprehensive is the coverage of real-world objects.
- Logical Consistency. This is an aspect of the internal consistency of the dataset, in terms of topological correctness and the relationships that are encoded in the database
- Lineage. This aspect of quality is about the history of the dataset, how it was collected, and how it evolved.
- Semantic Accuracy. This measure links the way in which the object is captured and represented in the database to its meaning and the way in which it should be interpreted.

- Usage. This is a fitness-for-purpose declaration that should help potential users in deciding how the data should be used.
- Temporal Accuracy. This is a measure of the validity of changes in the database in relation to real-world changes and also the rate of updates.

The main focus of this research is to assess the quality of VGI, in particular OSM, in the context of educational planning for the city of Islamabad, Pakistan. Data completeness, positional accuracy and attribute accuracy are the measure of quality selected for this case study. Data completeness is selected to provide detailed analysis about the coverage area of VGI in the city of Islamabad. Incomplete data does not reflect the true picture which is essential for educational planners to plan for the new educational facilities. Similarly, the attribute accuracy is important to check the consistency in the names and addresses of educational institutes. Finally, positional accuracy is selected in the context of educational planning to strategically plan the educational institutes in the city of Islamabad.

1.7. Objectives

The objective of this study is to assess the quality of OSM data for the capital of Pakistan, the Islamabad, for educational planning in three aspects:

- Completeness
- Positional
- Attribute

Chapter 2

MATERIAL AND METHODS

2.1. Study Area/Dataset

For this research, Islamabad, the federal capital of Pakistan has been selected as a study area. An important dataset for educational planning is the location of existing educational institution. This is used to assess the relationship between space and educational requirements. In addition to location, the attribute information is also relevant. The number of schools, colleges, universities, type of facilities, accessibility, teachers' skills, population, land use, water availability, and administrative boundaries all play a role in educational planning. For example, administrative boundaries are a crucial spatial dataset for addressing educational challenges by showing the scope and jurisdiction of local government bodies responsible for educational services. This can assist in identifying the places where immediate steps are required to improve educational deficits. Mapping and spatial analysis related to education can be performed within the administrative boundary of a local council. We have found that accurate administrative boundaries can be obtained from both VGI and authoritative data sources.

Data concerning the population in a given geographical area is required in order to assess the number of schools with respect to the number of pupils in that area. Furthermore, the identification of type of settlement is also crucial for educational improvements. For instance, industrial areas do not require schools as opposed to residential areas. Furthermore, the categorization of low, medium and high density areas is also required. Such population data is not typically collected using crowdsourcing and is more reliable if collected using authoritative data sources such as a national census.

The information about the road network in a particular area is very important. The better the road network, the better the accessibility. While planning a new educational facility, the existing road network can be used to identify optimal sites which are accessible and near to roads. The road network data are obtained from both VGI and authoritative datasets and their positional accuracy is also compared in our study. The price of land in a particular area is also a crucial requirement. Educational decisions are typically made in line with tight budgetary requirements. The affordability further enriches the geospatial educational data repository. Utilities such as electric, gas, water, telecom, sewerage also form core datasets which can be considered for locating new educational sites. These datasets are normally authoritative and are not shared with general public. Other Points of Interest (PoIs) are also important. This data can contain hospitals, police stations, bus stops, mosques, hostels, restaurants, railway station, airport, etc. POI data is largely contributed by volunteers based on their experience and can be easily obtained from VGI platforms such as OSM.

2.2. Data Collection

Four datasets belonging to the capital city were obtained. Schools and colleges data were obtained from Federal Board of Intermediate Secondary Education (FBISE). This data contain the list of schools and colleges in the capital. The Universities list was obtained from the Higher Education Commission (HEC). Settlements and road network

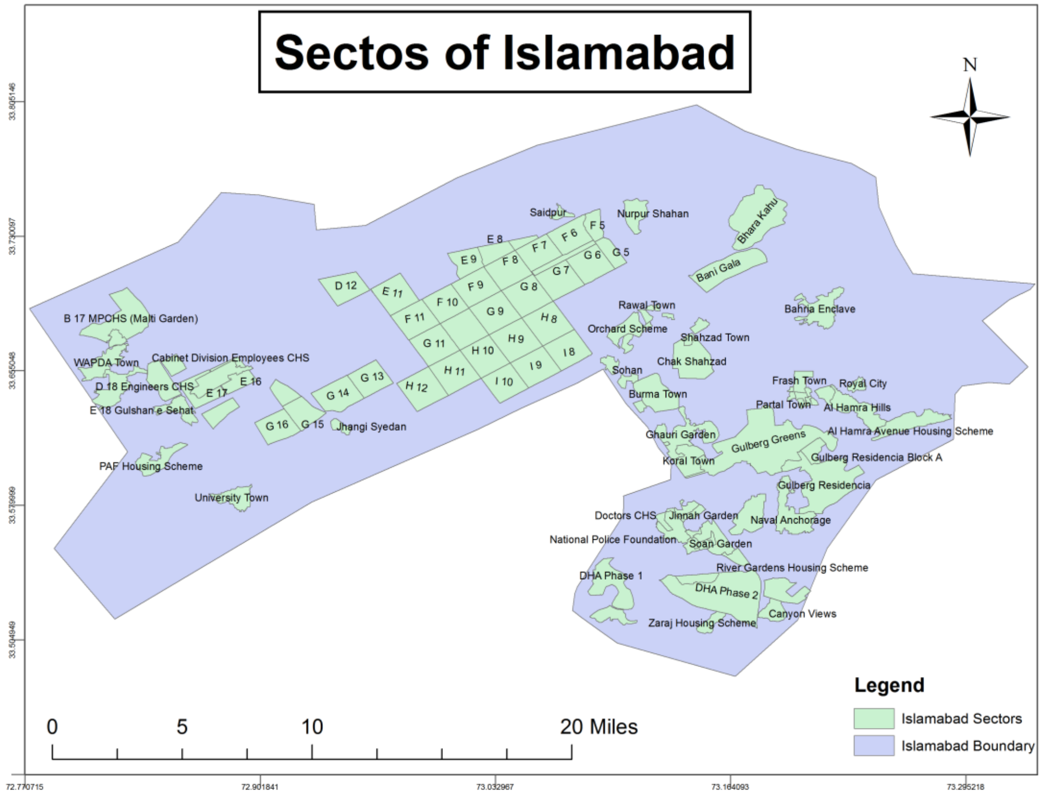

Figure 2.1. Study area - Islamabad

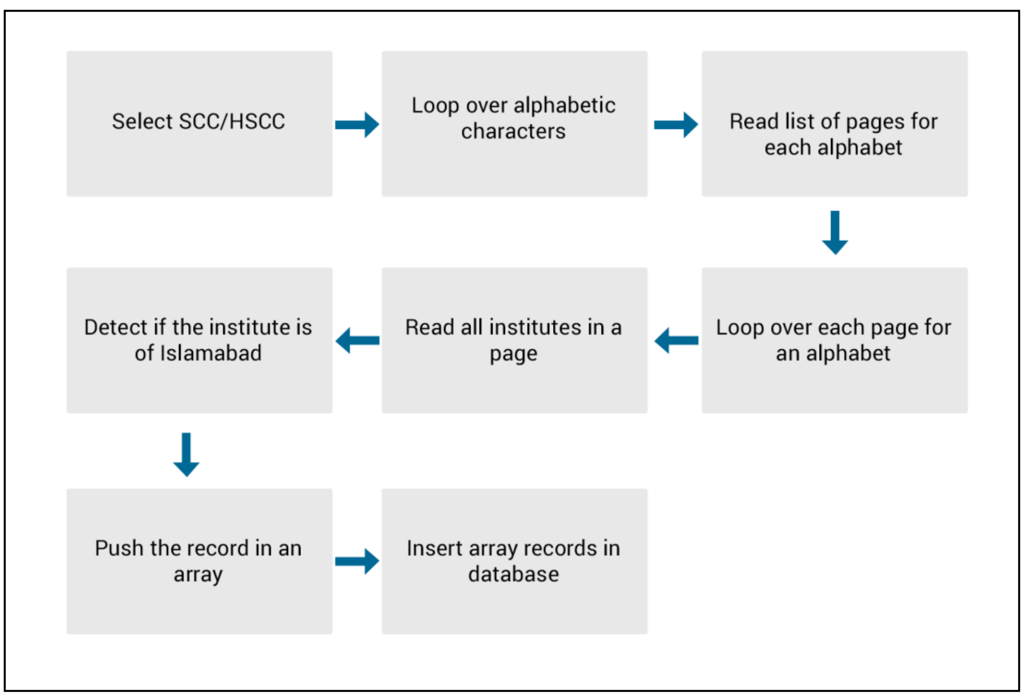

Figure 2.2. Scrape data.

were acquired from SoP. All these datasets are used for spatial analysis and are compared to VGI data.

2.2.1. Web Scrapping from FBISE website

FBISE has defined two categories for affiliated institutes. Either the institute is affiliated at Secondary School Certificate (SSC) level or at Higher Secondary School Certificate (HSSC) level. The attribute information such as institution name, address, affiliation date, website etc. of affiliated schools and colleges are also present on FBISE website. FBISE has affiliated institutes in Pakistan and worldwide. A script was written which scrapes and parses data from the FBISE database. The flow of the script is shown in Figure 2.2.

2.2.2. OSM Data Extraction

After extracting the required data from the FBISE data source, the next step was to collect the similar data from OSM. We got OSM data from *Geofabrik*. The downloaded data size for Pakistan is 28.1 MB as per (12-05-2016). Figure 2.3 outlines the steps of data extraction from *Geofabrik* into a local database for analysis.

2.3. Quality Assessment Methods

There are a number of methods which has been used in past to assess each parameter of quality assessment of VGI data. The methods which best fits in our case are discussed briefly.

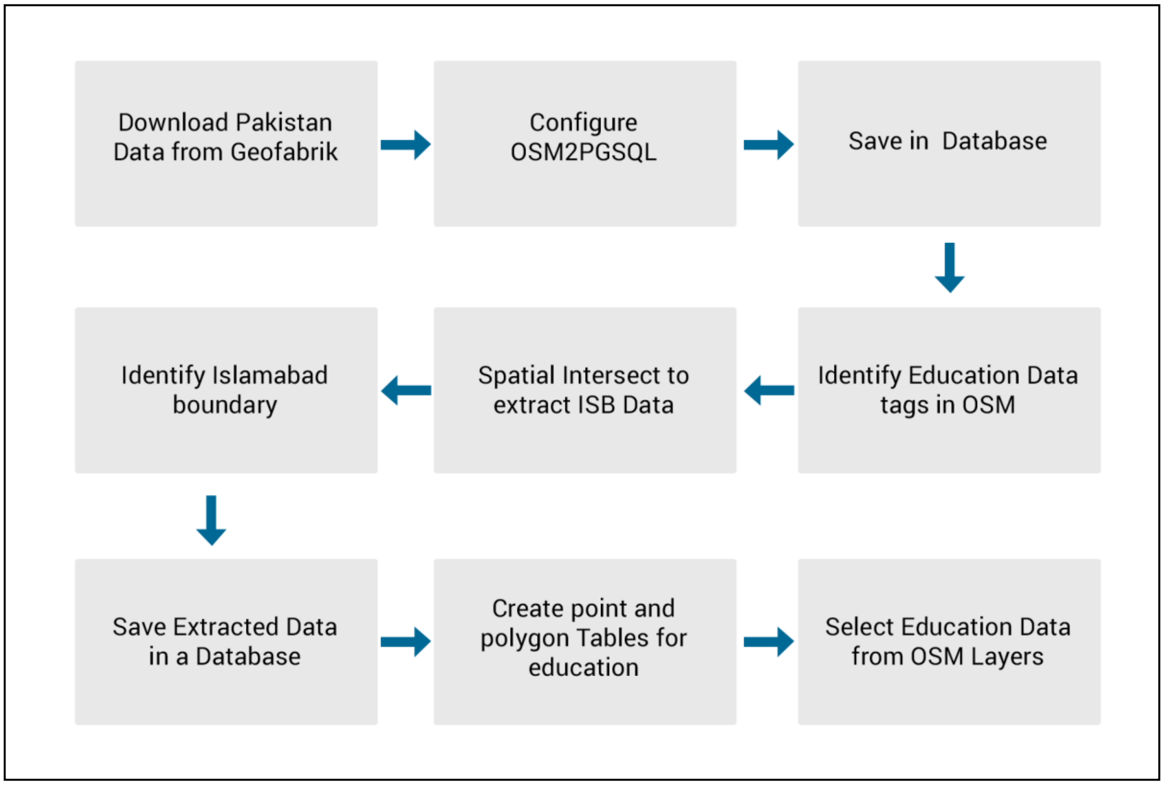

Figure 2.3. OSM data extraction.

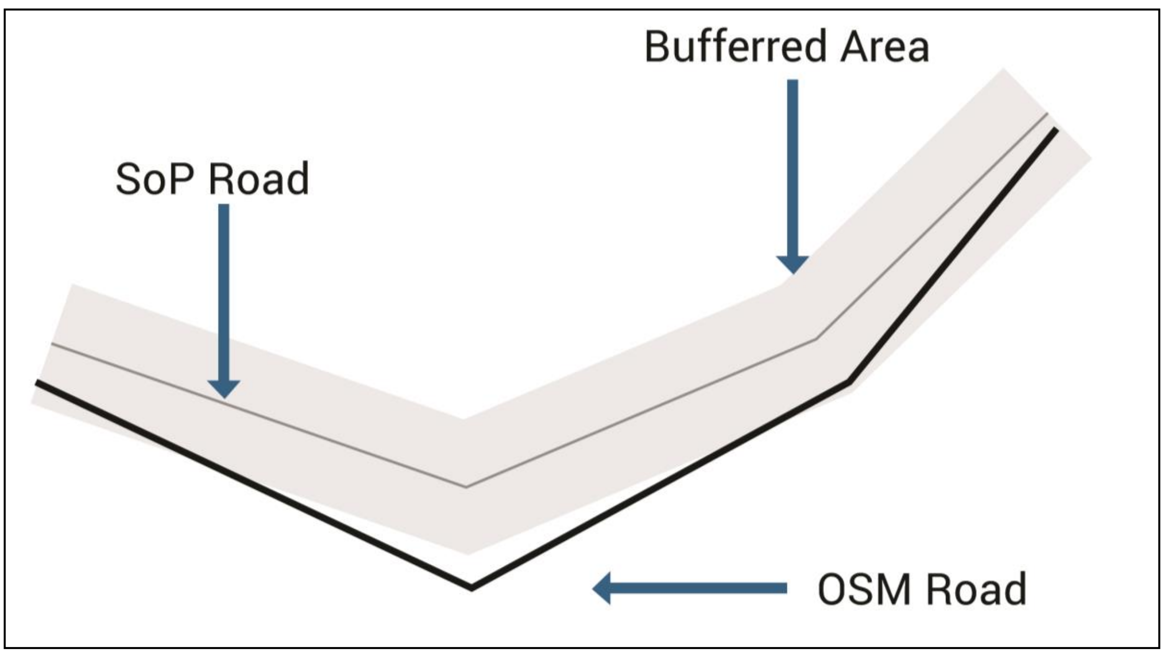

Figure 2.4. Buffered method for Positional accuracy.

2.3.1 Buffering Method

This method was purposed by Goodchild & Hunter (1997). It is used to assess the positional accuracy of two datasets. In this method a buffer is created against the true dataset and then its overlap with the tested data is calculated. We used this method to check the positional accuracy of roads. The concept is shown in the figure 2.4

2.3.2 Haversine Fourmula

The haversine formula determines the great-circle distance between two points on a sphere given their longitudes and latitudes. It is frequently used in navigation. We used this formula in positional accuracy parameter. We calculated distance between the OSM locations and the surveyed location with the help of this formula. The equations used are given below

$\alpha = osm\_lng - gps\_lng$

$\beta = osm\_lat - gps\_lat$

$a = sin^2(\beta/2) + cos(gps\_lng) * cos(osm\_lat) * sin^2(\alpha/2)$

$c = 2 * \tan^{-1}(\text{sqrt}(a), \text{sqrt}(1-a))$

$d = R * c$ (where R is the radius of the Earth)

2.3.3 Lavenshein Distance

In information theory and computer science, the Levenshtein distance is a string metric for measuring the difference between two sequences. Informally, the Levenshtein distance between two words is the minimum number of single-character edits (i.e. insertions, deletions or substitutions) required to change one word into the other.

We used Lavenshtein distance to measure the attribute accuracy. A lavenshtein distance greater than 7 is considered as an unmatched string.

2.3.4 Hausdorff Distance

In mathematics, the Hausdorff distance, or Hausdorff metric, measures how far two subsets of a metric space are from each other. Informally, two sets are close in the Hausdorff distance if every point of either set is close to some point of the other set. The Hausdorff distance is the longest distance you can be forced to travel by an adversary who chooses a point in one of the two sets, from where you then must travel to the other set. In other words, it is the greatest of all the distances from a point in one set to the closest point in the other set.

PostGIS, a spatial extension of PostgreSQL database, provides a function called *ST_HausdorffDistance* to calculate the distance between two features. The function returns the Hausdorff distance between two geometries. The units returned are as per the spatial reference system of the geometries. We used this function to calculate distance between NUST buildings of OSM and SoP data.

2.3.5 Bidimensional Regression

Bidimensional regression is a method for comparing the degree of resemblance between 2 planar configurations of points. The Euclidean version of bidimensional regression is

$$\begin{pmatrix} A' \\ B' \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} + \begin{pmatrix} \beta_1 & -\beta_2 \\ \beta_2 & \beta_1 \end{pmatrix} * \begin{pmatrix} X \\ Y \end{pmatrix}$$

The above equation has four parameters where two parameters capture the magnitude of the horizontal ($\alpha 1$) and vertical ($\alpha 2$) translation. The remaining two parameters ($\beta 1$ and $\beta 2$) are used to derive the scale (Φ) and angle (θ) values by which the original coordinates are transformed to derive the least squares fit. The scale transformation indicates the magnitude of contraction or expansion, and the angle determines how much and in which direction the predicted shape rotates with respect to the referent. A generic example of Bidimensional regression is shown in Figure 2.5. We used bidimensional regression to check the degree of resemblance between OSM and SoP road.

2.4. Research Approch

In this research, educational data is assessed across three principle axis, namely: completeness, attribute accuracy and positional accuracy. Figure 2.6 illustrates the methodology for quality assessment used in this study.

2.4.1 Data Completeness

Data completeness is an important element in the quality assessment of VGI for educational Planning. It is a measure of the degree to which features are present in both the VGI and authoritative data. VGI often do not contain complete coverage of existing educational infrastructure and therefore do not reflect the true picture for educational planners. For data completeness, the authoritative data of schools, colleges and universities were obtained from FBISE and HEC and the overlap between this ground truth and OSM is calculated. The concept of data completeness is presented in Figure 2.7.

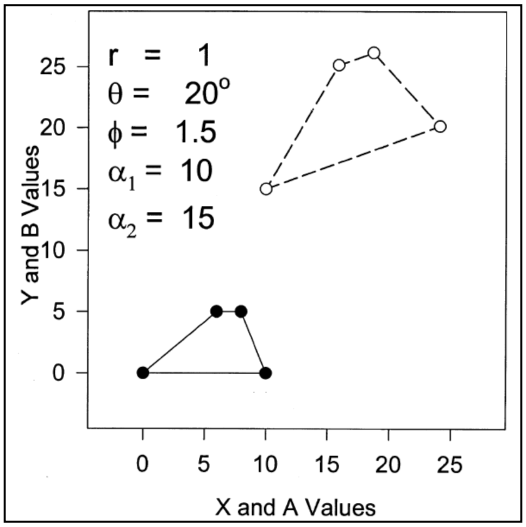

Figure 2.5. A generic example of Bidimensional regression.

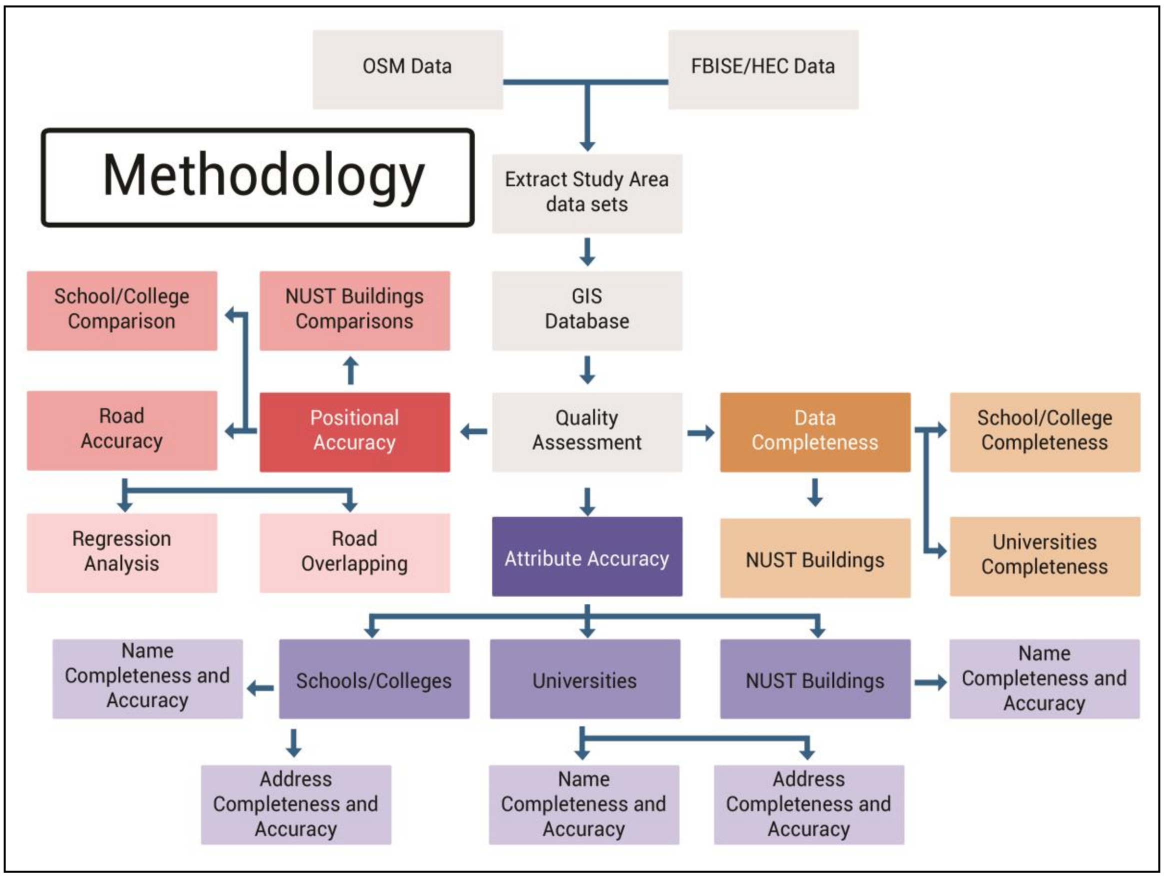

Figure 2.6. OSM quality assessment methodology for educational planning

For data completeness we assessed the quality in three dimensions. We checked the completeness for schools/colleges, universities and a test case for the buildings of NUST.

Data completeness for schools/colleges data were calculated using ArcGIS. The flow of the procedure is shown in the Figure 2.8

For universities, and the NUST buildings we calculated the data completeness manually since the data set was very small.

2.4.2 Attribute Accuracy

Attribute accuracy concerns the non-spatial characteristics of VGI. For educational datasets, it is important to check the consistency in the names, addresses and types of educational institutes. Therefore a comparison is made between the attributes listed in both authoritative and VGI data. The concept of attribute accuracy can be seen in Figure 2.9.

For Attribute accuracy, we assessed the quality in three aspects. We checked the name and address presence and correctness for schools/colleges and universities. While a test case for just name presence and correctness for the buildings NUST. As shown in Figure 2.10

We used Lavenshtein distance to check if the attribute part of OSM data is accurate or not. A levenshtein distance greater than 7 was considered as a mismatch string.

2.4.3 Positional Accuracy

Positional accuracy determines how accurately facilities are located within the

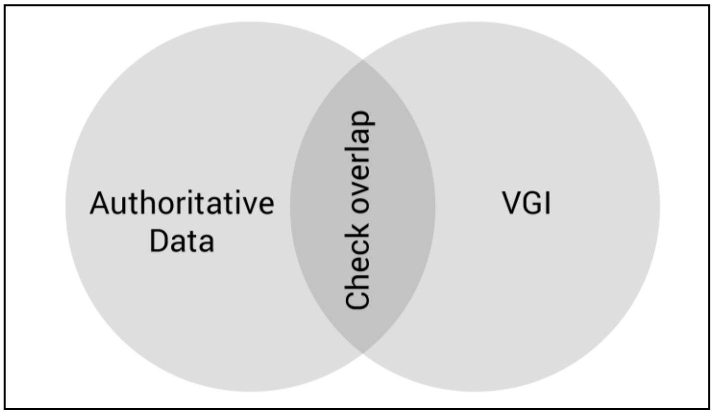

Figure 2.7. Concept of data completeness.

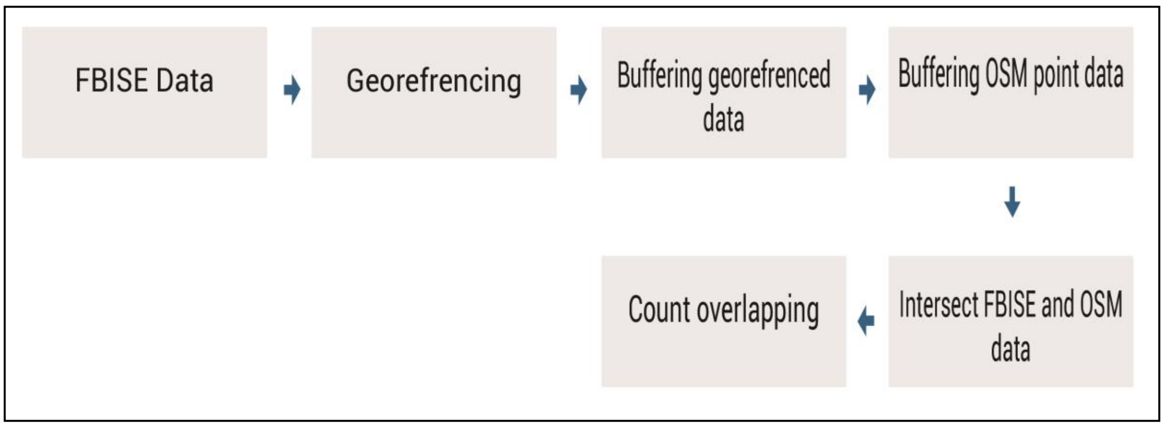

Figure 2.8. Steps of assessment of Data completeness.

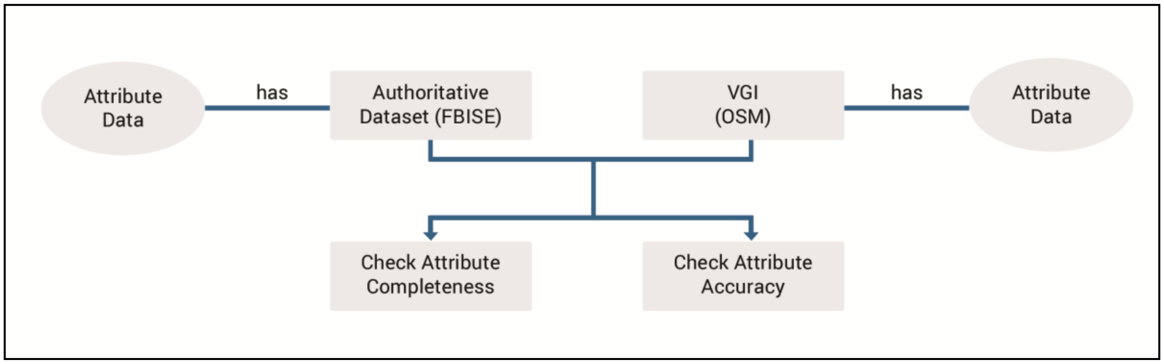

Figure 2.9. Concept of Attribute Accuracy.

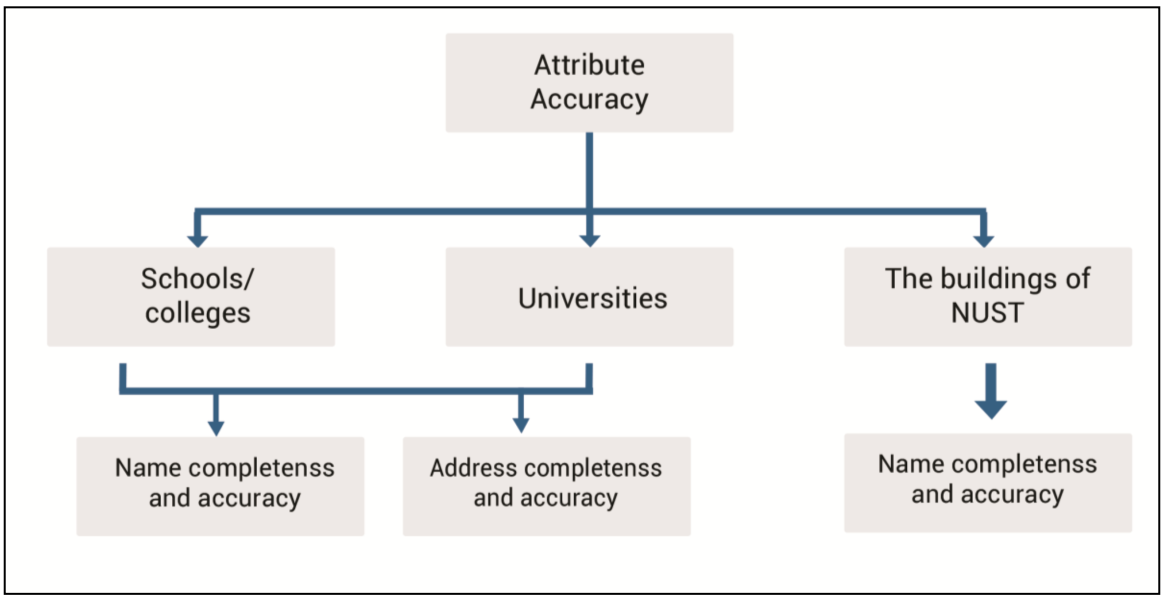

Figure 2.10. Different Aspects of Attribute accuracy.

real-world coordinate system. For educational planning, we assess positional accuracy in three dimensions as shown in Figure 2.11.

Firstly, the positional accuracy of roads is compared against authoritative data. The roads are essential components in educational planning and facilitate accessibility and routing. For this comparison we used buffer methodology as discussed in the previous section. It was observed that SoP considers the single line object for a two-way road while in OSM a two-way road is represented as two features. Goodchild & Hunter, (1997) proposed to create a buffer in order to measure the positional accuracy of linear features in two datasets. Therefore, a buffer is created against OSM. This helps determine the positional accuracy of OSM.

ArcMap was used to apply the buffer for positional accuracy of OSM and SoP roads. Firstly, the roads were separated for all *8* selected sectors of both OSM and SoP data. It was visually observed that both OSM and SoP roads for each sector represents almost the same area and length. Both the datasets were imported into ArcMap. A buffer of *15 meters* was taken against SoP dataset using the built-in buffer tool and the results were saved as a polygon shapefile. After buffering SoP dataset, the clip tool was used to select those parts of OSM roads which fall in the buffered polygon. The clipped roads of OSM were saved as a separate shapefile. The length of clipped roads were calculated and then divided by the length of original OSM roads. This provided the percentage of OSM roads which fall within the *15 meters* buffer of SoP roads. A flow chart of the procedure is shown in Figure 2.12

To check the geometrical distortion of OSM dataset, we selected 522 similar

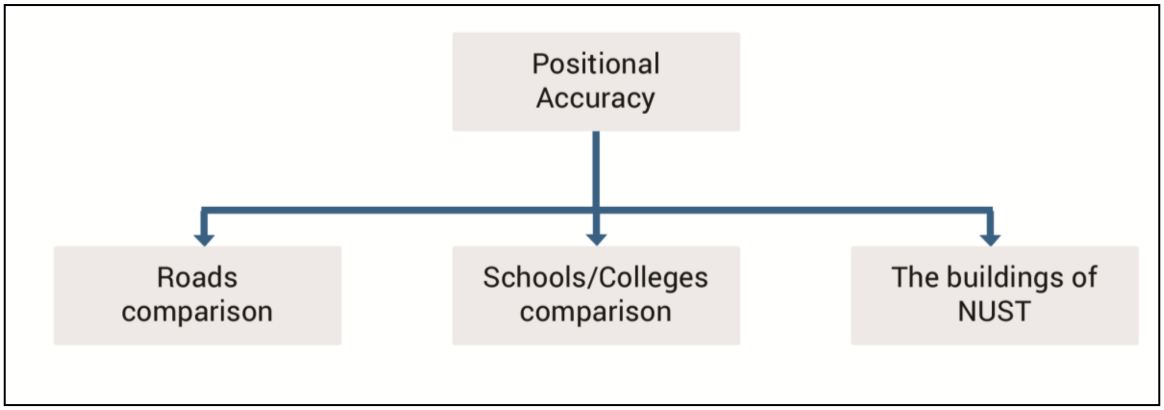

Figure 2.11. Different aspects of Positional Accuracy.

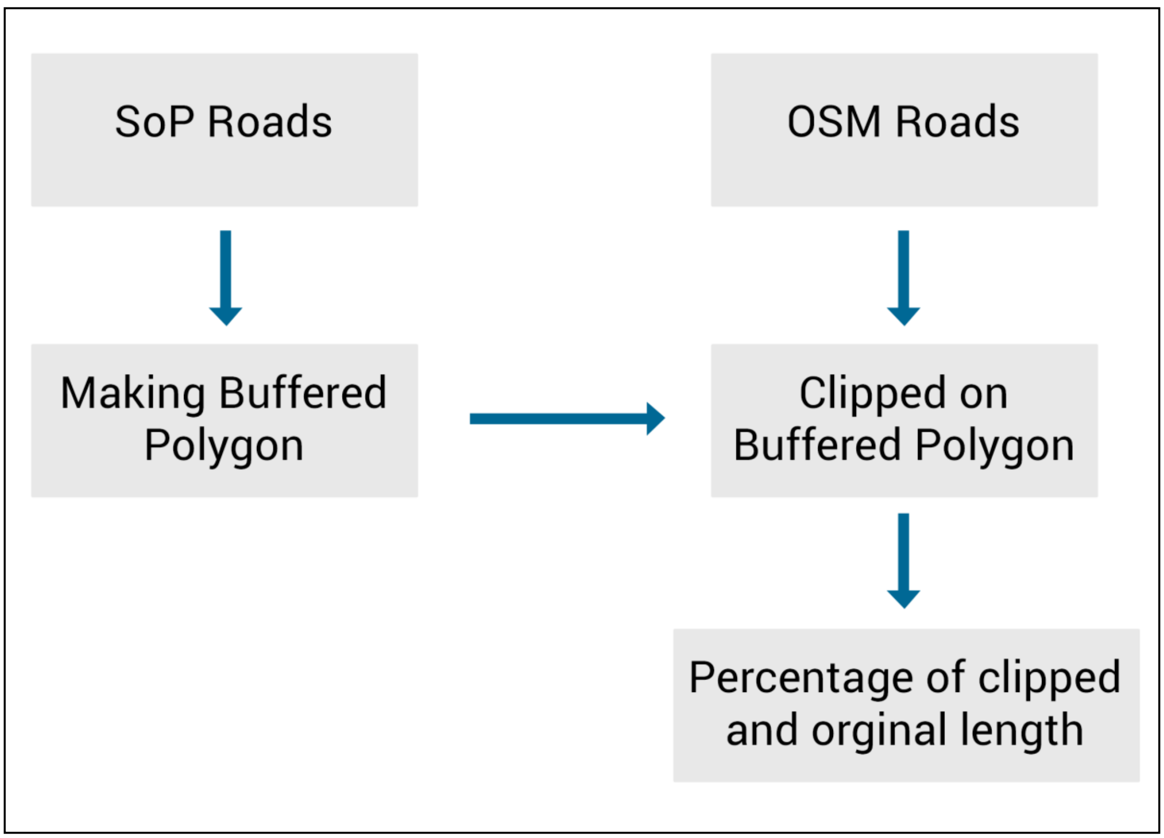

Figure 2.12. Steps for assessment of Positional accuracy of roads.

points from each SoP and OSM dataset where each point is the intersection of some road junction. We applied the bidimenstional regression method purposed by (Tobler, 1994; Friedman and Kohler, 2003).

Secondly, positional accuracy of schools and colleges of OSM data was compared with survey data. Since the SoP data was having the building outline of institution while OSM have boundary information, a GPS survey was used to assess the positional accuracy of schools and colleges across the G-9 sector of Islamabad. This sector was selected as it contains the maximum number of schools and colleges. Since OSM have both points and polygons marked as schools and colleges, the positional accuracy of points is not considered. This is due to the fact that within the boundary of an educational institute, a point is marked anywhere. Therefore it is challenging to measure the positional accuracy of a point. Figure 2.13 shows the polygons which are selected for a survey. The GPS location was marked along the boundaries of these educational institutes near the main entrance gates. Figure 2.13 also illustrates the marked surveyed points along the institutions boundaries.

As GPS points are marked near the main gates along the boundaries of institutions, a perpendicular point against the marked GPS point is selected from the OSM buildings. The distance between the GPS points and the respective boundaries of the polygons are calculated using Haversine Formula (Sinnott, 1984). The concept is shown in a flowchart in Figure 2.14.

Finally, the case of a single university is assessed as a measure of positional accuracy of individual building footprints within educational institutes. OSM building

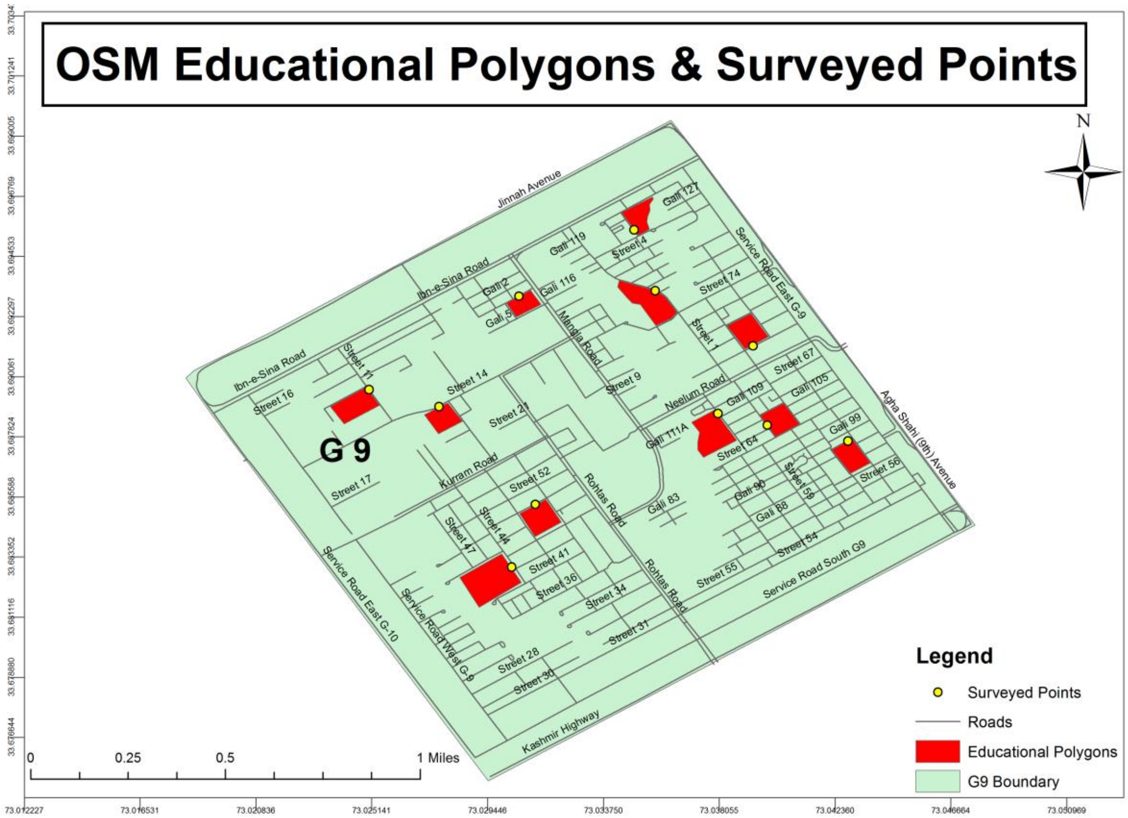

Figure 2.13. OSM Educational polygons & surveyed points.

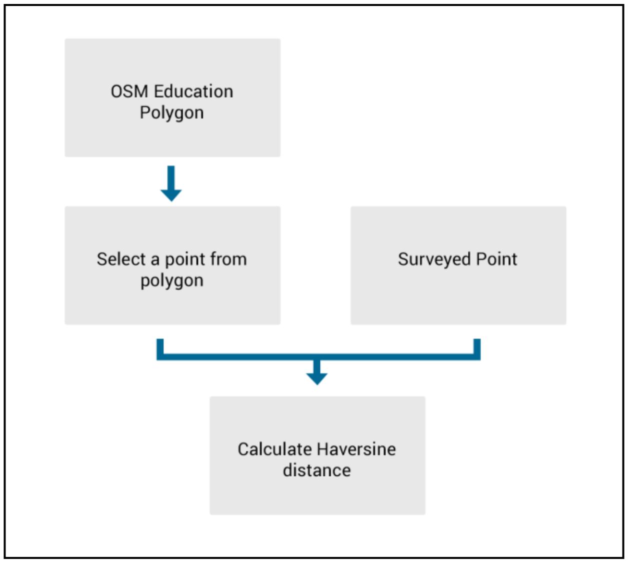

Figure 2.14. Calculating Haversine distance

polygons were extracted from downloaded data while the settlement data was obtained from SoP. The settlement data provides some information of NUST buildings. There were *13* buildings which were part of SoP data. A spatial query was performed in *PostGIS* to calculate the Hasudorff distance (Hausdorff, 1914). The flow of the procedure is shown in Figure 2.15.

2.5. Bench Marking

As for educational planning the complete picture of existing educational facilities is essential to plan for the new educational facilities. Therefore criterion for completeness is defined as the data from authoritative source and VGI source should overlap *100 %* as far as the educational facilities within the dataset are concerned. Similarly, for the positional accuracy *less than 50 m*eters error in distance between the authoritative source and VGI source is acceptable as existence of any facility is sufficient to recognize the evidence of educational facility in the vicinity. For the attribute accuracy, more than *70 %* overlap in the attributes of educational facilities from authoritative source and VGI source is acceptable for the evidence of educational facility in the vicinity.

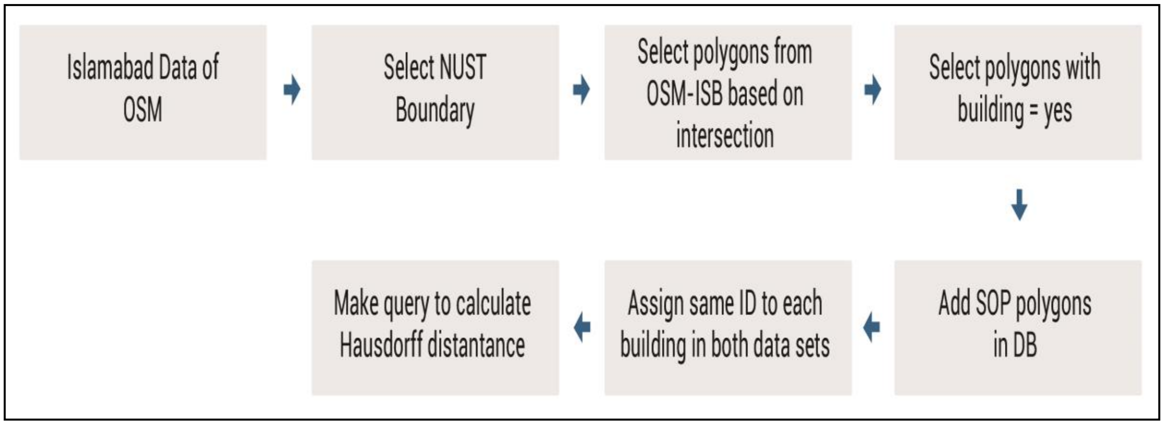

Figure 2.15. Hausdorff distance calculation.

Chapter 3

RESULTS AND DISCUSSION

3.1. Data Completeness

For this study, OSM has been used as an example of VGI dataset. In the OSM dataset for Islamabad, *181* features were tagged as educational. These include schools, colleges and universities. *64%* of these were polygons while the remaining *36%* were marked as points (see Figure 3.1). Interestingly, in some cases (approximately 20 times), the same educational institute is marked as point and polygon at the same time. This is an interesting research dimension to explore where an amenity is marked either as point, line or polygon. No educational points were marked as linear features in the given study area. Overall, in addition to data completeness, these factors become crucial when planning an educational move. For example, there are spatial analyses which are applicable on point features only. Similarly, the boundary or the area of each educational institute is measured using polygons.

For the school and college coverage, *97* polygons and *33* points in OSM were tagged as schools/colleges. On *20* occasions both point and polygon were representing the same school or college. As per the authoritative data of FBISE, there are total of *424* educational institutes. Table 3.1 shows the percentages of data completeness. *8%* of polygons were matched in OSM as compared to FBISE while *4%* of points were matched. The overall accuracy of data completeness with respect to data existence was found to be *41%*. With this low percentage of completeness in OSM, the outcome can affect decision making. Out of a total *424* schools and colleges as per the FBISE list, only

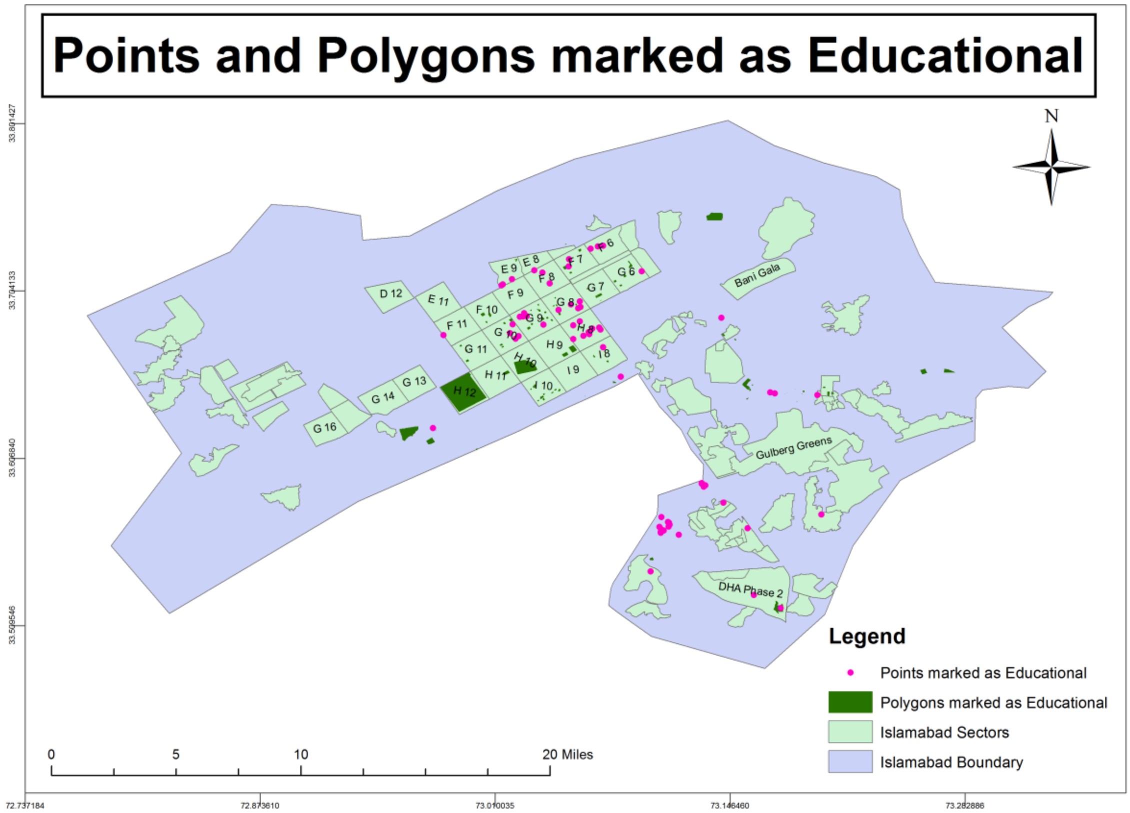

Figure 3.1. Points and polygons marked as educational institutes.

Table 3.1. Data completeness of polygon and point data tagged for schools and colleges

	Number of Features in OSM	Number of Features in OSM matching Authoritative Data	Correctness Percentage	Completeness Percentage
Polygons	97	37	38%	8%
Points	33	17	51%	4%
Total	130	54	41%	13%

130 were present in OSM. Furthermore, only *54* matched to FBISE. This is equal to *13%* only which shows a poor coverage of schools and colleges in OSM for the study region.

In the case of universities in Islamabad, data completeness was also evaluated. *29* polygon features of OSM were tagged as Universities while only *2* Universities were tagged as points. As per the authoritative data, *21* universities were affiliated with HEC. No occurrence was found where a University was marked as both point and polygon. There were other observations such as almost 53% polygons and 50% points were representing a true picture of Universities while the rest were the departments of respective Universities. The overall percentage of correctly marked Universities on OSM is 61 % while 90% of HEC affiliated universities is present in OSM as shown in Table 3.2.

A case for data completeness of buildings within NUST, one of the largest Universities of the country is presented. In order to extract the buildings from OSM, a spatial query was created in *ArcGIS,* which performs the spatial intersection on all polygons within the NUST boundary. In the OSM schema, the polygons have an attribute called *building* which accepts "yes" or "no" as a Boolean value. Figure 3.2 shows the extracted buildings of NUST. As per the authoritative data, NUST contains a total of *40* buildings. In OSM, there was *100%* coverage which is very good. However these include buildings which are geographically present but their building attribute is marked as "no" or they were empty. This is due to a lack of awareness and knowledge amongst volunteers while marking these spatial features.

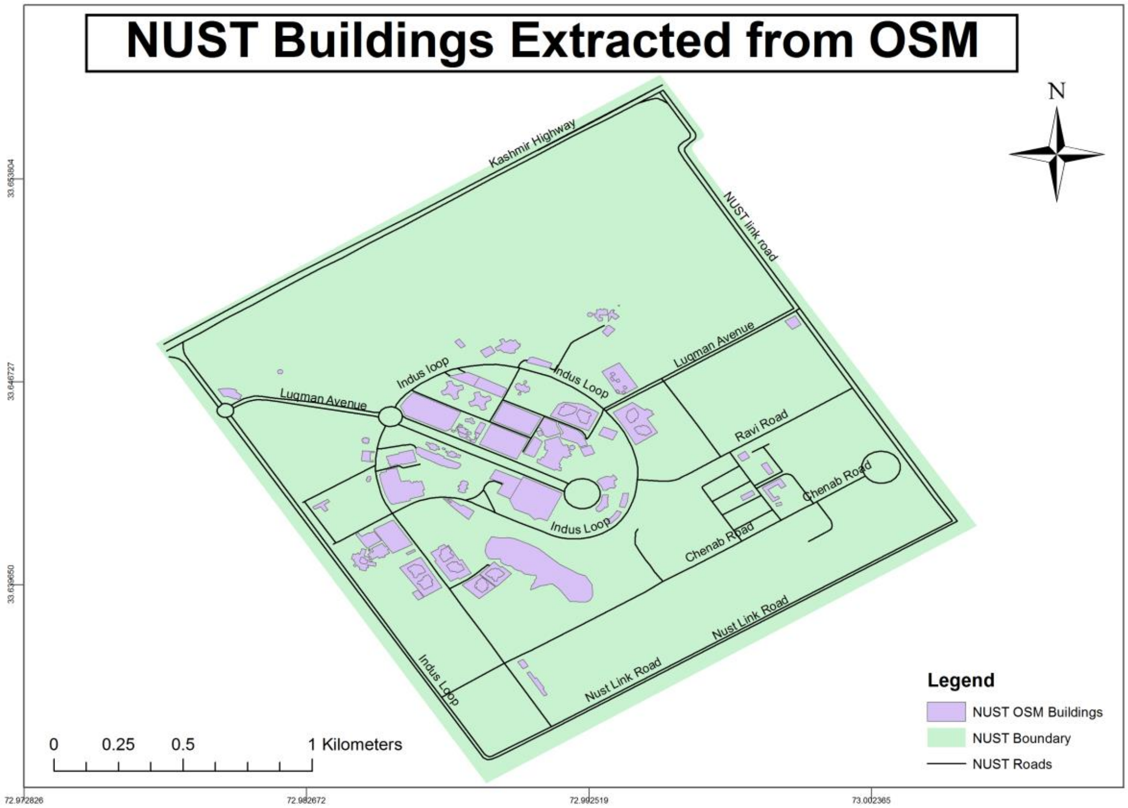

Figure 3.2. OSM polygons in NUST.

Table 3.2. Data completeness of polygon and point data tagged for universities

	Number of Features in OSM	Number of Features in OSM matching Authoritative Data	Correctness Percentage	Completeness Percentage
Polygons	29	18	58 %	86%
Points	2	1	3 %	4%
Total	31	19	61 %	90%

3.2. Attribute Accuracy

As described in the previous section, there were only *54* educational institutes present in OSM matching the authoritative data. To assess attribute accuracy, the total number of names and addresses of the records present were measured. Moreover, it has been checked if the correct attribute is reflected in OSM.

When compared using Lavenshtein distance, it was found that for schools and colleges 88.8% of the names were present and 64 % were correctly matched with FBISE data while 37.0% of the addresses of schools and colleges were present and 14% of them were matched with FBISE data. The results have been shown in Figure 3.2. On the other hand, for universities, the name was present for all marked feature i.e., 100 % while 88% matched with HEC data while 78% of them have address present and 52% were correct. The results have been shown in Figure 3.3. Lastly, NUST a test case for university buildings, the name of the all the buildings were present and all of them were correct. The overall details of Attribute Accuracy are shown in Table 3.3. Achieving the attribute accuracy can be a challenging task. It depends on volunteers and what attribute they are specifying while creating a geographic feature. In the case of attribute accuracy for educational institutes, some names were missing while for the addresses, street names were often omitted and so a complete match could not be obtained. Attribute accuracy is required for efficient educational planning. We consider attribute accuracy as inaccurate when the Lavenshtein distance is greater than 7.

3.3. Positional Accuracy

A comparison of OSM and SoP roads of *8* sectors of Islamabad was performed. The authoritative data is considered as highly accurate prepared by professionals. These datasets normally go through rigorous quality measures and checks.

The percentages of buffered-overlaps in *8* sectors of Islamabad are shown in Table 3.4. It can be seen that the lowest percentage is *84.2%* while the highest is *96.1%*.The average is *90.2%*. It can be concluded that OSM road network is up to *90%* accurate. This is reasonably a high percentage and corresponds to other jurisdictions. Therefore the OSM road network is very commonly used in routing studies even in developing countries.

We also used Bidimensional regression to compare the degree of resemblance between OSM and SoP roads dataset. Estimated parameters are shown in table 3.5

These parameters indicate how the OSM pattern must be transformed to get the SoP pattern. Negative value of a_1 indicate horizontal east-to-west shift while the positive value of a_2 indicate a vertical south-to-north shift.

Furthermore the contraction or expansion parameter (Φ) is slightly greater than 1 hence represent a negligible expansion. While negative value of θ refers to a clockwise rotation of the OSM to get the SoP pattern.

The close to one value of r^2 indicate a strong degree of resemblance and the small value of DI (Distortion index) is indicate that a very small distortion would happen on the way of regressing the OSM to SoP dataset.

Positional accuracy defines the reality of how accurate the data is marked in VGI with respect to their actual position in real world. Therefore, to strengthen this argument,

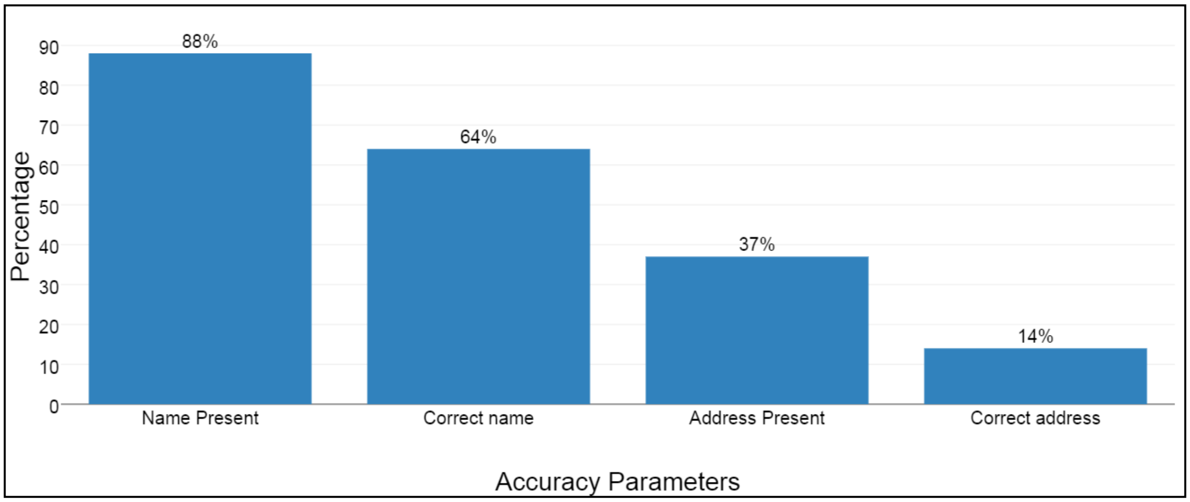

Figure 3.3. Results of Attribute accuracy for Schools/Colleges.

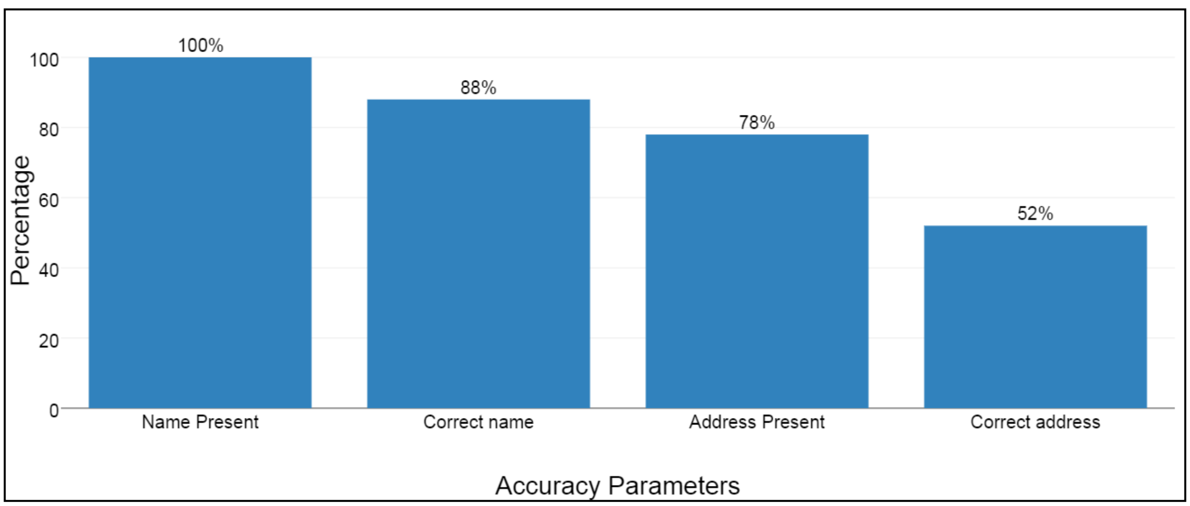

Figure 3.4. Results of Attribute accuracy for Universities.

Table 3.3. Correct names and addresses percentage

	Name Present	**Correct name**	**Address Present**	**Correct address**
Schools/Colleges	88%	64%	37%	14%
Universities	100%	88%	78%	52%
NUST buildings	100%	100%	-	-

Table 3.4. Percentage of overlap of OSM and SoP datasets for various sectors of Islamabad

Sector	**Percentage of overlap**
G9	90.2%
F8	87.5%
H8	95.7%
G8	85.3%
F8	89.5%
H8	96.1%
G7	84.2%
F7	93.6%

Table 3.5. Results of Bidimensional Regression.

	a_1	**a_2**	**b_1**	**b_2**	**Φ**	**Θ**	**DI**	**r^2**
OSM	-0.122	0.0324	1.0012	-0.000999	1.0012	-0.057	0.0038	0.9

the positional accuracy of OSM is measured by taking some samples while checking their positions (coordinates) using a GPS survey.

The result of survey's distance from the institution is shown in Table 3.6 along with the GPS accuracy. It can be seen that the minimum value for inaccuracy is 1.3 meters while maximum value is 10.6 meters with an average of 5.23 meters. Keeping in mind the GPS average inaccuracy of 4.43 meters, it can be said that the positional accuracy of schools and colleges lie between 0.8 meters to 9.66 meters which is acceptable.

The positional accuracy is further evaluated inside the NUST campus. The results of the PostgreSQL query are shown in Table 3.7. It can be seen that OSM buildings are on average *4.2 meters* within SoP buildings while there is a variation from *3.1 meters* to *5.6 meters* overall.

The positional accuracy is valuable in educational planning. Questions such as where are new schools, colleges and buildings within a large University needed? What and where are the optimal routes for students to commute to educational institutes? If digitization is carried out with care and the high resolution imageries are available, an accurate representation of data can be achieved.

The results for positional accuracy described here reinforce previous studies which have focused on analyzing the accuracy of the OSM road network. A small offset between the location of educational institutes in OSM and ground truth was detected, however the anomaly would not affect educational planning. When attribute accuracy and completeness were assessed, the results indicate a severe mismatch between authoritative datasets and OSM. The differences make OSM insufficient for educational planning in

Table 3.6. Distance between GPS point and OSM boundary with GPS accuracy at each point

Sample #	**Distance (meter)**	**GPS Inaccuracy (meter)**
1	5.9	3.75
2	5.2	4.1
3	10.6	4
4	1.6	4.5
5	6.8	7
6	2.9	3.5
7	4.2	2.89
8	9.7	4.3
9	2.1	3.7
10	7.3	8.32
11	1.3	2.75

Table 3.7. Buildings name of NUST and the Hausdorff distance between OSM and SoP buildings

NUST Building	**Hausdorff Distance (meters)**
Admin Office	5.2
Habib Bank Limited	3.5
Rumi Hostels	4.9
Masjid-e-Rehmat	5.6
IGIS	3.9
Concordia 2(C2)	3.7
SCME	4.6
Concordia 1(C1)	3.5
NUST Villas	5.3
ISRA Apartments	3.3
IQRA Apartments	3.7
Masjid-e-Taqwa	3.1
RIMMS	4.8

Islamabad and other spatial and or non-spatial datasets or VGIs like Wikimapia need to be included to provide reliable results for educational planning.

Chapter 4

CONCLUSION AND RECOMMENDATIONS

4.1 Conclusion

When official spatial data from national agencies is expensive to access, slow to change and presented in difficult to use formats, VGI is seen as a viable alternative for spatial data. The quality of the OSM for a planning task in the study region was tested in several dimensions. Completeness, positional accuracy and attribute accuracy were chosen as key requirements for educational planning. For completeness, there are large gaps in the VGI for educational institutes. OSM data for Educational Planning is below the bench mark of 100% data completeness. For example, only 13% of schools and colleges are present in the OSM data. Similarly, for attribute accuracy, OSM data for educational planning is below the bench mark of 70% overlap for schools and colleges but satisfies for Universities. Finally, the positional accuracy of institutes and individual buildings within them was assessed. The offset was found to be insignificant and a high degree of overlap with official data was discovered. Positional accuracy of OSM data for educational planning satisfies the bench mark of at max 50 meter inaccuracy.

While the positional accuracy of the institutes found in OSM is satisfactory, the measures of completeness and attribute data are not sufficient for an educational planning task. In order to determine the location for a new educational institute, complete knowledge of existing educational institutes is required along with reliable attribute information describing the name and type of each educational institute. The results support previous studies which highlight the positional and topographical accuracy of OSM is generally acceptable, however this thesis has highlighted the problems with using OSM for specific tasks, in this case educational planning. It is anticipated that crowdsourcing will be effective at capturing changes to local areas faster than official mapping sources which are not updated frequently and require rigorous testing before publishing. However, without completeness, this metric is irrelevant for education planning.

4.2 Recommendations for further research

The study presented here did not include an assessment of the temporal accuracy of the OSM data. This can be assessed in future studies. Also, OSM was the only VGI source used to test the viability of VGI for educational planning while a single region of Pakistan was the only area analyzed. Further VGI projects, such as *Wikimapia (http://wikimapia.org)* will be explored to determine if additional information can be collected to support planning in regions where OSM data is missing or erroneous. The research will also be extended to other areas to compare the quality of VGI for educational planning in other cities. There is a need to set bench marks and analyze more deeply OSM data to make it a part of NSDI.

REFERENCES

1. Ali, A, & Ahmad, M. (2013). Geospatial Data Sharing in Pakistan: Possibilities and Problems? In Proceedings of Global Geospatial Conference 2013, Addis Ababa, Ethiopia.
2. Ballatore, A., & Zipf, A. (2015). A Conceptual Quality Framework for Volunteered Geographic Information. In International Workshop on Spatial Information Theory (pp. 89-107). Springer International Publishing.
3. Barron, C., Neis, P., & Zipf, A. (2014). A Comprehensive Framework for Intrinsic OpenStreetMap Quality Analysis. Transactions in GIS, 18(6), 877-895.
4. Bartoschek, T., & Keßler, C. (2013). VGI in Education: From K-12 to Graduate Studies. In Crowdsourcing Geographic Knowledge (pp. 341-360). Springer Netherlands.
5. Ciepłuch, B., Mooney, P. & Winstanley, A.C. (2011). Building Generic Quality Indicators for OpenStreetMap. In 19th Annual GIS Research, UK.
6. Camboim, S., Bravo, J. & Sluter, C. (2015). An Investigation into the Completeness of, and the Updates to, OpenStreetMap Data in a Heterogeneous Area in Brazil. ISPRS International Journal of Geo-Information, 4(3), 1366-1388.
7. Campagna, M., Floris, R., Massa, P., Girsheva, A. & Ivanov, K. (2015). The Role of Social Media Geographic Information (SMGI) in Spatial Planning. In Planning Support Systems and Smart Cities (pp. 41-60). Springer International Publishing.
8. Connors, J. P., Lei, S. & Kelly, M. (2012). Citizen Science in the Age of Neogeography: Utilizing Volunteered Geographic Information for Environmental Monitoring, Annals of the Association of American Geographers, 102(X): 1-23.
9. De Leeuw, J., Said, M., Ortegah, L., Nagda, S., Georgiadou, Y. & DeBlois, M. (2011). An Assessment of the Accuracy of Volunteered Road Map Production in Western Kenya. Remote Sensing, 3(12), 247-256.
10. Degrossi, L. C., de Albuquerque, J. P., Fan, H., & Zipf, A. (2016). A Conceptual Model for Quality Assessment of VGI for the Purpose of Flood Management. In Proceedings of 19th AGILE. Helsinki.
11. Elwood, S., Goodchild, M. & Sui, D. (2012). Researching Volunteered Geographic Information: Spatial Data, Geographic Research, and New Social Practice. Annals of the Association of American Geographers, 571-590.
12. D’Antonio, F., Fogliaroni, P., & Kauppinen, T. (2014). VGI Edit History Reveals Data Trustworthiness and User Reputation. Proceedings of AGILE. Spatin.
13. Girres, J. F. & Touya, G. (2010). Quality Assessment of the French OpenStreetMap Dataset. Transactions in GIS, 14(4), 435-459.
14. Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. GeoJournal, 69(4), 211–221.

15. Goodchild, M. F. & Li, L. (2012). Assuring the Quality of Volunteered Geographic Information. Spatial Statistics, 1, 110-120.

16. Goodchild, M. F., & Hunter, G. J. (1997). A Simple Positional Accuracy Measure for Linear Features. International Journal of Geographical Information Science, 11(3), 299-306.

17. Goranson, C., Thihalolipavan, S. & di Tada, N. (2013). VGI and Public Health: Possibilities and Pitfalls. In Crowdsourcing Geographic Knowledge (pp. 329-340). Springer Netherlands.

18. Haklay, M. (2010). How Good is Volunteered Geographical Information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets. Environment and Planning B: Planning and Design, 37(4), 682-703.

19. Hall, P. (1996). Cities of Tomorrow: An Intellectual History of Urban Planning and Design in the Twentieth Century. Oxford: Blackwell.

20. Hausdorff, F (1914). Grundzüge der Mengenlehre, Leipzig: Veit, ISBN 978-0-8284-0061-9 Reprinted by Chelsea in 1949.

21. Horita, F. E. & De Albuquerque, J. P. (2013). An Approach to Support Decision-Making in Disaster Management based on Volunteer Geographic Information (VGI) and Spatial Decision Support Systems (SDSS). In Proceedings of the 10th International Conference on Information Systems for Crisis Response and Management, Baden-Baden, Germany (Vol. 1215).

22. Hung, K.-C., Kalantari, M. & Rajabifard, A. (2016). Methods for Assessing the Credibility of Volunteered Geographic Information in Flood Response: A Case Study in Brisbane, Australia. Applied Geography, 68, 37-47.

23. Ludwig, I., Voss, A., & Krause-Traudes, M. (2011). A Comparison of the Street Networks of Navteq and OSM in Germany. In Advancing Geoinformation Science for a Changing World (pp. 65-84). Springer Berlin Heidelberg.

24. Keßler, C., & de Groot, R. T. A. (2013). Trust as a proxy measure for the quality of volunteered geographic information in the case of OpenStreetMap. In Geographic Information Science at the Heart of Europe (pp. 21-37). Springer International Publishing.

25. Kresse, W. & Fadaie, K. (2004). ISO Standards for Geographic Information, Springer.

26. Masli, M., Bouma, L., Owen, A., & Terveen, L. (2013). Geowiki+ route analysis= improved transportation planning. In Proceedings of the 2013 Conference on Computer Supported Cooperative Work Companion (pp. 213-218). ACM.

27. Memon, G. R. (2007). Education in Pakistan: The key issues, problems and the new challenges. *Journal of Management and Social Sciences*, *3*(1), 47-55.

28. Mooney, P., Corcoran, P., & Winstanley, A. C. (2010, November). Towards quality metrics for OpenStreetMap. In Proceedings of the 18th SIGSPATIAL

International Conference on Advances in Geographic Information Systems (pp. 514-517). ACM.

29. Okpala, D. C. (2001). Spatial Information: the Basic Tool for Sustainable Human Settlements Development Planning and Management. International Conference and Spatial Information for Sustainable Development. Nairobi.
30. Oort, P.A.J. van (2006) Spatial Data Quality: From Description to Application, PhD Thesis, Wageningen University, p. 125.
31. Poser, K., & Dransch, D. (2010). Volunteered Geographic Information for Disaster Management with Application to Rapid Flood Damage Estimation. Geomatica, 64(1), 89-98.
32. See, L., Mooney, P., Foody, G., Bastin, L., Comber, A., Estima, J., & Liu, H. Y. (2016). Crowdsourcing, Citizen Science or Volunteered Geographic Information? The Current State of Crowdsourced Geographic Information. ISPRS International Journal of Geo-Information, 5(5), 55.
33. Sinnott, R.W., 1984, Virtues of the haversine: Sky and Telescope, v. 68, no. 2, p. 159.