VOLUNTEERED GEOGRAPHIC INFORMATION: AN ASSESSMENT OF DATA QUALITY IN MALAWI

MASTER OF SCIENCE (INFORMATICS) THESIS

GIOVANNIE MAKONDI

UNIVERSITY OF MALAWI

JULY 2024

VOLUNTEERED GEOGRAPHIC INFORMATION: AN ASSESSMENT OF DATA QUALITY IN MALAWI

MASTER OF SCIENCE IN INFORMATICS THESIS

By

GIOVANNIE MAKONDI

BSc (Management Information Systems) -University of Malawi, The Polytechnic

Submitted to the Department of Computing, School of Natural and Applied Sciences, in partial fulfilment of the requirements for the degree of Master of Science in Informatics

UNIVERSITY OF MALAWI

JULY 2024

DECLARATION

I, the undersigned, declare that this thesis is my original work and has not been submitted to any other institution for similar purposes. Acknowledgements have been made where other people's work has been used. I bear the responsibility for the contents of this paper.

Giovannie Makondi

Full Legal Name

Signature

26 July 2024

Date

CERTIFICATE OF APPROVAL

The undersigned certifies that this thesis represents the student's work and has been submitted with his approval.

Signature: ___________________________ Date: ______________________

Kondwani Godwin Munthali, PhD (Senior Lecturer)

Supervisor

DEDICATION

To God who gifts us life, identity, strength, endurance, and healing.

To a selfless helpmate – Maness.

To family and friends who ask how high when told to jump.

ACKNOWLEDGEMENTS

I am grateful to everyone who contributed to the success of this thesis. Your support has been the fuel to the realisation of this dream. Without your support, this dream would be impossible.

A special thank you to my supervisor Dr Kondwani Godwin Munthali for the untiring and professional mentorship and support throughout the study period. Your commitment to ensuring the success of this thesis will forever be appreciated.

I also thank all my lecturers and UNIMA Informatics 10 cohort classmates. You have all immensely contributed to the knowledge that has been the foundation of this study. Learning and keeping it together throughout the COVID-19 pandemic would not have been easier without your presence and support.

I want to thank my mentors and colleagues at Malawi University of Business and Applied Sciences, Mr Goodal Nyirenda, Dr Patrick A. Chikumba and Mr Dickson Chinguwo for all their support and encouragement.

Thank you!

ABSTRACT

Volunteered Geographic Information (VGI) is touted as the modern-day approach to spatial data acquisition. Around the globe, VGI is slowly being integrated with authoritative spatial data while completely replacing it in other cases. However, the issue of quality in VGI remains combative, mainly from a lack of specialised skills of its contributors. VGI has great potential in revolutionising spatial data sourcing, hence the need to establish adequate confidence in the status of VGI quality to boost its uptake, particularly in Malawi. This interpretive philosophical study that followed a mixed-methodology aimed at ascertaining the status of VGI quality in Malawi through analysis of VGI expert user experiences and systematic quality testing of VGI. Interviews, questionnaires, and online spatial data repositories as secondary data sources were used for data collection. Narrative analysis extracted themes from narratives of expert user experiences with VGI quality. Analytical Hierarchical Processing (AHP) was used to rank six key data quality dimensions in the order of priority to determine a quality dimension on which VGI quality was tested. As a result, the OpenStreetMap (OSM) Education Facilities dataset for Malawi was tested for horizontal accuracy against the Malawi Ministry of Education Primary Schools dataset. The test used the National Standard for Spatial Data Accuracy (NSSDA) methodical approach to estimate positional accuracy and the results were validated using Euclidean distance buffers as an application of Horizontal Positional Error (HPE). The study observed that VGI suffers from presumed notions of lack of quality primarily attributed to misinformation about VGI quality, however VGI used by the expert users in their practice surpassed their quality expectations. It was also discovered that VGI expert users prioritised accuracy among key data quality dimensions. From the quality test, the OSM Education Facilities dataset for Malawi was found to have a Circular Standard Error (CSE) of 19.20 metres and an NSSDA of 47.0150 metres for featured Primary schools. This study found VGI to be within tolerable levels of accuracy when compared against authoritative spatial data.

Keywords: Data, Spatial data, Quality, VGI, Users, Malawi, OSM, AHP, NSSDA

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS AND ACRONYMS

AHP	Analytical Hierarchical Processing
CDNGI	Chief Directorate: National Geospatial Information
CSE	Circular Standard Error
DoDMA	Department of Disaster Management Affairs
GIS	Geographical Information System
GPS	Global Positioning System
HPE	Horizontal Positional Error
ISO	International Organisation for Standardisation
MASDAP	Malawi Spatial Data Platform
MCDA	Multi-Criteria Decision Analysis
NSDI	National Spatial Data Infrastructures
NSPDR	National Spatial Planning Data Repository
NSSDA	National Standard for Spatial Data Accuracy
OpenDRI	Open Data for Resilience Initiative
OSM	OpenStreetMap
RCMRD	Regional Centre for Mapping of Resources for Development
RMSE	Root Mean Squared Error
SABAP	Southern African Bird Atlas Project
SADC	Southern African Development Community
SDI	Spatial Data Infrastructure
UNDP	United Nations Development Program
VGI	Volunteered Geographic Information

CHAPTER 1

INTRODUCTION

In the 21st century, governments, businesses, and individuals harness the power of technology to produce and accumulate huge amounts of data. Data is usually at both ends of business processes, natural occurrences, social interactions, and research projects. Every email sent, every click on a camera and each call on mobile phones comes with a trail of data in what could be described as the greatest data and information revolution of all time. This data explosion has led to the emergence of Big Data analytics – a field that deals with data whose volume and exponential growth cannot be captured, managed, and processed by traditional data management tools such as relational databases (Meng & Ci, 2013). In 2018, the global business magazine Forbes reported that an estimated 2.5 quintillion bytes of data were being generated every day. With the world holding an estimated 33 Zettabytes of data that year (Forbes, 2018), further projections suggest that the world's data would grow to 175 Zettabytes by 2025 (Coughlin, 2018). These projections seem to hold as it is believed that the world was host to 79 Zettabytes of data at the end of 2021, representing over 120% growth in data volume from the estimations of the year 2018 (Djuraskovic, 2022). However, up to 90% of the new data contributing to the data growth is thought to be duplicate data (Desjardins, 2019; Djuraskovic, 2022; Marr, 2020)

Like many other social and scientific fields, the field of Geographic Information Systems (GIS) is not an exception to the data accumulation spasm. Various studies show that spatial data associated with different subjects and objects is continuously collected by various public and private institutions and is often considered proprietary to those institutions (Johnson et al., 2017). Organisations such as GIS Tech Consultants, Catalyst Spatial Consulting, Tom-Tom, Radar, C12 Consultants and 28East collect and manage spatial data commercially. Most commercial spatial data is generated by well-trained individuals operating under well-financed institutions. However, the cost of generating and managing such data keeps increasing, leading to a quest for more financially viable ways of collecting spatial data (Stage, 2009). In addition, spatial data acquisition requires planning, training, and time which is not feasible in times of crisis (Stieglitz et al., 2018). During environmental or socio-political crises, GIS experts have limited time to acquire spatial data to aid in formulating various interventions (World Bank, 2014).

With cost and time as some of the limiting factors on spatial data acquisition in recent years, Volunteered Geographical Information (VGI) has emerged. VGI is spatial data collected and shared by volunteers in their communities using internet-based mapping systems (Cooper et al., 2012) at the expense of government mapping agencies, research non-governmental organisations or private entities involved in spatial data collection at a cost. VGI is often called crowdsourced spatial data, citizen participatory mapping, and sometimes open spatial data (See et al., 2016). Despite the minor differences between VGI, citizen participatory mapping, open spatial data and crowdsourced spatial data, they are all underpinned by involving citizens in collecting and sharing spatial data (See et al., 2016). VGI can also be considered as spatial data contributed by volunteers at no fee, where contributors have no special training skills (Goodchild & Li, 2012). Various authors argue that VGI primarily addresses the problem of time restrictions and increasing costs of acquiring the data, among other challenges in spatial data acquisition (Goodchild & Li, 2012).

With the citizenry having increased access to mobile devices and mapping technologies even in the remotest areas across the world, communities have the ability to generate and share spatial data. Consequently, a vast amount of spatial data is becoming openly available to access, download, analyse and share through various platforms. To increase the availability and accessibility of spatial data, various governments and organisations across the globe have taken the initiative to make spatial data available to stakeholders through Spatial Data Infrastructures (SDIs).

In Africa, where the majority of the countries are classified as developing countries (Essoungou, 2011), up to 80% of national strategic planning and developmental decision-making processes in government are based on spatial data (Muya, 2017).

Location intelligence, which is the insight gained from visualising and analysing geospatial data (ESRI, 2023)**,** has become very important in dealing with almost every economic, social, and political development aspect (Nkwae & Nichols, 2006). This dependency on spatial data not only necessitates the need for investments in National Spatial Data Infrastructures (NSDI) by African governments but also calls for the need to expand the horizon for spatial data acquisition beyond the traditional ways (Mwungu, 2017; Nkwae & Nichols, 2006). In countries where NSDIs are successfully implemented, there is efficiency in the collection and use of spatial data and saving of financial resources which are usually limited in developing countries (Chikumba, 2019). Over the last decade, several NSDIs have been enacted in the African region, and some have gone as far as having active geo-data portals. Some of the currently active geo-data portals in Africa include the Africa GeoPortal, the Regional Centre for Mapping of Resources for Development (RCMRD), the National Spatial Planning Data Repository (NSPDR) and Chief Directorate: National Geospatial Information (CDNGI) Spatial portals in South Africa and the Malawi Spatial Data Platform (MASDAP) in Malawi (Gardner & Mooney, 2018; Haklay et al., 2014; Mwungu, 2017; Muya, 2017). These spatial data portals provide a platform for the practice and hosting of VGI through integration (Genovese & Roche, 2010).

Several VGI projects have also been realised across Africa within the last decade. Some of the notable projects include the Map Kibela project in Kenya, the Open Data for Resilience Initiative (OpenDRI), OpenStreetMap (OSM) mapping for refugees in Malawi and the Southern African Bird Atlas Project 2 (SABAP2) covering countries in Southern Africa. The integration of VGI in various NSDIs and their respective spatial data portals across the world offers government ministries, departments, and agencies expanded and limitless opportunities in the utilisation of less costly spatial data for their development projects (Genovese & Roche, 2010; Haklay et al., 2014).

Whilst data unavailability is fast becoming an uncommon problem, data whose volume grows exponentially with time comes with many challenges. The challenges include inaccuracy, data inconsistency, disorganisation of data stores, lack of metadata, poor quality controls at data entry and reliability of this data (Roumeliotis, 2020). When combined, the challenges of accuracy and reliability coupled with the inability to legitimise the data sources result in compromises in the data quality. Regrettably,

suspicious data has less value in the age of the data-driven world, and in the worst cases, poor-quality data leads to massive business losses due to ill-informed business decisions (Wassén, 2019). Most data quality concerns are attributed to the vast amount of data sources at the disposal of businesses and people today and these concerns affect the usability of most of the data in today's data-driven world (Roumeliotis, 2020).
The definition of data quality hinges on the conditions of its application (Veregin, 1999), however, various authors define data quality as a group of traits starting from accuracy, completeness, reliability, relevance, and timeliness (Fan, 2015; Harris, 2011; Sarfin, 2021). Data quality is also largely influenced by who, how and where the data is generated, handled, stored, and used (Vosoughi et al., 2018). The existence of continuous changes in data attributes entails sustained efforts in maintaining quality. With too much digital data available, ensuring that the spatial data is accurate is a big challenge yet most data users do not put much thought into the quality of the data they are using (Fan, 2015; Lee et al., 2006)**.** In many projects, accuracy, completeness, and other quality issues of GIS data downloaded from various opensource, commercial or government sources are often overlooked (Santini, 2019). Users are more inclined towards getting the result and overlooking the quality of that result most of the time.

While VGI continues to gain momentum globally, it also comes with growing concerns about its quality and the subsequent effects the quality has on the success of various GIS and GIS-supported projects. Of particular concern is the accuracy surrounding spatial data in digital form, the main form in which VGI is shared. While many physical maps include a map reliability or confidence rating which aids the users in deciding the suitability for the use of the map, this information is rarely included in spatial data that is published in digital form (Greenfeld, 2013). Concerns about data quality and incomplete representation of data are considered as the two main barriers to embracing VGI for decision-making (Ferster et al., 2018). These concerns could be the reason why two decades after the concept of VGI was introduced in Southern Africa, VGI uptake remains low. Governments, aid agencies and various development partners still struggle to source spatial data for development planning and time-sensitive interventions such as disaster mitigation, response and recovery, yet VGI boasts of the enormous potential to be both a reliable and sustainable source of spatial data and information for various applications (Yilma, 2016).

1.1 Problem statement

Various studies have tackled matters of spatial data and VGI quality in the dimensions of theme, space and time, particularly exploring issues of accuracy, completeness, precision and consistency as components of spatial data quality (Greenfeld, 2013; Hunter et al., 2003; Pascual, 2011; Veregin, 1999). While strides have been made in discussing spatial data collaboration and spatial data management for various fields in Malawi and the African region (Chikumba & Chisakasa, 2018; Cooper et al., 2012; Muya, 2017; Sekhula, 2013) there have not been significant conversations regarding the assessment and status of VGI quality, hence the status of VGI quality in Malawi remains relatively unknown. This study seeks to bridge this gap by attempting to ascertain the status of VGI quality in Malawi.

The goal for this study was thus to ascertain the current state of VGI quality in Malawi by learning the pre-usage and post-usage VGI quality perceptions of its expert users through their narratives, discovering the key data quality dimension of VGI as considered by the expert users and evaluating the quality of VGI through comparison with authoritative data.

Allowing the citizenry to contribute spatial data voluntarily is a very innovative mechanism for producing and distributing spatial data. With advancements in technologies for mobile devices, Web 2.0, and Global Positioning System (GPS), VGI can potentially redefine and simplify spatial data acquisition. However, VGI continues to suffer from a presumed notion of inaccuracy, leading to the constant questioning of the quality of VGI by various players in the field of GIS. While VGI is a favourable solution to the problem of high costs and time limitations associated with data acquisition in various location intelligence-driven projects such as natural disaster emergency responses and health and socio-economic interventions among others, the status of the quality of VGI and its performance against authoritative spatial data remains insufficiently studied in Malawi. This position significantly hampers the adoption of VGI by various GIS stakeholders in the country.

1.2 Research objective

The study's main objective was to determine the status of Volunteered Geographic Information quality in Malawi. The specific objectives are:

- To explore VGI quality perceptions and observations by its expert users before and after use.
- To identify key data quality dimensions of VGI as considered by its expert users.
- To examine the quality performance of VGI against authoritative spatial data.

1.3 Research questions

In striving to bridge the knowledge gap on the state of the quality of VGI in Malawi, the study attempted to answer the following research questions:

a) What are the expert users' perceptions and observations on VGI quality?

b) What data quality dimensions are considered most important by the expert users of VGI?

c) What is the quality performance of VGI against authoritative spatial data?

1.4 Thesis structure

This thesis is divided into five chapters. **Chapter one** introduced the study topic by providing a background to the study. Also included in the chapter is the problem statement, research questions and related study objective. **Chapter two** discusses various literature on VGI, data quality, data quality dimensions, spatial data sources and spatial data quality testing. The study also draws a conceptual framework from Juran’s theory of quality. **Chapter three,** comprising the methodology, explains how the study was carried out, focusing on the study philosophy, approach, and strategy adopted. The chapter also discusses the study setting, study population, sampling techniques, data collection tools, data analysis methods and tools and ethical considerations. **Chapter four** presents the findings and discusses the study findings concerning quality status for VGI in Malawi based on expert user narratives and methodical quality test. **Chapter five** presents a summary of the study, recommendations, and a conclusion to the study.

1.5 Chapter Summary

This chapter introduced the topic of this study and provided a general view of what is discussed in the study. It provided a comprehensive background to the study and presented the problem statement, research objectives, research questions, and the structural composition of the thesis.

CHAPTER 2

LITERATURE REVIEW AND CONCEPTUAL FRAMEWORK

This chapter presents a review of the literature on VGI. It analyses the VGI developments from the global scale to the local context. The chapter also defines and discusses related concepts, including data and data quality, spatial data quality and Juran's Trilogy from which a conceptual framework is drawn.

2.1 Data quality

With huge amounts of data being generated worldwide, one of the top topics on data is the question of how much good data is out there. Universally, good data is believed to be data that conforms to the standards of data quality (Harris, 2011; Mahanti, 2018). Data quality describes the state of accuracy, completeness, timeliness, validity and consistency of data that makes it fit for its purported use (Lee et al., 2006; Longley et al., 1999). The fitness of data for use also incorporates how accurately the data represents the real-world idea. Unfortunately, it is believed that more than 25% of critical data in the world's top companies is flawed and that 50% of records in a customer's database may be obsolete and inaccurate within two years (Fan, 2015). Furthermore, most of the world's data is full of inconsistencies, duplicated, primarily incomplete, inaccurate and obsolete at the time of use (Bielecka & Burek, 2019). Regrettably, there is no single fix for most data quality challenges because data quality is a multi-dimensional concept comprising accuracy, completeness, conformity, consistency, coverage, timeliness, and uniqueness among others (EDM Council, 2021). Since the success of GIS operations on a particular set of spatial data is determined by the user of that data, the responsibility of ensuring data quality is gradually shifting from producer to consumer in what is called "Fitness for Use" (Longley et al., 1999). It follows then that any usable data is quality data from the lenses of the user (Mahanti, 2018).

In business, high-quality data is a highly regarded asset that ensures customer satisfaction and increased revenue and is a tool for competitive advantage (Lee et al., 2006). At the same time, the cost of dirty data is huge. In 2015, it was estimated that American businesses missed $600 billion of revenue due to poor data quality (Fan, 2015). In the modern world, the success of new inventions and technologies highly depends on the quality of data (Lotame, 2019) and technologies such as machine learning, automation and artificial intelligence also thrive on quality data (Hillier, 2021). With quality data in projects and workplaces, there is less backtracking due to errors and fewer double checks lead to substantial time savings, increased productivity, and profits (Wassén, 2019).

2.2 Spatial Data

Often referred to as geographic information or geospatial data, spatial data can be defined as data referencing a particular geographical location (Shin et al., 2018). Various authors have discussed different types of spatial data based on different metrics. These metrics include but are not limited to how spatial data is captured (Olaya, 2018), who captures it (Longley et al., 1999) and the format in which it is stored, which includes vector and raster formats (Spencer & Wilkes, 2019). Another common classification of spatial data identifies two types of spatial data by the nature in which they represent what they capture (Olaya, 2018). This classification identifies spatial data as being geographic and geometric. Geographic data is considered as information that is mapped around the earth with emphasis on the spherical shape of the earth (Zola & Fontecchio, 2021). Geographic data expresses the relationship between a specific location or object to the idea of latitude and longitudinal. In addition, Geographic data is broken down into spatial and thematic components (Shin et al., 2018). The spatial component is concerned with the where of an object, while the thematic component is concerned with the what of the spatial component of the object (Greenfeld, 2013). On the other hand, geometric spatial data is considered as spatial data that is presented on a two-dimensional flat surface (Zola & Fontecchio, 2021). Some definitions of geometric spatial data categorise it as a sub component of cartesian systems that use coordinates to measure the position of a point from a defined origin along perpendicular axes with the main goal of mapping the earth on a flat surface without distortion (Linz, 2022).

Within the historical context, paper-based maps were the only source of spatial data at the emergence of GIS practice (Appel, 2019). Over the years, various techniques for spatial data acquisition have emerged (Diggelen & Enge, 2015). Remote sensing has been identified as a source of spatial data subdivided into electromagnetic radiation, sensors, and photogrammetry (Olaya, 2018). Other common sources of spatial data include printed cartography, GPS, Metadata and VGI (Appel, 2019). In recent years, spatial data sources are classified as either authoritative or non-authoritative (Dorn et al., 2015). Within this context, spatial data is available in two distinctive modes of commercial and free spatial data. For cost-associated categories, the sources can be subdivided by the type and category of data they host. Some sources can exclusively be host to point of interest data, imagery, street data and demographics while other sources host different types of spatial data at the same time (Greenfeld, 2013). The different categorisations of spatial data show that users have a variety of spatial data sources at their disposal.

2.2.1 Spatial data quality

Various scholars and GIS practitioners agree that there is no such thing as the perfect GIS data (Pascual, 2011). This sentiment is echoed and bemoaned in the inherent complexity of the actual geographical world, which makes a dream for its perfect digital representation almost impossible (Longley et al., 1999). Consequently, the quality of spatial data and its presentation, such as maps, will always face some percentage of doubt regardless of who and where it is generated (Crowe, 2017). The reality of imperfect spatial data makes it worse for VGI, considering that a significant portion of the scepticism on its quality emanates from the fact that VGI contributors are largely not verified and the quality of the data itself is not checked as would be in commercially organised mapping exercises (Fonte et al., 2015). This argument portrays the image that commercially organised mapping exercises are bound to produce high-quality spatial data compared to citizen participatory mapping. However, it is evident that any spatial data, regardless of its source, has inevitable quality concerns (Crowe, 2017). Concerns for spatial data quality in recent years date back to the late 90s and have not been limited to VGI. These concerns arise from increased data production by the private sector and its exclusion from the quality standards conformity requirements, growth in the adoption of GIS as a decision-support tool and an increased reliance on secondary data sources (Longley et al., 1999; Veregin, 1999).

While there is a holistic view of spatial data quality that is characterised by the phrase “fitness for use” (Greenfeld, 2013), spatial data quality remains a multi-dimensional concept just like any other type of data (Bielecka & Burek, 2019). A dimension can be defined as a quantifiable characteristic of an object (Black & Nederpelt, 2020) and in spatial data quality, such quantifiable characteristics include accuracy, correctness and validation stamps (Dasgupta, 2012). Other authors also identify completeness, consistency, usability, and temporal quality as elements that make up spatial data quality, however, most of the discussions on spatial data quality is based on spatial accuracy as a key dimension of quality (Spencer & Wilkes, 2019). While recognising that spatial accuracy is a key dimension of data quality it is important to appreciate spatial data quality as the multi-dimensional concept that it is. It is thus erroneous to address data quality issues as a single-faceted concept (Fan, 2015). Global data management think tanks mainly address seven key data quality dimensions as part of their data management capability assessment models (EDM Council, 2021). These dimensions include accuracy, completeness, conformity, consistency, coverage, timeliness, and uniqueness (EDM Council, 2021). Collectively, over 50 data quality dimensions, including accuracy, clarity, availability, completeness, currency, validity, traceability, and uniqueness are discussed as vital data quality dimensions (Black & Nederpelt, 2020).

In the exploration of the various dimensions of spatial data quality, it is also important to understand what the commonly occurring quality dimensions translate to. For instance, accuracy has been defined by numerous authors as the extent to which data or information on a map or any digitally presented form matches the ground truth or accepted values (Black & Nederpelt, 2020; Cooper et al., 2012; Fonte et al., 2015; Veregin, 1999).

2.2.2 Spatial data quality testing

The concept of spatial data quality testing emphasises the need for spatial data to undergo examinations for quality within the confinements of data quality assessment approaches. Various spatial data quality testing approaches have been discussed (Du et al., 2016). Among the commonly cited approaches is the application of the International Organisation for Standardisation (ISO) standard for geographic data quality on the test data to check its conformity (Maulia, 2018). Another common approach is what is identified as the "trust" methodology which looks at the properties of who captures the data (Dasgupta, 2012). With the trust approach, there are two general classifications of spatial data collectors: those considered official sources, commonly referred to as authoritative sources, and those considered as experienced casual collectors (Dasgupta, 2012; Hunter et al., 2003).

While one set of scholars is interested in frameworks for spatial data quality and testing from the producer's perspective, other spatial data testing methodologies have shown interest in the quality issues that can be introduced during spatial operations (Senaratne et al., 2016). Furthermore, these methodologies also look at the exacerbation of quality issues from the source point of spatial data down the hierarchy of spatial operations. The argument around this school of thought is that when quality issues, also called errors, are not taken care of at a particular level, they only grow in nature via transformative operations of GIS (Heuvelink, 1999; Santini Ron, 2021).

2.2.3 Errors and error sources in spatial data

Errors in general can be defined as conclusions demonstrably incorrect from a rational point of view (Brown et al., 2018). That is, anything in the range of mathematical mistakes, incorrect statistical procedures and statements not supported by data can be called errors. On the other hand, spatial data errors are defined as the imprecision and inaccuracies of spatial data (Pascual, 2011) where precision refers to a GIS database's level of exactness and measurement for its description and accuracy the degree to which information on a map or database matches true values (Goodchild & Li, 2012). By dissecting the concepts of precision and accuracy, further subcategories of spatial errors referred to as location and topological errors are discovered (Hunter et al., 2003).

Although there are more apparent sources of errors in spatial data, including the age of data, area cover, map scale, the density of observation, relevance, format, accessibility and cost, the largest source of errors is the data itself (Bielecka & Burek, 2019; Brown et al., 2018).

2.2.4 Error propagation in spatial data

In spatial data processing, when data stored in a GIS database contain errors and is used as an input to a GIS operation, the errors are transferred to the output of the operation in a process called error propagation (Pascual, 2011). Consequently, when errors are propagated in spatial operations, the output of such operations may not be as reliable (Fonte et al., 2015). It is important to note that this transfer of errors is very generic to any spatial data, including that collected by professionals as demonstrated in Figure 2.1. For data like VGI, whose quality is more questionable, there is a greater need to understand more about the types of errors associated with it, their sources and techniques for dealing with error propagation (Goodchild & Li, 2012).

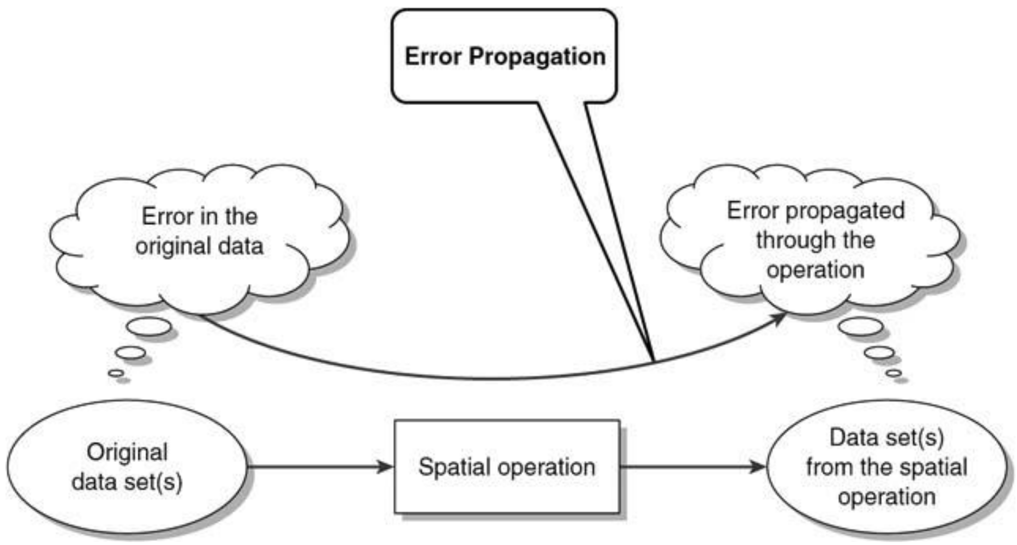

Figure 2.1 A visualisation of error propagation

2.3 Volunteered Geographic Information - VGI

Volunteered Geographic Information (VGI) is defined as the geographic data collected by untrained private citizens with little to no formal qualifications in creating and distributing spatial information (Goodchild, 2007; Yan et al., 2020). Over the past decade, there have been various debates around what VGI is and its relationship to various terms used interchangeably with VGI, including participatory mapping and crowdsourced spatial data. In the course of these debates, one underlying characteristic of VGI observed in various studies is the involvement of non-skilled volunteers in the generation of such data (Cooper et al., 2012; Goodchild, 2007; Ferster et al., 2018; Yilma, 2016; Young et al., 2020). Within this broader understanding there are two classifications of VGI, namely explicit and implicit VGI (Senaratne et al., 2016). The former is about map-based data such as OSM, while the latter is concerned with text-based data such as tweets and image-based data such as Flicker. Using a narrowed down perspective, VGI has been branded using the Coote and Rackham eight aspects of the data namely source, purpose, collection, cost, management, licensing, access and quality (Coote & Rackham, 2008). These eight aspects of data that distinguish VGI from traditional spatial data are summarised in Table 2.1.

Table 2.1 Coote and Rackham aspects of data

Aspect	**Description**
Source	VGI is a product of voluntary collaborative activities dominated by inexperienced communities without any financial or monetary reward.
Purpose	The purpose and motivation for the creation of VGI by its users is very diverse and usually unique to individual contributors or their communities
Collection	The collection of VGI is highly random, real-time, and prevalently powered by the local knowledge of contributors. Its distribution across space and time may be clustered and exhibit characteristics of irregular coverage influenced by the frequency of activity.
Cost	While some associated costs may be other than monetary and dependent on the custodian, VGI is available as an open-source resource. It is believed that the free distribution of information may boost user participation.
Management	VGI is managed either collaboratively by members of a VGI community or by custodians who mobilise VGI communities.
Licensing	VGI boasts of remarkably high shareability but pays less attention to legal issues associated with users' privacy and security, attracting much criticism from scholars.

Aspect	Description
Access	VGI is highly accessible, in many instances offering users the freedom to copy, share and transmit VGI data if they credit its contributors
Quality	VGI often falls short of metadata describing its quality which brings about a lot of scepticism on its reliability.

VGI can further be classified into three other categories based on the motivation of the contributors (Yilma, 2016). These categories include civic VGI, social networking VGI and market-driven VGI.

2.3.1 VGI Quality

Fundamentally, VGI suffers from a presumed lack of quality largely attributed to position accuracy and content inaccuracy (Tusker et al., 2018), with inexperience and lack of verification for contributors among the reasons. Contributors can potentially share misleading and false information (Vosoughi et al., 2018), and all kinds of errors can be introduced at any stage of the data handling and storage (Santini Ron, 2021) before being made available to the user (Pascual, 2011). Consequently, that calls for methodologies for testing the quality of VGI by identifying the types and sources of errors in VGI data. In addition, techniques that can be applied to limit the transfer of errors from where the data is generated to users must be identified (Ferster et al., 2018). In all this, there must also be consideration for the players who create and manage the platforms on which VGI data is shared (Tusker et al., 2018).

Literature identifies two main approaches to assessing VGI quality (Fogliaroni et al., 2018). The first one being based on the assumption that authoritative data is always of high-quality as compared to VGI (Maulia, 2018; Senaratne et al., 2016). From this assumption, VGI quality is consequently tested by comparing it against authoritative data. This approach is believed to address the key quality dimensions of accuracy, validity, completeness and timeliness, among other dimensions of quality (Fogliaroni et al., 2018). The method of comparing VGI data to authoritative data is supported by various studies with some calling it the matching of crowdsourced and authoritative geospatial data where corresponding spatial data features are identified between different spatial datasets (Du et al., 2016). The other approach is where the quality of VGI is assessed by its evolution, addressing the changes that may happen to data overtime (Fogliaroni et al., 2018).

2.3.2 VGI projects in Africa

Various VGI projects and initiatives have shaped Africa in the last two decades. Some of the notable projects include Tracks4Africa, established in the early 2000s and the Southern African Bird Atlas Project 2 (SABAP2. In addition, there is the Open Cities Africa initiatives by Open Data for Resilience Initiative (OpenDRI), which has implementations for Mapping for Resilience in Uganda, Mapping for detecting invasive armyworm species in Malawi, both supported by the World Bank, Mapping for disaster risk management in Zanzibar and Mapping for Urban and Coastal Flooding in Seychelles. All these are projects that involve voluntary community participation in the collection and dissemination of spatial information. In collaboration with the Humanitarian OpenStreetMap Team (HOT), the Open Mapping Hub – Eastern and Southern Africa works with communities and organisations across twenty-three countries in open spatial data activities. VGI contributions and projects in developing countries (a feature that characterises most African countries) are mostly sporadic and usually a response to some disaster or humanitarian crises (Mahabir et al., 2017).

In the Southern African region, the concept of VGI surfaced in the early 2000s (United Nations Economic Commission for Africa, 2017). Since then, several short-term and long-term VGI projects have taken place. Among the significant ongoing VGI projects in the Southern African region are extensions of global projects such as OSM and Wikimapia (Yilma, 2016). In collaboration with the Open Mapping Hub, HOT supports projects in Malawi, Namibia and Zambia, where communities and organisations create and update open map data in OSM. As discussed in chapter one, SABAP2 is also one of such successful VGI projects exclusive to the Southern African region. The Open Cities Africa initiative also runs VGI-related projects in a few countries, including Tanzania and the islands of Seychelles and Madagascar (Global Facility for Disaster Reduction and Recovery, 2020). Another notable VGI implementation in Southern Africa is the iCitizen project in Southern Africa (Yilma, 2016). This is a VGI project whose aim is to collect data on public service and infrastructure problems to help the authorities make informed decisions on tackling problems (Yilma, 2016).

VGI in Malawi surfaced around late 2011 through a collaborative project between disaster risk and management stakeholders who had an interest in the mapping of the flood-prone southern district of Nsanje (Mhone, 2021). These stakeholders included the Department of Disaster Management Affairs (DoDMA), the Department of Surveys and the World Bank. While this project started with the agenda of spatial data collaboration among authoritative spatial data custodians in Malawi, it evolved and bought into the concept of open spatial data powered by volunteer and community participation (Mhone, 2021). At the end of the year 2012, the Malawi Spatial Data Platform was born. Within the last decade, strides have been made to establish and consolidate the position of VGI under the banner of Open Spatial data.

In 2015, the World Bank trained communities in 15 flood-affected districts to assist in mapping affected areas. In 2016 flood mapping continued with more technical assistance from HOT (Mhone, 2018). By this time, MASDAP had also become a prominent host of crowdsourced spatial data in Malawi. It was also within the same year, 2016, that Youth Mappers Malawi, operating under the University of Malawi was born. The subsequent years also saw some remarkable developments in VGI practice platforms. mHub partnered with MASDAP on several mapathon projects from which the Malawi Mappers community was born (Mhone, 2021). In no time, Malawi Mappers partnered with Google and launched Google local guides mapping community for Malawi.

The year 2018 saw the birth of more Youth Mappers chapters and the growth of the Malawi OSM community beyond academia. Through these entities, international organisations such as Red Cross and Doctors without Borders could obtain VGI for their response programs to the disaster of Cyclone Idai (UNDP, 2020). In 2020, other significant developments towards improved and expanded open spatial data capabilities came to fruition. Map Malawi project was born, and UNICEF established the Africa Drone and Data Academy (ADDA) in Malawi. At the end of 2021, efforts were being made to formalise OSM Malawi and make it the mother body for all crowd-sourced spatial data initiatives (Mhone, 2021).

This trail of developments in volunteer involvement in spatial data collection and sharing highlights the efforts made towards embracing VGI in Malawi (UNDP, 2020). While the concept of VGI is relatively in its development stages in Malawi (Gardner & Mooney, 2018), notable signs of interest have been ignited within the GIS and volunteer communities, as evidenced by the number of success stories of Mapathons that have been conducted across the country.

Through the Accelerator labs Malawi, the United Nations Development Program - UNDP facilitated the mapping of buildings and roads in Area 25 Township of the capital city Lilongwe (UNDP, 2020). In the year 2021, HOT statistics for Malawi indicate that Malawi had over 949 community mappers with slightly above 180,000 map edits, over 143,000 and 2,400Km building and roads mapped, respectively (HOT, 2021).

2.4 Conceptual Framework

The concept of data quality which concerns the level of accuracy, completeness and uniqueness of a particular data set can be described as a directly proportional concept. The higher the accuracy, completeness and uniqueness of a data set, the higher the data quality. In an effort to assess VGI quality from a different perspective, this study proposed a framework for determination of spatial data quality through partial adoption of the Juran’s Trilogy. From the wholistic view of Juran’s theory on quality (Juran & Godfrey, 1951), the underlying processes of quality planning, control and improvement demonstrate that quality can be achieved by design, therefore ensuring spatial data quality must be intentional.

Spatial data quality must be planned at the source by understanding the inherent quality issues that are attached to the particular spatial data source (Quality planning) (Juran & DeFeo, 2010). Since each spatial data source has its unique challenges, the source becomes an independent variable and the perceived quality issues a dependent variable in the quest to measure the level of data quality for any spatial data set. The reality that spatial data may not be perfect at the source demands provision of quality control measures as spatial data is being worked on at different stages of its life cycle (Quality control) (Juran & DeFeo, 2010). However, the quality of the data presented to the user will also depend on the spatial data quality management skills of the responsible data custodian and the efficiency of the quality control techniques used at the various stages of handling and processing the spatial data. The skills and techniques become the moderating variable in the proposed framework.

In summary, understanding the spatial data source and its quality issues determines what must be done to improve the quality of spatial data throughout its life cycle. In this context, the quality of VGI is a product of the relationship between the sources of the data as an independent variable (planned/design) and the perceived spatial data quality issues as a dependent variable, with both being moderated by the quality improvement technique and the skills of the VGI custodian (control). As the cycle repeats, new insights from previous quality inspections must be incorporated for continuous improvement (Juran & DeFeo, 2010). The proposed framework is visualised in Figure 2.2. This relationship among the identified variables provides a holistic approach to quality as per Juran's Trilogy which comprises quality planning, control and improvement (Juran & Godfrey, 1951).

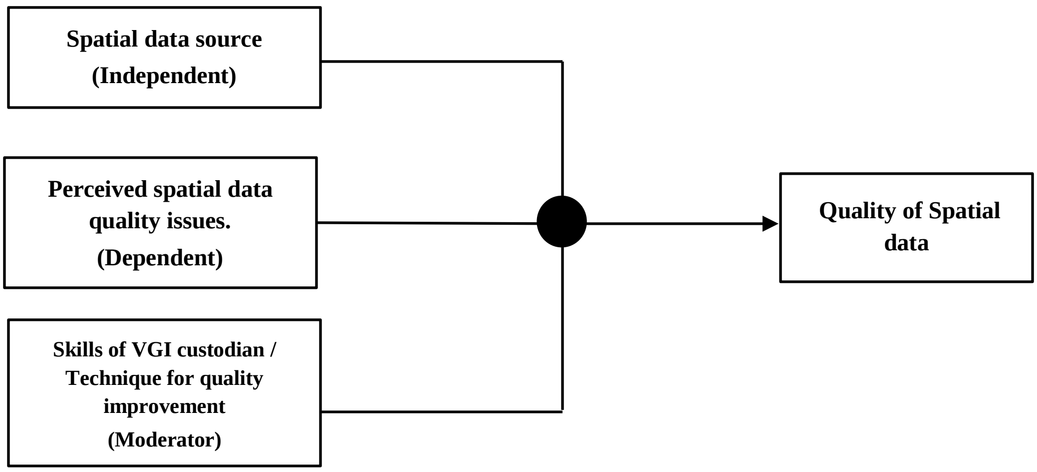

Figure 2.2: Conceptual Framework for Spatial Data Quality, Adopted from Juran's Trilogy

2.5 Chapter Summary

This chapter introduced the topic of this study and provided a general view of what is discussed in the thesis. It provided a comprehensive background to the study and presented the problem statement, research objectives, research questions, and the structural composition of the thesis.

CHAPTER 3

RESEARCH DESIGN AND METHODOLOGY

This chapter explains how the study was conducted. It outlines the study's design, philosophy, approach, and strategy. Furthermore, it discusses the instruments used for collecting data and how the collected data was analysed. The rationale for the choices is discussed by highlighting the choice's suitability in tackling the study objectives, fitness to the entire study design, and benefit to the rest of the study.

3.1 Research Philosophy

Different research philosophies, including positivism, ontology and pragmatism can be adopted and ably deliver on different research problems, this study followed the interpretivism philosophy. Thompson (2015) states that the interpretivism research philosophy is a viewpoint which assumes that "social reality is not singular or objective but is rather shaped by human experiences and social contexts" (Thompson, 2015). It can therefore be added that individuals are intricate and complex since different people experience and understand the same "objective reality" in different ways and have personal reasons for their reactions (Gemma, 2018). Since quality is a very subjective concept (Juran & DeFeo, 2010), the interpretive philosophical approach and its underlying principles were best fit for a better exploration of the subject of Volunteered Geographic Information quality. Data or information quality is categorised as a multi-dimensional or multi-faceted concept. Among the many dimensions of data quality, scholars agree on six primary dimensions or attributes that characterise data quality (Brown et al., 2018; Sarfin, 2021). These primary dimensions include accuracy, validity, timeliness, completeness, consistency and uniqueness (Ballatore & Zipf, 2015; Fan, 2015).

While all these dimensions can be tested individually or in combination, they are all measured from the data's ability to satisfy the needs of a particular user (Juran & DeFeo, 2010; Lee et al., 2006; Veregin, 1999). Ultimately, the classification of that data set on the quality scale will likewise be influenced by and derived from the user's interpretation of his or her experience with that data set. All forms of data may be expected to meet some predefined quality measures and standards, but data quality requires judgement (Hillier, 2021).

Firstly, in a GIS, the attribute accuracy of a geographical feature will constantly be subject to subjective interpretation derived from experiences. A feature in a dataset whose land use attribute is labelled as residential based on the knowledge of the data producer would rightly be identified as such by one user, yet another will identify the same as being incorrect, with both observations attributed to the user's experience with ground truth and interpretation of that experience with the feature not preceding the influence of time on the user's knowledge. Consequently, the attribute accuracy of the data will be a reasonable outcome of their experience with the feature. It must be noted that while there may be no correlation between time and location, despite attributes often changing over time (Mooney & Corcoran, 2012).

Secondly, spatial data accuracy is affected by formatting that may include the date and time formatting. For instance, a data set with European date formatting being used by an American with no prior European experience, and the date is of primary interest, is a likely cause of interpretation dilemma. This is also called semantic ground, which describes the conceptualisation of things and a common language to describe them (Ballatore & Zipf, 2015).

It is evident therefore that to understand the concept of data quality, spatial data quality, or VGI quality in the case of this study, the interpretational contexts, and experiences of the population on the use of VGI was important. The interpretive paradigm is underpinned by observation and involves a less structured methodology that facilitates close interaction with the study sample (Antwi & Hamza, 2015).

3.2 Research Approach

Mixed methods research (MMR), is a method that involves collecting, analysing, and integrating qualitative and quantitative data in a single study (Leavy, 2017). In this approach, the quantitative and qualitative phases of the study inform the other in both directions, and the results are integrated (Wisdom & Creswell, 2013). In phase one of the study, qualitative interview data was collected and analysed. In the next phases, a variable for a quantitative test was identified and quantitative data in form of GPS coordinates were collected and tested for quality in phase two and three respectively. The last phase involved interpretation and validation of the results from phases one and three. Figure 3.1 shows a complete MMR implementation flow-chart for the study.

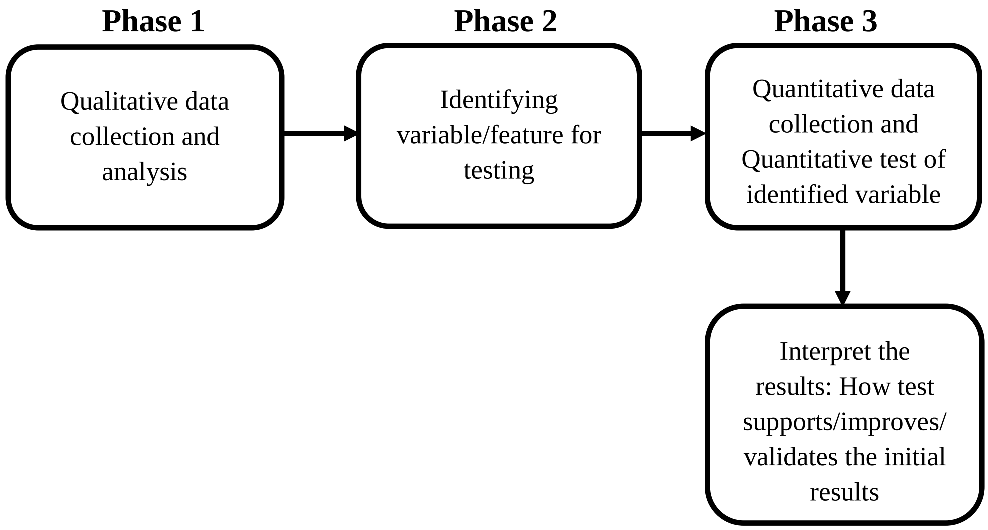

Figure 3.1 Exploratory sequential design

The researcher adopted a case study strategy which is an intensive longitudinal study that uses an in-depth examination of a small population (Bhattacherjee, 2012). This strategy seeks to provide a complete and accurate account (Marczyk et al., 2005) and derivation of detail and contextualisation of inferences (Žukauskas et al., 2018). This strategy enabled the researcher to ask "how", "where", and "why" types of questions as an outsider in alignment with the fundamental principles of the interpretivism philosophy that guided the study.

3.3 Study site

This study was conducted in Malawi in the southern part of the African continent with a land coverage of 118,484 square kilometres and an estimated population of slightly over 19.3 million in the year 2020 (World Bank Group, 2022) . Malawi lies at 13.2543º S and 34.3015º E in the great rift valley and has a tropical climate. Malawi belongs to the Southern African Development Community – SADC, the region's inter-governmental organisation that fosters integration and social-economic, political and security cooperation among its 16 member countries. Figure 3.2 shows the position of Malawi within the Southern African region.

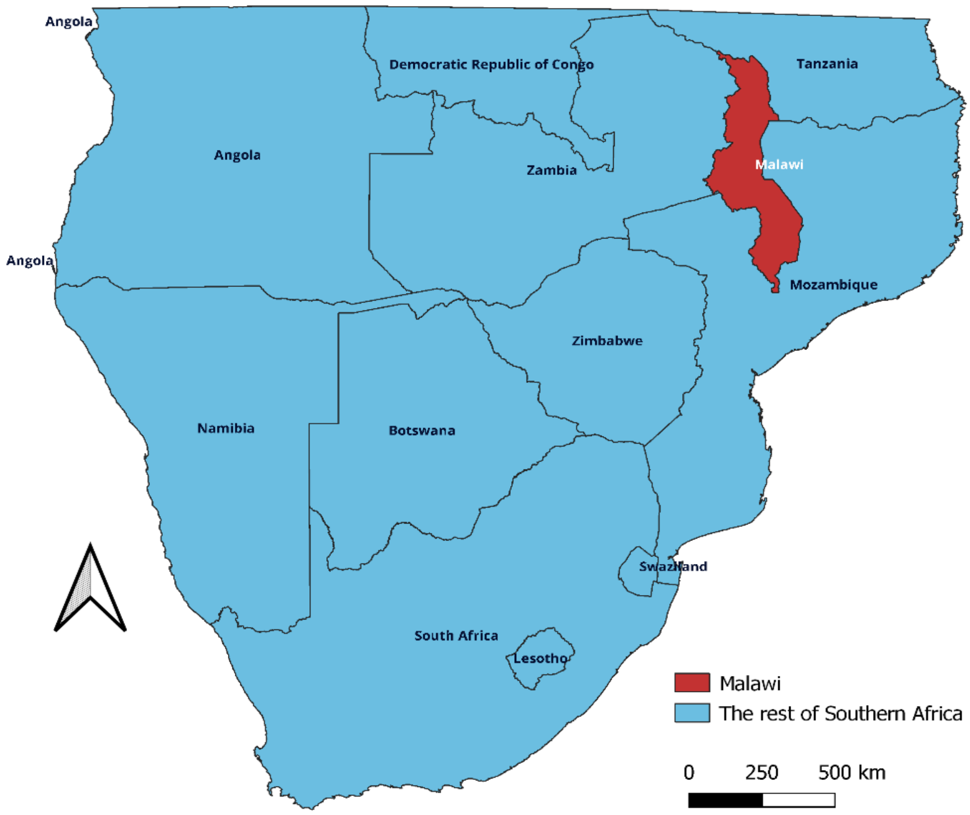

Figure 3.2 Map of Southern Africa showing the position of Malawi (red) in the region– Produced in QGIS

3.4 Population and Sampling design

3.4.1 Study Population

Hu (2014) defines a study's targeted population as a population subsection from which a sample is obtained. The study's targeted population was two-dimensional. On the one hand, it consisted of VGI expert users identified by their professional careers and experiences in GIS, spatial data management, utilisation of VGI, and involvement in open spatial data projects. Special attention was given to the expert users' interests and interaction with VGI from governmental, non-governmental and private institutions concerned with spatial data collection, hosting, applications and systems development and training. Targeted participants were primarily identified by their professional titles relevant to the study. Table 3.1 presents the study participants' profiles and the type of VGI they had used in their careers.

Table 3.1 Participants' profiles and type of VGI used.

Participant's alias	**Professional Job title**	**Level of Education**	**Type of VGI used**
David	GIS Expert/Data manager	Master's Degree	Spatial Framework data (Spatial Datasets) and thematic data
George	Senior Systems / Software Developer	Bachelor's Degree	Spatial Framework data (Spatial Datasets)
Angela	GIS Expert/Data manager	Master's Degree	Spatial Framework data (Spatial Datasets)
Alinafe	GIS/Cartography Officer	Master's Degree	Spatial Framework data (Spatial Datasets), thematic data and Gazetteer data
William	GIS Consultant / Senior lecturer	Doctor of Philosophy	Spatial Framework data (Spatial Datasets), thematic data and Gazetteer data

The other dimension of the population were the primary schools whose GPS coordinates were to be methodically tested for quality within selected datasets.

3.4.2 Sampling methods

In this study, both non-probability and probability sampling methods were used in the qualitative and quantitative phases of the study respectively. Under non-probability sampling, there are various types of sampling methods: convenience, consecutive, snowballing, quota, and purposive. The researcher used purposive and snowballing non-probability sampling methods as described by Lohr (2019) to select participants who were interviewed in the qualitative phase of the study.

Levy (2008) defines probability sampling as a method in which every member of a study's targeted population has a chance of being chosen to represent the population. For the quantitative part of the study, Systematic probability sampling was employed to select primary schools from both the OSM Education Facilities for Malawi and

Ministry of Education Primary schools' datasets. The GPS coordinates for the selected schools were used in the spatial quality test that was conducted.

3.5 Data Collection

The underlying principles of qualitative research approaches are rooted in their ability to extract depth of meaning, people's subjective experiences, and the process through which they construct such meanings (Leavy, 2017; Wisdom & Creswell, 2013). To complement the overall approach and strategy employed in the study, semi-structured interviews were used to collect the qualitative data. This study involved engaging and extracting experiences from participants through open conversations. Additionally, they provided room for a flexible enquiry, which helped in obtaining an in-depth understanding and opinions of the participants on VGI within the participant's natural settings. To ensure a higher validation of meanings extracted from the first interviews, the study interviewed each participant twice, this allowed the researcher to share interpretations of the initial interviews and AHP results with participants as recommended by Pessoa et al., (2019)

The study singled out six key dimensions of quality on which respondents were asked to express their opinions on the relative importance of each data quality dimension against the other. The dimensions of quality included, accuracy, validity, timeliness, completeness, consistency and uniqueness. These six dimensions formed a fifteen-pairs pairwise comparisons matrix. Table 3.2 shows the pairwise comparisons utilised in the study.

Table 3.2: Pairwise comparisons for Data quality dimensions in VGI

Accuracy	9	7	5	3	1	3	5	7	9	Validity
Accuracy	9	7	5	3	1	3	5	7	9	Timeliness
Accuracy	9	7	5	3	1	3	5	7	9	Completeness
Accuracy	9	7	5	3	1	3	5	7	9	Consistency
Accuracy	9	7	5	3	1	3	5	7	9	Uniqueness
Validity	9	7	5	3	1	3	5	7	9	Timeliness
Validity	9	7	5	3	1	3	5	7	9	Completeness
Validity	9	7	5	3	1	3	5	7	9	Consistency
Validity	9	7	5	3	1	3	5	7	9	Uniqueness
Timeliness	9	7	5	3	1	3	5	7	9	Completeness

Timeliness	9	7	5	3	1	3	5	7	9	Consistency
Timeliness	9	7	5	3	1	3	5	7	9	Uniqueness
Completeness	9	7	5	3	1	3	5	7	9	Consistency
Completeness	9	7	5	3	1	3	5	7	9	Uniqueness
Consistency	9	7	5	3	1	3	5	7	9	Uniqueness

The relative importance pairwise comparison chart in Table 3.2 takes the form of a competition draw in which every component faces the other once and a relative importance assessment is drawn from the responses provided by the respondents where only one digit is circled on each comparison row.

To assess the quality of VGI against authoritative spatial data, quantitative data in form of GPS coordinates were collected from the OpenStreetMap education facilities for Malawi dataset and the Malawi Ministry of Education Primary Schools dataset published on Humanitarian Data Exchange (HDX) and MASDAP respectively. The coordinates for various features were systematically selected to serve as input data for the spatial data quality test exercise. The details of the datasets are summarised in Table 3.3.

Table 3.3: Account of published spatial datasets collected from secondary sources.

Name of the data set	**Publisher**	**Type of data file**	**Source Domain**	**Attribution**
OSM Education-Facilities for Malawi	Humanitarian Data Exchange (HDX)	Shapefile	VGI	OpenStreetMap
Ministry of Education Primary Schools	MASDAP	Shapefile	Authoritative	Ministry of Education -Justin Saunders (GIS Expert/ Consultant)

3.6 Sample size

The sample size for the qualitative component of this study was five, which was arrived at using the “Guidelines by experts” approach, or “Rule of thumb” (Kumar et al., 2020).

This aligned with mixed-method approach employed (Creswell D. J., 2009; Creswell & Creswell, 2018).

Quantitatively, the population consisted of various spatial features such as roads, points of interest and buildings mapped by VGI contributors and collectively making up various VGI datasets that can be tested for quality. Since VGI is relatively new in Malawi, the study set the targeted feature count at 2000 features for every selected dataset. Cochran formula (Cochran, 1977) was then used to determine how many geographical features represented by planimetric coordinates were to be tested as shown by:

$$n' = \frac{n}{1 + \frac{z^2 \times \hat{p}(1 - \hat{p})}{\varepsilon^2 N}} \tag{3.1}$$

where

Z is the Z score

Ɛ is the margin of error

N is the population size

$\hat{p}$ is the population proportion

n' is the number of geographical features represented by the planimetric coordinates.

Table 3.4 shows the parameters set for calculating the sample size in the study as guided by the Cochran formula. Based on this, a minimum sample size of 27 primary schools was obtained for the data quality test. The study selected 50 primary school points from both data sets. The primary schools are listed in Appendix G.

Table 3.4: Sample size calculation parameters

PARAMETER	**VALUE**
Margin of Error	7%
Confidence Level	90%
Z-Score	1.65
Population Proportion	5%

During the preliminary processing of the datasets, an examination of the attribute table of the OSM education facilities for Malawi dataset in QGIS revealed that the features in the dataset were only referenced by their OSM ID; hence a further mapping exercise of the OSM IDs to their respective GPS coordinates was done using MapCarta.

3.7 Data Analysis

The study applied qualitative and quantitative data analysis methods following the *QUAL to Quan* exploratory sequential design (Edmonds & Kennedy, 2017) specifically inductive narrative analysis, descriptive statistics, and Root Mean Squared Error (RMSE).

The inductive narrative analysis approach was ideal for the study due to its strong association with qualitative methods, ability to digest subjective reasoning and in-depth extraction of meaning from participant stories (Reichertz, 2014). The study focused on discovering the expert users' perceptions and interpretation of experiences with VGI quality before and after using VGI, with special attention on how those narratives were told. The narrative analysis process started with the transcription of interviews from audio to readable text using openly available oTranscribe, Google docs, as well as Microsoft word in combination with Voice in Voice plugin for Google Chrome and VB-Cable virtual Audio device. Subsequently, narrative blocks were coded as "life events" by identifying verbal constructs comprising entrance and exit talks. In its nature, narrative analysis has a **dual** layer of interpretation between what is said and how it is said (Riessman, 1993). To extract depth of meaning between these layers verbatim transcripts of the interviews were used to capture filler words, pauses, stray utterances and phrases such as "hmm", "well", "No way!", "exactly", "you know", and "in the end" to establish entry points, main points, exit points and, more importantly, the tones in the narrations. Consequently, for each inductively formed life-event code, e.g., "Narrative about adopting VGI for GIS practice", emerging narrative themes were identified. The identified themes from the narratives assisted in filtering through the participants' stories to establish major similarities and differences between their narratives about their experiences with VGI quality.

Secondly, identifying key data quality dimensions of VGI as considered by its expert users, required participants to respond to discrete relative ranking of the importance of the quality dimensions with as minimal bias as possible. To achieve this, Analytic

Hierarchy Process (AHP) was used. AHP, a Multi-Criteria Decision Analysis (MCDA) technique that provides a systematic approach to decision-making when ranking multiple criteria evaluates available alternatives, harnesses the ability of humans to make comprehensive decisions based on comparative judgements about small problems given little to no room for the combining of factors to be measured. AHP was ideal for the type of data involved in the study because it combines the precision of mathematics and subjectivity of psychology. AHP also provides unbiased weighted ranking for accuracy, completeness, validity, timeliness, consistency, and uniqueness. The AHP analysis was based on a two-stage hierarchical structure comprising the goal in stage zero and the criteria/dimensions in stage one as shown in Figure 3.3**.** For each pairwise comparison created on a survey questionnaire, the AHP used the scale shown in Table 3.5 to compare objectives.

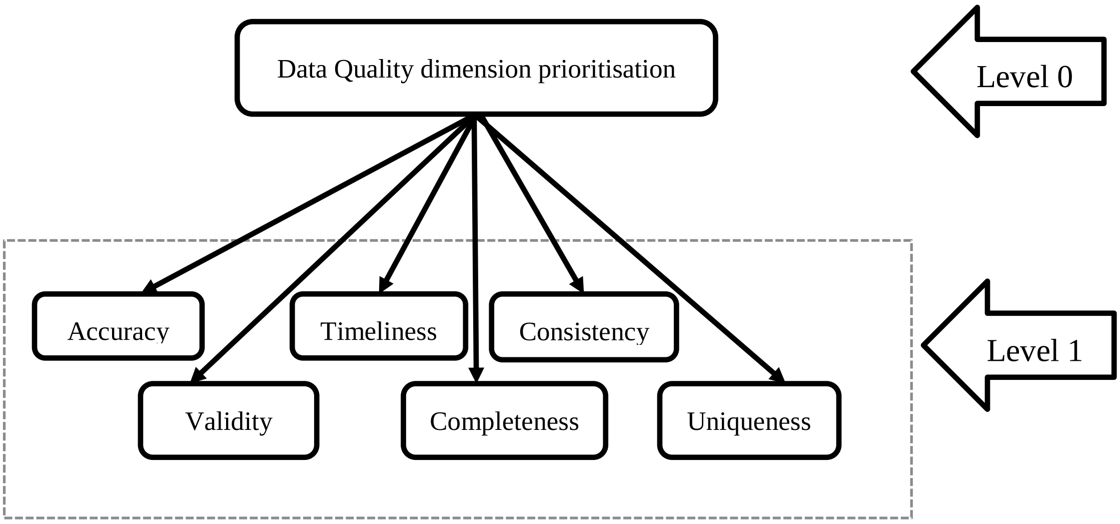

Figure 3.3: The hierarchical structure of AHP in the study

Table 3.5: Intensities in Saaty's scale of importance

Strength of importance	**Classification**	**Explanation**
1	Equal importance	Judgement entails equal importance between criteria
3	Moderate importance of one judgement over the other	The judgement shows a slight preference for one criterion over the other

Strength of importance	**Classification**	**Explanation**
5	Strong importance	Judgement strongly favours one criterion over the other
7	Extraordinarily strong importance	The judgement shows very strong favour of one criterion over the other
9	Extremely important	The judgement shows that one criterion is extremely preferred over the other
2,4,6,8	Intermediate values between two adjacent judgements	Choice shows compromise because an absolute verdict cannot be given between two adjacent judgements.

In processing the responses, the data quality dimensions were coded C1 to C6 for accuracy, validity, timeliness, completeness, consistency, and uniqueness, respectively, and the respondents were coded R1 to R5 for the first to the fifth respondent respectively. Responses of all participants were consolidated into one pairwise comparison matrix following which geometric means were then used to normalise the consolidated comparison matrix before calculating the final priority values of the data quality dimensions also called weights. To find the geometric mean for each criterion, a product of the values in each row of the matrix was calculated followed by the 3rd root of the calculated product which became the geometric mean. To calculate the weights, a sum of all geometric means for all data quality dimensions was calculated. Thereafter, each geometric mean was divided by the sum of the geometric means to get the weight for each data quality dimension.

Based on this matrix a two-stage hierarchy prioritisation model was developed for the calculation of weighted ranks and a Consistency Index. For acceptability, a consistency ratio (CR) was calculated by using:

$$CR = \frac{CI}{RI} \tag{3.2}$$

where CI is the consistency index defined by $\frac{(Principal\ Eigen\ Value - n\ )}{(n-1)}$ with n being the size of the comparison matrix and RI is the random index as defined in the random consistency index table (Saaty, 1988).

To assess the quality of the VGI, the National Standard for Spatial Data Accuracy (NSSDA), a statistical and testing methodology for positional accuracy of spatial data (Federal Geographic Data Committee, 1998), was used. It is an instrumental foundation of many national standards for different countries and is compatible with ISO-TC-211 standards relating to data quality (ISO, 2019). The NSSDA uses Root Mean Squared Error (RMSE) to estimate the positional accuracy of a spatial dataset. RMSE is defined as the square root of the mean of squared differences between planimetric coordinate values of a test dataset and planimetric coordinate values obtained from an independent source of higher accuracy for identical points (Federal Geographic Data Committee, 1998). In this study the data quality test was limited to horizontal positional accuracy with RMSE calculated for individual latitude and longitude as follows:

1) For latitudes:

$$RMSE_x = sqrt\left[\sum(X_{vgi,i} - X_{auth,i})^2 /n\right] \qquad (3.3)$$

2) For Longitudes:

$$RMSE_y = sqrt\left[\sum(Y_{vgi,i} - Y_{auth,i})^2 /n\right] \qquad (3.4)$$

where: $X_{vgi,i}$ and $Y_{vgi,i}$ are the planimetric coordinates of the i^{th} checkpoint in the VGI dataset, $X_{auth,i}$ and $Y_{auth,i}$ are the planimetric coordinates of the i^{th} checkpoint in the authoritative dataset, n represents the total number of checkpoints tested, and $1 \leq i \leq n$.

The horizontal RMSE was calculated as follows:

$$RMSE_r = sqrt\left[\sum\left((X_{test.i} - X_{check.i})^2 + (Y_{test.i} - Y_{check.i})^2\right)\right] \qquad (3.5)$$

where

$X_{test.i}$ and $X_{check.i}$ are the longitudinal coordinates of the i^{th} checkpoint in the test and check datasets while, $Y_{test.i}$ and $Y_{check.i}$ are the latitudinal coordinates of the i^{th} checkpoint in the test and check datasets

The prescription of the NSSDA states that when $RMSE_x$ is equal to $RMSE_y$, under the assumption that systematic errors have been eliminated as much as possible and that errors are normally distributed, the NSSDA horizontal accuracy, at 95% confidence level, be calculated as:

$$Accuracy_r = 1.7388 \times RMSE_r \quad (3.6)$$

In the test, $RMSE_x$ was not equal to $RMSE_y$, but $RMSE_{min}/RMSE_{max}$ was between 0.6 and 1; hence a Circular Standard Error (CSE) was calculated with error at 39.35% confidence level, using:

$$CSE \sim 0.5 \times (RMSE_x + RMSE_y) \quad (3.7)$$

After finding the Circular Standard Error, the horizontal accuracy according to the NSSDA was approximated using:

$$Accuracy_r \sim 2.4477 \times 0.5 \times (RMSE_x + RMSE_y) \quad (3.8)$$

3.8 Ethical considerations

The study's assessment of its data and participants requirements concluded that it was not a data sensitive study hence no internal research board special approvals were obtained. Nevertheless, necessary permissions to engage participants within the stipulated permissible provisions of academic research were obtained and verbal or written consent for voluntary participation of participants was sought. All the participants provided verbal consent to be involved in the study. In addition, participants' identities remained discrete, and anonymity was achieved using codes in the AHP analysis of respondents and pseudonyms in the analysis of participant's narratives. The study also ensured the right handling of collected study data by ensuring strict security of data storage media and all physical evidence such as printed interview transcripts and notes and proper disposal of the data at the end of the study (Morse & Niehaus, 2016). Lastly, the study results were shared with the participants.

3.9 Chapter Summary

This chapter discussed how the research study was conducted. It began with the study's research design, which consisted of the research philosophy, approach and strategy adopted. The chapter also discussed the study population, the instruments used for data collection and how the collected data was analysed. The rationale for the choices was discussed by highlighting their suitability in tackling the specific research objectives and their fitness for the rest of the study.

CHAPTER 4

FINDINGS AND DISCUSSION

This chapter presents the findings and discussion of the study in line with the study objectives. The discussion is based on results from interviews that were conducted, AHP analysis and the methodical quality test that was conducted.

4.1 VGI quality perceptions and observations by its expert users before and after use

Through the expert narratives the study identified five narrative blocks coded as life events and fifteen themes emerged from these life events. These narrative blocks and themes are summerised in Table 4.1. Non computational efforts of measuring the quality of spatial data require judgement which is shaped by the experience each unique individual has with that particular type of data (Hillier, 2021; Thompson, 2015). Actual encounters with particular types of spatial data such as VGI can therefore be considered an opportunity for building a perspective and perception that may be considered authentic and uninfluenced. Without those, the narrative about the quality of spatial data is driven by assumptions.

The narrative blocks formulated in this study were instrumental in understanding the experiences that shaped how each of the expert user participants perceived the quality of VGI. Without these, the conclusions drawn from the expert user narratives would be unfounded.

*Table 4.1**:** Narrative blocks (Life events) and emerging themes*

Narrative block coded as life-event	**Emerging narrative theme**
Narratives about the expert's journey into GIS	*Passion* *Scheme* *Destiny*
Narratives about pre-exposure VGI quality perception	*Negativity* *Positivity* *Curiosity*

Narrative block coded as life-event	Emerging narrative theme
Narratives about deciding to try VGI.	*Dilemma* *Directive* *Status quo* *Riding the tide*
Narratives about post-exposure VGI quality observation	*Amazement* *Satisfaction*
Narratives about adopting VGI for GIS practice.	*Comfortability* *Scepticism* *Work in progress*

While the five narrative blocks contributed to the researcher's understanding of the experts' journey into VGI practice, the study identified two key narrative blocks that were directly linked to the perceptions the participants had on VGI quality. These narrative blocks were the narratives about pre-exposure VGI quality user perception and the narratives about post-exposure VGI quality observations.

4.1.1 User perception of VGI quality before exposure

In this study, participant's previous exposure to stories about the quality of VGI was found to have a bearing on how they perceived the quality of VGI before going into practice. Participants mentioned the influence the exposure had to their understanding of VGI quality. From the narrative block of "pre-exposure VGI quality user perception" three themes of negativity, positivity, and curiosity emerged.

The study has shown that the participants' first impression of VGI quality before practice was negative. Participants believed that VGI was very poor-quality spatial data. This impression was attributed to the influence that authoritative practitioners had over the narrative of VGI quality and participants' limited access to VGI itself at the beginning of their careers and interaction with VGI. Most participants felt very discouraged about the quality of VGI as they were getting into VGI practice.

> *I was very hesitant to pursue OSM and the use of crowd-sourced spatial data in general because of the picture other professionals painted regarding the data quality of such sources. Most of these*

recommendations were from people I looked up to. There were numerous issues of trust, inaccuracy, and reliability.

- Said a Senior Systems Developer.

These findings echo the observations expressed by Dasgupta, (2012), where VGI is portrayed as very disorganised, inefficient, and unreliable spatial data by the authoritative experts. About a decade ago, when most participants launched their careers in GIS, VGI was heavily alienated for its quality (Cooper et al., 2012).

This finding demonstrates that users who had not been exposed to any information about the lack of quality of VGI went into the practice with a more positive and inquisitive attitude towards the probable quality of VGI. This finding was discovered in narratives of expert users exposed to environments where VGI was already in use and had registered successes. For them, little to no encounters with negative stories about VGI's lack of quality positively shaped their perception of VGI quality as they began practicing. The contrast between the themes of negativity and positivity demonstrates that things that are perceived as realities are a product of experiences.

Another theme discovered from the narrative about user perception before VGI exposure was that of curiosity. The study's findings have shown that curiosity emerged from the narrative of both participants who had been subjected to negative reviews of VGI quality in the early stages of their practice and those that were completely void of VGI quality reputation. The participants who had been subjected to negative reviews; the huge discouragement attached to VGI quality turned into a pull factor that motivated them to incorporate VGI into their GIS practice. For participants who were completely unaware of VGI's quality reputation, the entire idea of crowdsourced and free spatial data was a motivation to test VGI. All participants were curious to experience VGI and appreciate VGI quality first hand.

When it comes to technology there is always something that attracts people. The pros and the cons, either way. Only when you have explored and experienced it is when you can give an approval or not. My first experience with VGI was a "let's see how this goes" kind of approach. The negativity around it attracted me more.

- Said a GIS and Data Management expert.

This discovery from the study demonstrates why VGI practice has continued to gain momentum besides the doubt that surrounds the status of its quality. Regardless of the reputation of VGI, users continue to explore VGI's capabilities in becoming an alternative to authoritative spatial data sources. This echoes the observations expressed by Ferster et al (2018) where the progressive account of the growth of VGI practice is captured.

4.1.2 User perception of VGI quality after exposure

The second narrative block that was directly linked to the perceptions of participants on VGI quality was "post-exposure VGI quality observation". From this narrative block, two themes were discovered. These included amazement and satisfaction.

While it is expected that the narrative on VGI quality be driven by practitioners who have used VGI as a source of spatial data, this study discovered that the negative narrative about VGI quality is mostly driven by commercial GIS contributors and practitioners who had not used VGI themselves. The study participants who had experience negative reviews about VGI quality were of the view that there existed a battle for relevance in the spatial data markets between commercial spatial data collectors and VGI practitioners. Since VGI is a competing source of spatial data, the study participants perceived commercial practitioners as biased in assessment of VGI quality as the emergence of VGI threatens the market share and the long-term relevance of commercial spatial data sources. The participants in this study showed that the bad reputation of VGI was mostly driven by hearsay within the GIS community and less from practical experience.

> *It was very surprising to learn that the one who discouraged me in adoption of VGI had not used VGI before. Quite strange!*
>
> - Said a GIS and Data Management expert.

These findings provide insight into the relationship between VGI quality and user experience and why various authors have argued for user-based and computational approaches to VGI quality assessment as compared to a producer's perspective. Within the GIS landscape, authoritative spatial data producers are believed to have an upper hand in driving the narrative about the quality of various spatial data sources (Fogliaroni et al., 2018; See et al., 2016). They are alleged to place their interests first (Crowe, 2017) regardless of the fact that there is no perfect GIS data (Pascual, 2011). It is thus believed that user-centred approaches to spatial data quality assessment must be embraced to eliminate such likely biases (Tusker et al., 2018; Yan et al., 2020) **.**

While the emerging themes had their emphasis drawn to the difference between the participants' circumstances that shaped their perception of VGI quality at the beginning of their VGI practice, the findings on the narrative about VGI quality observations post-exposure converged towards the theme of satisfaction. All the participants in the study expressed approval of the quality of VGI encountered within their period of practice.

> *It was interesting when I made my first desktop application using OSM data, the results were great and beyond my quality expectations.*
>
> - Said Senior Systems Developer.

> *VGI has more than proved its worth in many circumstances. In my view, I see that VGI has done a lot of good than damage.*
>
> - Said a GIS Consultant and Senior Lecturer

> *Overall, I can say that I never looked back from the day I got introduced to VGI such as OSM. VGI delivers.*
>
> - Said a Senior Cartography officer.

These findings which portray a shift in user perceptions before and after a user's exposure to VGI validates the belief that VGI suffers from enormous presumed notions of lack of quality as argued by various authors (Fonte et al., 2015; Genovese & Roche, 2010; Greenfeld, 2013; Young et al., 2020). In recent years, numerous authors have argued against this presumed lack of quality in VGI. Tusker et al. (2018) argued that the quality of VGI should not just be based on the properties producers called contributors in VGI but that the data must be tested to validate arguments about its inaccuracy and unrealiability (Tusker et al., 2018). Another argument by Fogliaroni et al. (2018) emphasises the need for computed trustworthiness as a valid approximation of VGI quality (Fogliaroni et al., 2018; Lee et al., 2006).

Overall, the findings from these key narrative blocks and themes demonstrate a shift in perception on the quality of VGI by the participants. Starting with a negative perception when they had not used VGI to being satisfied with the quality of VGI after use, it can thus be concluded that their present take on VGI quality in Malawi has been shaped by experience.

4.2 Key data quality dimension of VGI as considered by expert users.

Table 4.2 shows the comparison matrix findings which formed part of the AHP analysis for the key data quality dimensions evaluated in the study.

Table 4.2: Comparison Matrix of data quality dimensions

Comparison Matrix adopted from SuperDecision software. (+ve digits in favour of roll items, -ve digits in favour of column items)					
Inconsistency	Validity	Timeliness	Completeness	Consistency	Uniqueness
Accuracy	3	4.51	6.12	7.22	9
Validity		1.93	3	8.14	5.72
Timeliness			2.29	4.16	3.94
Completeness				1	3.5
Consistency					1

On the relative importance comparison of completeness and consistency, the value 1 shows that the participants believed that completeness and consistency had equal importance. A similar observation was also made on the comparison of uniqueness and consistency whose comparison shows uniqueness being equal in strength as shown by the value 1. The findings further demonstrate that accuracy was considered to be 3 times more important than validity, 4.51 times more important than timeliness, 6.12 times more important than completeness, 7.22 times more important than consistency and 9 times more important than uniqueness. This finding shows that accuracy dominated all the five dimensions on the relative importance comparison scale followed by validity which was 1.93 times more important than timeliness, 3 times more important than completeness, 8.14 times more important than consistency and 5.72 times more important than uniqueness. This finding shows that validity was the second most important criteria for spatial data quality after accuracy. The comparison matrix was further translated into weighted rankings to have obtain overall rankings of the six criteria as perceived by the VGI expert users in the study. The weights are summarised in Table 4.3.

Table 4.3: Calculated weights of criteria from VGI expert users' perception

Criteria	Weights
Accuracy	0.46794
Validity	0.23501
Timeliness	0.13962
Completeness	0.07463
Consistency	0.04557
Uniqueness	0.03724

Further to the weighted ranks for the quality dimensions in Table 4.3, the study found a consistency ratio of 0.03963 which was used to validate the correctness of the AHP model. Since the consistency ratio was less than 0.1, it demonstrated that the opinions of the participants were consistent, thus the AHP model passed the consistency test and was accepted.

4.2.1 Accuracy: The dominant dimension

The results from the AHP analysis show that the participants prioritised accuracy over the other five key data quality dimensions. This finding is similar to discoveries from other studies (Black & Nederpelt, 2020; Maulia, 2018; Spencer & Wilkes, 2019; Veregin, 1999) where accuracy has emerged as a relatively significant dimension of data quality. Authoritative Data management bodies such as the EDM Council and DAMA recognise accuracy as the leading dimension of data quality (EDM Council, 2021). Goodchild and Li (2012) also identified accuracy as a dominant measure of spatial data quality, which can be broken down into subcategories of position or thematic accuracy (Goodchild & Li, 2012).

Various studies (Ballatore & Zipf, 2015; Cooper et al., 2012; Fan, 2015; Haklay et al., 2014; Pascual, 2011; United Nations Economic Commission for Africa, 2017) on spatial data quality and VGI quality in the last decade have also recognized accuracy as a critical dimension of spatial data quality. It is also recognized as the most cited data quality dimension in a review of scholarly literature spanning two decades from 1995 to 2015 (Mahanti, 2018). The AHP analysis in this study produced similar observations among VGI expert users in Malawi confirming the relevance of previously established trends (Ballatore & Zipf, 2015).

In follow-up interviews that shared the AHP results with the participants, all participants described accuracy as the lifeline of data quality. The participants pointed out that the other lowly ranked quality dimensions, such as consistency and uniqueness can be resolved by applying various operations such as merging and filtering of the data based on the skills of data custodians and users, while accuracy cannot be negotiated.

4.3 Quality performance of VGI against authoritative spatial data

4.3.1 The Datasets

A horizontal accuracy test to assess the quality of VGI against authoritative data deemed to be of high accuracy was conducted. As discussed in chapter 3, the test was conducted on the OSM education facilities dataset for Malawi. The dataset included different themes of education facilities, including kindergarten, primary schools, secondary schools, and higher learning institutions. Part three of the Spatial Positioning Accuracy Standards of the Federal Geographic Data Committee (FGDC) the FGDC-NSSDA-STD-007.3-1998 (Federal Geographic Data Committee, 1998, p. 5) recommends reporting separate accuracies for composite datasets with multiple themes and geographic areas that may contain different accuracies. Since the OSM Education facilities dataset for Malawi contained multiple themes for the education facilities, the NSSDA provided room for a separate test of any of the themes under education facilities. The results presented in this section are for the quality test on primary school facilities.

The OSM dataset for Education facilities in Malawi was found to have metadata that included information such as the source, date for the dataset's last update, the expected update frequency and methodology of collection. The full metadata is shown in Figure 4.1.

Source	OpenStreetMap contributors
Contributor	Humanitarian OpenStreetMap Team (HOT)
Date of Dataset	October 06, 2022-October 06, 2022
Updated	6 October 2022
Expected Update Frequency	Every month
Location	Malawi
Visibility	Public
License	Open Database License (ODC-ODbL)
Methodology	Volunteered geographic information
Caveats / Comments	OpenStreetMap data is crowd sourced and cannot be considered to be exhaustive
Tags	EDUCATION EDUCATION FACILITIES - SCHOOLS GEODATA
File Format	SHP GEOPACKAGE KML GARMIN IMG

Figure 4.1: Metadata for OSM dataset for Education facilities in Malawi

Figure 4.1, shows that the dataset did not have any metadata describing the status of its quality. This highlights one of challenges that is faced by users when accessing published spatial data from various platforms in Malawi. Without such facts, it is difficult for users to easily determine the extent in which a dataset satisfies their quality needs (Mahanti, 2018). Similar observations are made in the article *Spatial Metadata in Africa and The Middle East* (Cooper & Gavin, 2005), metadata on spatial data quality helps users in identifying acceptable accuracies for their field of application (Federal Geographic Data Committee, 1998).

To test VGI for quality using the NSSDA standard and its associated metrics an authoritative dataset deemed of higher quality had to be identified. The choice of primary school theme on the OSM dataset for Education facilities led to the selection of a Malawi Ministry of Education Primary (MoE) Schools dataset from MASDAP. This dataset represented an equal sphere and themed authoritative dataset of higher quality. Figure 4.2 shows the published metadata for the MoE Primary Schools dataset.

Title	Primary Schools
Abstract	Location of Primary Schools (2013)
Owner	Justin Saunders
Created	November 20th 2019
Published	January 10th 2017
Last Modified	November 5th 2022
Resource Type	dataset
Category	location
Keywords	features primary_schools_ll
Regions	Malawi
Point of Contact	Justin Saunders
Language	eng
Supplemental Information	No information provided

Figure 4.2: Metadata for Ministry of Education Primary Schools dataset

Figure 4.2 does not show metadata on quality of the MoE dataset. The commonly applied techniques of spatial data quality assessment quoted in the study capitalise on the principle of trust that comes with authoritative spatial sources. Regrettably, VGI sources lack this type of trust hence the missing metadata on quality becomes an issue while the same treatment hardly applies to authoritative data sets.

The first step in the horizontal accuracy tests was to identify identical points whose coordinates would be used for the quality test. The two datasets were examined to establish points representing the same feature on the ground. According to the FGDC-NSSDA-STD-007.3-1998, a horizontal accuracy test should be conducted on planimetric coordinates of well-defined points with corresponding coordinates of the same points from an individual/authoritative dataset deemed to be of high accuracy (Federal Geographic Data Committee, 1998). In order to achieve this goal, the paired datasets were loaded in the open-source GIS software QGIS for visual inspection. The various point features were then closely inspected using a zoom tool in QGIS to establish the proximity of the featured points in relation to their registered names in the datasets. From this a systematic selection of the points guided by the study's methodology. Figure 4.3 and Figure 4.4 show overlays of the two datasets.

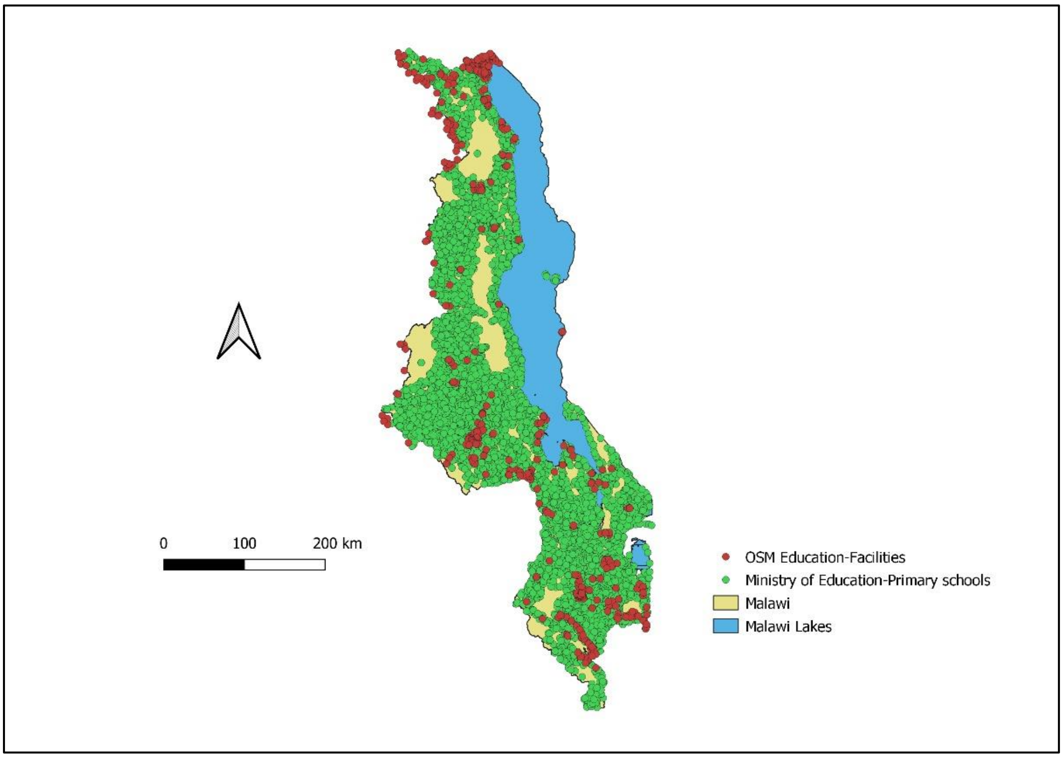

Figure 4.3: Map showing OSM and MoE primary schools' datasets

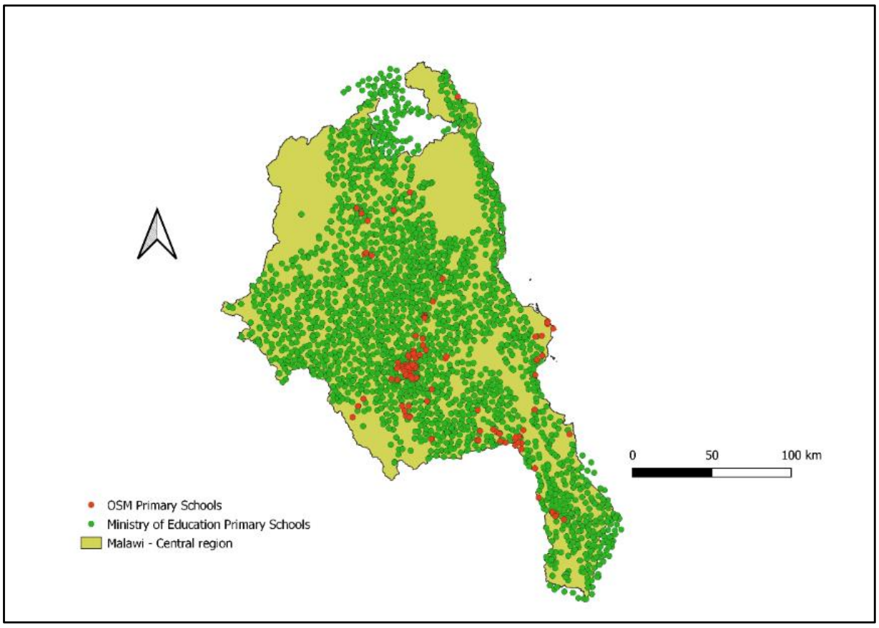

Figure 4.4: Map extract showing differences in concentration of captured schools between OSM and MoE primary schools' datasets in central Malawi.

Figure 4.3 and Figure 4.4 reveal that more school locations were recorded in the MoE dataset compared to the OSM one. The MoE dataset attributes table revealed 5296 entries against 767 in OSM. While these statistics were for unprocessed raw datasets, which may include duplicate features and unnamed features, this discovery brought to light issues of incompleteness in VGI datasets and that more work was to be done in improving the coverage of the VGI in the education sector. In addition, visualisation of the OSM dataset showed the dataset included primary schools from neighbouring countries. This very evident in the boarder districts of the northern part of Malawi. Randomly selected primary schools from the sample were also visualised to see how the points looked side by side on a map. This is demonstrated in Figure 4.5

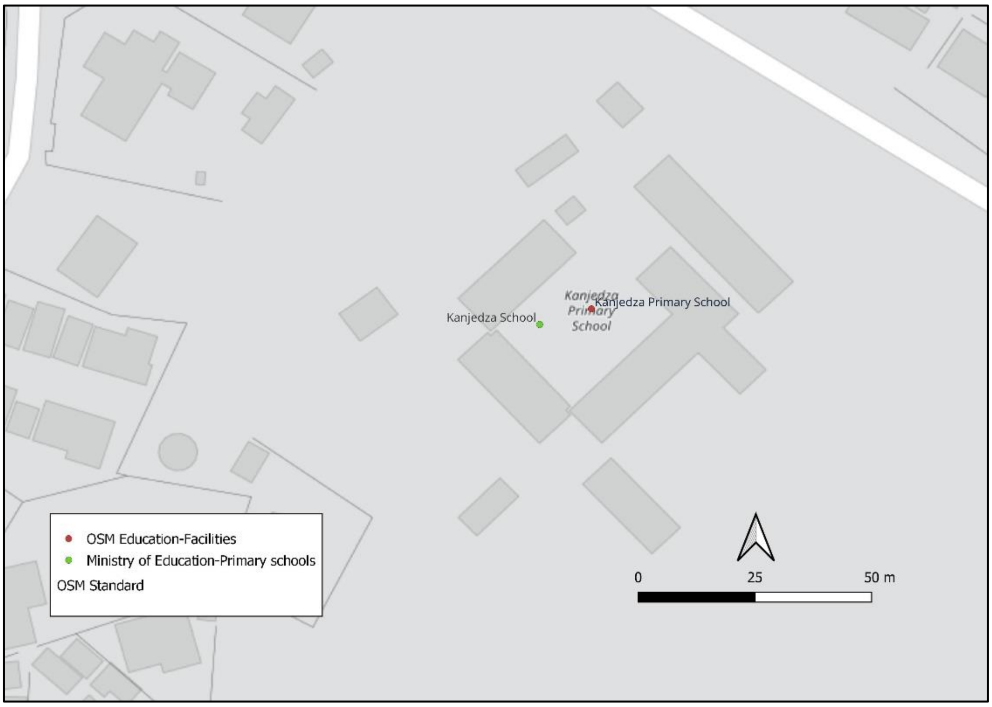

Figure 4.5: Map extract showing two points representing Kanjeza Primary School in Blantyre.

4.3.2 NSSDA - Horizontal Accuracy

The results in Table 4.4 show that RMSE for latitudes (RMSE$_{X}$) was not equal to RMSE for longitudes (RMSE$_{Y}$. FGDC-NSSDA-STD-007.3-1998 standard states that where RMSE$_{X}$ is not equal to RMSE$_{Y}$, accuracy must be reported using the CSE at 39.35% confidence level and the NSSDA accuracy at 95% confidence level in ground distances (Federal Geographic Data Committee, 1998, p. 11). The NSSDA methodology of accuracy test also recommends that test results be reported in ground distances.

Table 4.4: Horizontal Accuracy test results: OSM vs MoE Primary school RMSE

Parameter	Degree Minutes	Metres
$RMSE_X$ (Latitude)	0.00014	15.38558
$RMSE_Y$ (Longitude)	0.00012	23.03014
RMSE (combined)	0.00025	27.69663
Circular Standard Error (CSE)	0.00017	19.20786
NSSDA (ACC_r) (OSM Horizontal Accuracy)	0.00042	47.01508

The results in Table 4.4 show that RMSE for latitudes ($RMSE_{X)}$ was not equal to RMSE for longitudes ($RMSE_Y$. FGDC-NSSDA-STD-007.3-1998 standard states that where $RMSE_X$ is not equal to $RMSE_Y$, accuracy must be reported using the CSE at 39.35% confidence level and the NSSDA accuracy at 95% confidence level in ground distances (Federal Geographic Data Committee, 1998, p. 11). Table 4.4 shows a reported CSE accuracy of 19.2078 metres at 39.35% confidence level and 47.0150 metres for NSSDA at 95% confidence level. Under the NSSDA this result meant that for every 20 points under the primary school theme in the OSM dataset, 19 points were to have an error equal to or lower than 47.0150 metres while one point was allowed to be outside the reported accuracy. Further evaluation of the sample points showed that only one point of the fifty had a location error higher than the reported NSSDA.

Further tests on an additional 450 points from the OSM dataset showed that only 17 points had an error higher than the reported NSSDA. This demonstrated that the OSM dataset passed the test in relation to the NSSDA standard (Federal Geographic Data Committee, 1998). Both the Circular Standard Error of 19.2078 metres and NSSDA of 47.0150 metres were found to be acceptable for various applications of GIS such as ground feature identification, travel, and resource planning (Federal Geographic Data Committee, 1998). According to the (Malawi Government- MEST, 2019) establishment of a primary school in Malawi requires a minimum 1 and 1.5 hectares of land for an 8-classroom and 16-classroom school respectively. With this being the minimum standard size of primary schools in Malawi, it was expected that each point representing a school in the VGI dataset would be within the boundaries of the school premises as long as it was within the reported NSSDA. This was confirmed in the visualisation of the points per Figure 4.6 which showed that the points representing

Blantyre Girls Primary school in each of the datasets were within the school campus. This showed to the correctness of the VGI dataset.

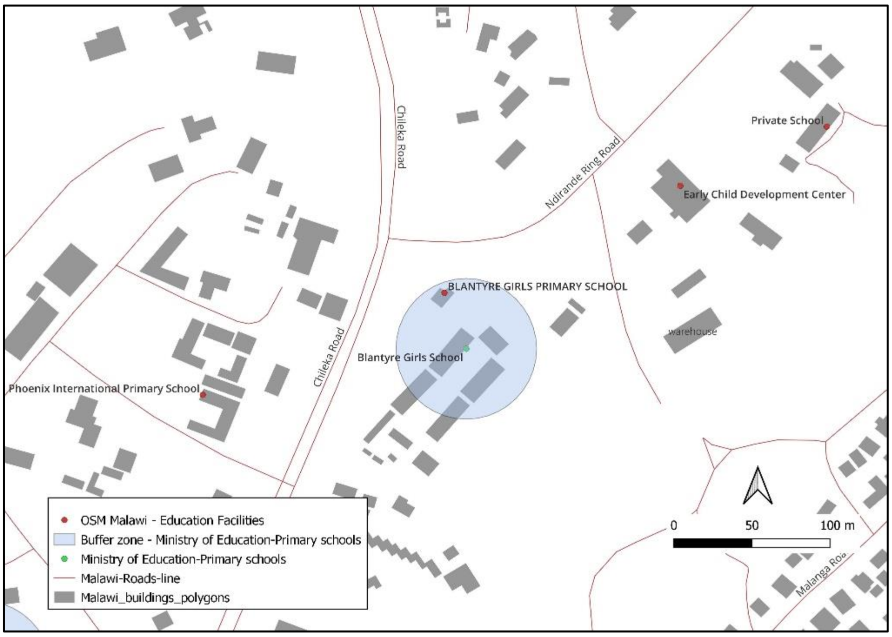

Figure 4.6: Map extract showing points for Blantyre Girls Primary School in Blantyre.

Based on the requirements of the NSSDA accuracy reporting (Federal Geographic Data Committee, 1998, p. 5) , The accuracy for the OSM is reported as:

Dataset: OSM Education Facilities for Malawi
Featured theme: Primary Schools

"Tested **19.2078** metres Circular Standard Error at 39.35% confidence level".

"Tested **47.0150** metres horizontal accuracy at 95% confidence level"

4.3.3 NSSDA validation

As a way of corroborating the OSM dataset's horizontal accuracy test results, the study partially applied Horizontal Positional Error (HPE) using Euclidean distance buffers to visualise the reported accuracy of 47.0150 metres. Figure 4.7 and Figure 4.8 show pairs of points for the same primary schools from both the VGI and authoritative datasets overlayed with a circular buffer of 47.0150 metres from the reported NSSDA. The buffering was done on the MoE dataset points as the dataset of higher accuracy. Figure

4.7 shows Karonga Demo primary school whose point (red) of the VGI dataset was visually within the circular buffer demonstrating a lower error than the NSDDA.

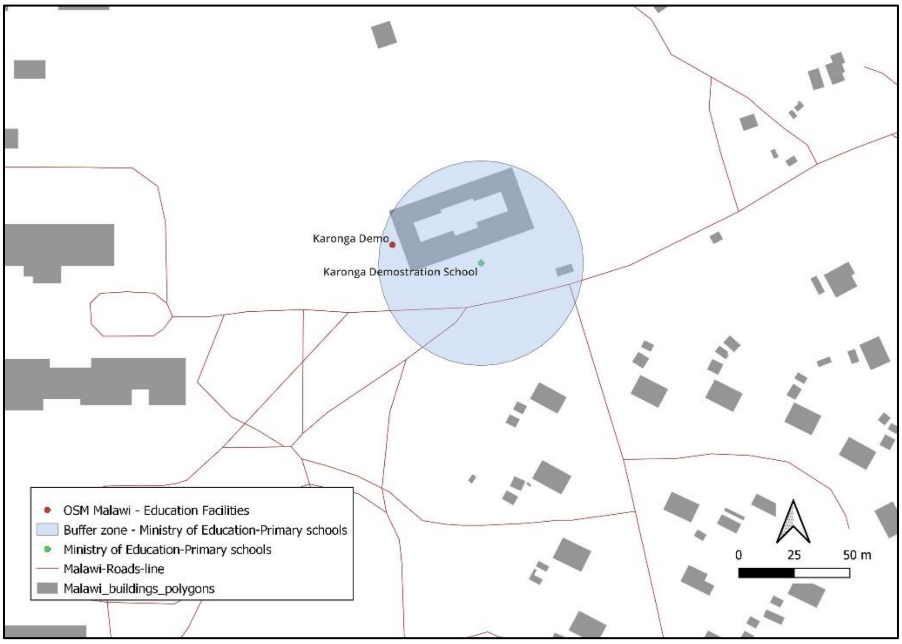

Figure 4.7: Map extract showing Euclidean circular buffer of Karonga Demonstration Primary School in Karonga District.

Figure 4.8 shows the points representing Mponda Primary School, the only VGI point that fell outside the NSSDA distance of 47.0150 metres among the 50 sampled points, demonstrating that this point had an error higher than the reported NSSDA.

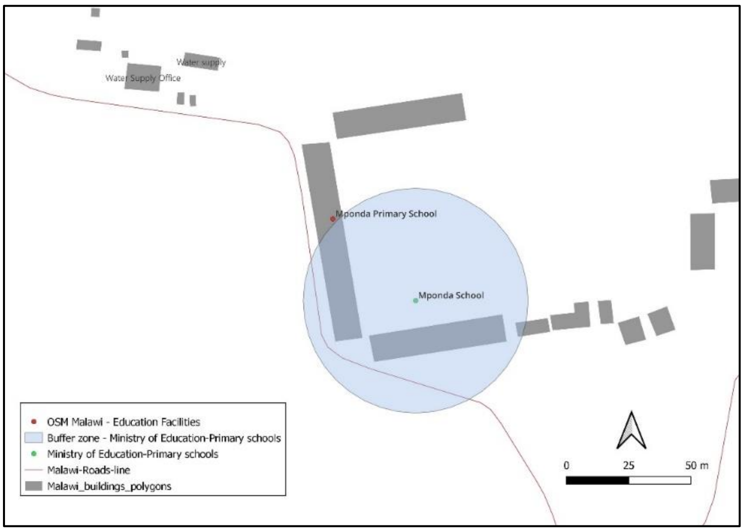

Figure 4.8: A visualisation of Mponda Primary School in Balaka district with the OSM point falling outside the 47.0150 metres buffer zone.

Further validation efforts using the measuring tool in QGIS found a measured cartesian distance between the OSM and MoE dataset points for Mponda Primary school in Balaka district, Southern Malawi to be 48.3474 metres. This was the only set of points with an error higher that the calculated NSSDA distance among the 50 sampled pairs of points.

Due to a lack of metadata describing the accuracy of the MoE dataset, the study could not report the accuracy of MoE dataset beyond its authoritative properties. In addition, it was beyond the scope of this study to establish the methods by which the two datasets were collected. Diggelen & Enge (2015) and Senaratne et al. (2016) suggest that the methods of spatial data collection which include the type of gadget, technology used and ground or desk data collection have a huge bearing on the quality of spatial data. However, this study focused on skills and experience which are the key differentiating factors between volunteer and authoritative contributors of spatial data.

4.4 Chapter Summary

This chapter presented the study's findings and a discussion based on those findings. The findings presented included the narrative blocks and themes emerging from the narrative analysis, the AHP analysis results and the outcome of the NSSDA test for horizontal accuracy. All these results were discussed in line with the research questions, and the corresponding objectives to the research questions were addressed.

CHAPTER 5

CONCLUSION AND RECOMMENDATION

This study aimed at ascertaining the status of VGI quality in Malawi by analysing expert user experiences and conducting methodical quality test of VGI. Resource constraints and time limitations in various GIS projects call for exploration of financially viable ways of acquiring spatial data. VGI therefore fits the profile for a source of spatial data that can address such challenges. As VGI continues to suffer from a presumed lack of quality across the globe, it has become imperative that the status of the quality of VGI be assessed to inform its usability in Malawi and elsewhere.

While various researchers have attempted to address the issue of VGI quality from a producer's perspective by examining the properties of VGI contributors, other researchers have criticised this approach and advocate for user-based spatial data quality assessment methodologies. In VGI quality management, application of quality standards such as NSSDA, CSE, and HPE can greatly help filter out inaccurate data within prescribed boundaries of accuracy for both the custodian and the end-user. This study applied a blended approach to spatial quality assessment by focusing on the areas of user perceptions and interpretation of VGI quality from their experiences and computational testing of VGI quality against authoritative spatial data. Findings from this study revealed that in Malawi, VGI suffers from presumed notions of lack of quality matching the global trends. This study has found VGI to be within tolerable levels of accuracy when compared against authoritative spatial data quality. As spatial data quality is a multi-dimensional concept, it has also been observed that expert users of VGI in Malawi prioritise accuracy over other quality dimensions despite examined VGI showing levels of incompleteness.

In order to strengthen the development of VGI practice, this study has suggested two strategies: by testing and publishing metadata on VGI quality and improving completeness of VGI datasets, the uptake of VGI in Malawi can be improved and the narrative about VGI quality be turned about with VGI certified a reliable source of spatial data. The involvement of the citizenry in the collection of spatial data remains crucial in fostering and accelerating the development agenda in Malawi where economic development is driven by spatial intelligence. The use of available free and quality spatial data not only saves time in project lifecycles but also saves finances that would have rather been used in the collection or purchasing of this data.

This study's findings from computational VGI quality tests are deemed acceptable for spatial applications that accept a maximum horizontal accuracy of approximately 50 metres. As the study only focused on the status of VGI quality in Malawi with emphasis on user perceptions and accuracy as a dimension of spatial data quality, further research may be required to address other dimensions of quality, such as completeness, whose prevalence has been noticed in the study.

Chapter Summary

This chapter presented a summary of the study. It also outlined the recommendations based on the findings and discussions presented in the study. Part of the presentation also included a conclusion to the study, expected contributions and areas of future research.

REFERENCES

Adler, E. S., & Clark, R. (2010). *An Invitation to Social Research: How It's Done.* Massachusetts: Cengage Learning.

Antwi, S. K., & Hamza, K. (2015). Qualitative and Quantitative Research Paradigms in Business Research: A Philosophical Reflection. *European Journal of Business and Management*, 217-225.

Appel, S. (2019). Geospatial Information Literacy Instruction: Frameworks, Competency and Threshold Concepts. *Journal of Map and Geography Libraries*, 134-151.

Ballatore, A., & Zipf, A. (2015). A Conceptual Quality Framework for Volunteered Geographic Information. *Conference on Spatial Information Theory XII* (pp. 90 -107). Santa Fee: UC Santa Barbara.

Bhattacherjee, A. (2012). *Social Science Research: Principles, Methods, and Practices.* Florida: Global Text Project.

Bielecka, E., & Burek, E. (2019). Spatial Data Quality And Uncertainty Publication Patterns And Trends By Bibliometric Analysis. *Open Geosciences*, 219-235.

Black, A., & Nederpelt, P. v. (2020). *Dimensions of Data Quality (DDQ).* Overbetuwe: DAMA NL Foundation.

Brown, A. W., Kaiser, K. A., & Allison, D. B. (2018). Issues with data and analyses: Errors, underlying themes, and potential solutions. *Proceedings of the National Academy of Sciences* (pp. 2561-2562.). Washington, D.C: National Academy of Sciences.

Bryman, A., & Bell, E. (2003). *Business research methods.* New York: Oxford University Press.

Chikumba. (2019). Geodata Maintenance and Collaboration in GIS Implementation in Health Sector in a Developing Country Context: The Case of DHIS2 GIS in Malawi. *Africa GIS 2019.* Kigali: Africa GIS Conference.

Chikumba, P. A., & Chisakasa, G. (2018). Towards Geodata Maintenance: A Case of DHIS2 GIS Implementation in Malawi . *Journal of Health Informatics in Developing Countries*.

Cochran, W. G. (1977). *Sampling Techniques.* John Wiley & Sons.

Cooper, A., & Gavin, E. J. (2005). Spatial Metadata in Africa and The Middle East. In A. H. Harold Moellering, & A. Crane, *World Spatial Metadata Standards* (pp. 123-139). Amsterdam: Elsevier Ltd.

Cooper, A., Coetzee, S., Kourie, D., Kaczmarek, I., Iwaniak, A., & Kubik, T. (2012). *Volunteered Geographical Information - The Challenges.* Cape town: UCT.

Coote, A., & Rackham, L. (2008). Neogeographic data quality – is it an issue? *AGI Conference* . ConsultingWhere Ltd.

Coughlin, T. (2018, November 27). *175 Zettabytes By 2025*. Retrieved from Forbes website: https://www.forbes.com/sites/tomcoughlin/2018/11/27/175-zettabytes-by-2025/?sh=20612e825459

Creswell, D. J. (2009). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches.* Los Angels: SAGE Publications.

Creswell, J. W., & Creswell, J. D. (2018). *Research Design: qualitative, quantitative, and mixed methods approaches.* California: Sage.

Crowe, J. (2017, October 30). *Bad Internet Maps: 'A Social Media Plague'*. Retrieved from The Map Room website: https://www.maproomblog.com/2017/10/bad-internet-maps-a-social-media-plague/

Dasgupta, A. (2012, July 24). *VGI: Democratisation of Geographic Information*. Retrieved from geospatialworld website: https://www.geospatialworld.net/article/vgi-democratisation-of-geographic-information/

Desjardins, J. (2019, April 17). *How much data is generated each day?* . Retrieved from World Economic Forum Website: https://www.weforum.org/agenda/2019/04/how-much-data-is-generated-each-day-cf4bddf29f/

Diggelen, F. v., & Enge, P. (2015). The World's first GPS MOOC and Worldwide Laboratory using Smartphones. *Proceedings of the 28th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2015)* (pp. 361 - 369). Tampa, Florida: Institute of Navigation.

Djuraskovic, O. (2022, January 10). *Big Data Statistics 2022: How Much Data is in The World?* Retrieved from firstsiteguide website: https://firstsiteguide.com/big-data-stats/

Dorn, H., Tornros, T., & Zipf, A. (2015). Quality Evaluation of VGI Using Authoritative Data—A Comparison with Land Use Data in Southern Germany. *ISPRS International Journal of Geo-Information*, 1657-1671.

Du, H., Alechina, N., Jackson, M., & Hart, G. (2016). A Method for Matching Crowd-sourced and Authoritative Geospatial Data. *Transactions in GIS*.

EDM Council. (2021, April 23). *Data Quality Dimensions.* Retrieved from EDM Council website: https://cdn.ymaws.com/edmcouncil.org/resource/collection/28B06485-3F13-498E-9856-FABC4CA29FAF/DCAM_DQ_Dimensions.pdf

Edmonds, W. A., & Kennedy, T. D. (2017). *Exploratory-sequential approach.* California: SAGE Publications.

ESRI. (2023, 03 31). *Location Intelligence: Insights that transform decision-making*. Retrieved from ESRI website: https://www.esri.com/en-us/location-intelligence/overview

Essoungou, A.-M. (2011, August 15). *Africa's least developed: lands of opportunity*. Retrieved from www.un.org/africarenewal: https://www.un.org/africarenewal/magazine/august-2011/africas-least-developed-lands-opportunity

Fan, W. (2015). Data Quality: From Theory to Practice. *SIGMOD Rec*, 7–18.

Federal Geographic Data Committee. (1998). *Geospatial Positioning Accuracy Standards, Part 3: National Standard for Spatial Data.* Virginia: Federal Geographic Data Committee. Retrieved from fgdc government website: https://www.fgdc.gov/standards/projects/FGDC-standards-projects/accuracy/part3/chapter3

Ferster, C., Nelson, T., Robertson, C., & Feick, R. (2018). Current Themes in Volunteered Geographic Information. *Comprehensive Geographic Information Systems journal*, 26-41.

Fogliaroni, P., D’Antonio, F., & Clementini, E. (2018). Data Trustworthiness and User Reputation as Indicators of VGI Quality. *Geo-spatial Information Science*, 213-233.

Fonte, C., Bastin, L., Foody, G., Kellenberger, T., Kerle, N., Mooney, P., . . . See, L. (2015). VGI Quality Control. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 317 - 324.

Forbes. (2018, May 21). *How Much Data Do We Create Every Day: The Mind-Blowing Stats Everyone Should Read*. Retrieved from Forbes website: https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/?sh=19e1f0d060ba

Gardner, Z., & Mooney, P. (2018). Investigating gender differences in OpenStreetMap activities in Malawi: a small case-study. *AGILE 2018.* Lund: AGILE 2018.

Gemma, R. (2018). Introduction to Positivism, Interpretivism and Critical Theory. *Nurse Researcher*, 14-20.

Genovese, E., & Roche, S. (2010). Potential of VGI as a Resource for SDIs in the North/South Context. *Geomatica*, 439-450.

Global Facility for Disaster Reduction and Recovery. (2020). *Open Cities Africa.* Washington, D.C.: GFDRR.

Gokhale, M. (2022, October 19). *Use of analytical hierarchy process in university strategy planning.* Retrieved from scholarsmine website - Missouri S&T Library and Learning Resources: https://scholarsmine.mst.edu/masters_theses/4608

Goodchild, M. (2007). Citizens as Sensors: The World of Volunteered Geography. *GeoJournal* , 211–221.

Goodchild, M. F., & Li, L. (2012). Assuring the Quality of Volunteered Geographical Information. *The Journal of Spatial Statistics*, 110-120.

Greenfeld, J. (2013). *GIS – Beyond just a pretty map.* New Jersey: New Jersey Institute of Technology.

Haklay, M., Antoniou, V., Basiouka, S., Soden, R., & Mooney, P. (2014). *Crowdsourced Geographic Information use in Government, Report to GFDRR (World Bank).* London: World Bank.

Harris, J. (2011, August 22). *The General Theory of Data Quality*. Retrieved from OCDQ website: http://www.ocdqblog.com/home/the-general-theory-of-data-quality.html

Heuvelink, G. (1999). Propagation of Error in Spatial Modelling with GIS. In P. A. Longley, M. F. Goodchild, D. J. Maguire, & D. W. Rhind, *Geographic Information Systems* (pp. 207 - 217). New York: John Wiley & Sons, Inc.

Hillier, W. (2021, July 13). *What is data quality and why is it important?* Retrieved from careerfoundry.com: https://careerfoundry.com/en/blog/data-analytics/what-is-data-quality/

HOT. (2021, April 20). *HOT Malawi*. Retrieved from hotosm organisation website: https://www.hotosm.org/where-we-work/malawi/

Hunter, G. J., Jones, S., Bregt, A., & Masters, E. (2003). Spatial Data Quality. *Advanced Geographical Information Systems*.

ISO. (2019, July 01). *ISO 19157:2013*. Retrieved from iso: https://www.iso.org/standard/32575.html

J.Ferster, C., Nelson, T., ColinRobertson, & Feick, R. (2018). Current Themes in Volunteered Geographic Information. *Comprehensive Geographic Information Systems*, 26-41.

Johnson, P. A., Siber, R., Scassa, T., Stephens, M., & Robinson, P. (2017). The Cost(s) of Geospatial Open Data . . *Transactions in GIS*, 434-445.

Juran, J. M., & DeFeo, J. A. (2010). *Juran's Quality Handbook: The Complete Guide to Performance Excellence 6th Edition.* New York City: McGraw Hill Professional.

Juran, J. M., & Godfrey, B. (1951, October 15). *Juran's Quality Handbook.* McGraw Hill. Retrieved from Study.com: https://study.com/academy/lesson/joseph-jurans-theory-definition-of-quality.html.

Kumar, D. S., Kumar, D. R., Govindaraj, D. M., & Prabhu, D. N. (2020). Sampling Framework for Personal Interviews in Research. *Palarch's Journal Of Archaeology Of Egypt/Egyptology, 17*(7), 7102-7114.

Leavy, P. (2017). *Research Design: Quantitative, Qualitative, Mixed Methods, Arts-Based, and Community-Based Participatory Research Approaches* . New York: The Guilford Press.

Lee, Y. W., Pipino, L. L., Funk, J. D., & Wang, R. Y. (2006). *Journey to Data Quality.* London: MIT Press.

Levy, P. (2008). *Sampling of Populations: Methods and Applications.* New Jersey: John Wiley & Sons, INC.

Linz. (2022, August). *Geospatial data types*. Retrieved from Linz government website: https://www.linz.govt.nz/guidance/data-service/linz-data-service-guide/data-and-schemas/geospatial-data-types

Lohr, S. (2010). *Sampling: Design and Analysis.* Boston: Brooks/Cole.

Lohr, S. L. (2019). *Sampling Design and Analysis 2nd edition.* New York: CRC Press.

Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (1999). *Geographical Information Systems.* New York: John Wiley & Sons, Inc.

Lotame. (2019, April 30). *Why Is Data Quality Important?* Retrieved from Lotame website: https://www.lotame.com/why-is-data-quality-important

Mahabir, R., Stefanidis, A., Croitoru, A., Crooks, A. T., & Agouris, P. (2017). Authoritative and Volunteered Geographical Information in a Developing Country: A Comparative Case Study of Road Datasets in Nairobi, Kenya. *International Journal of Geo-Information*.

Mahanti, R. (2018). *Data Quality: Dimensions, Measurement, Strategy, Management, and Governance.* Milwaukee: ASQ Quality Press.

Malawi Government- MEST. (2019). *Safer Schools Construction Guideline.* Malawi Government.

Marczyk, G., DeMatteo, D., & Festinger, D. (2005). *Essentials of Research Design and Methodology.* New Jersey: John Wiley & Sons.

Marr, B. (2020). *How Much Data Is There In the World?* Retrieved from Bernard Marr website: https://www.bernardmarr.com/default.asp?contentID=1846

Maulia, N. (2018). *Development of an update procedure for authoritative spatial data by the combination with crowdsourced information (Masters Thesis).*

Meng, X., & Ci, X. (2013). Big Data Management: Concepts, Technics and Challenges. *Journal of Computer Research and Development*, 146-169.

Mhone, C. (2018, January 9). *'Malawi Mappers' Mobilize to Improve Open Geospatial Data*. Retrieved from opendri organisation website: https://opendri.org/malawi-mappers-cmhone/

Mhone, C. (2021, November 23). *Building Sustainable Mapping Communities [Video]*. Retrieved from Humanitarian OpenStreetMap Team [Youtube channel]: https://www.youtube.com/watch?v=KmfEsZfBnX4

Mooney, P., & Corcoran, P. (2012). Using OSM for LBS – An Analysis of Changes to Attributes of Spatial Objects. In G. Gartner, & G. Ortag, *Advances in Location-Based Services* (pp. 165–179). Berlin: Springer .

Morse, J. M., & Niehaus, L. (2016). *Mixed Method Design: Principles and Procedures.* New York: Routledge.

Muya, D. A. (2017, May 15). *The Case for the Development of a National Spatial Data Infrastructure by African Countries*. Retrieved from africaland.org: https://africaland.org/2017/05/15/the-case-for-the-development-of-a-national-spatial-data-infrastructure-by-african-countries/

Mwungu, C. M. (2017). *Spatial Data Infrastructure in Africa: a technical and institutional analysis.* Nairobi: University of Nairobi.

Nkwae, B., & Nichols, D. S. (2006). *Spatial Data Infrastructures: African Experiences.* Fredericton: University of New Brunswick.

Olaya, V. (2018). *Introduction to GIS.* South Carolina: CreateSpace Independent Publishing Platform.

Pascual, M. S. (2011, November 6). *GIS Data: A Look at Accuracy, Precision, and Types of Errors*. Retrieved from GIS Lounge website: https://www.gislounge.com/gis-data-a-look-at-accuracy-precision-and-types-of-errors/

Pessoa, A., Harper, E., Santos, I., & Gracino, M. (2019). Using Reflexive Interviewing to Foster Deep Understanding of Research Participants' Perspectives. *International Journal of Qualitative Methods, 18*. https://doi.org/https://doi.org/10.1177/1609406918825026

Reichertz, J. (2014). Induction, Deduction,Abduction. In U. Flick, *Qualitative Data Analysis* (pp. 123- 135). London: Sage Publication.

Riessman, C. K. (1993). *Narrative analysis.* Sage Publications.

Roumeliotis, R. (2020, April 28). *The state of data quality: Too much, too wild and too skewed*. Retrieved from helpnetsecurity website: https://www.helpnetsecurity.com/2020/04/28/data-quality-issues/

Saaty, T. (1988). *The Analytic Hierarchy Process.* New York: McGraw-Hill.

Santini Ron. (2021, March). *GIS Errors, Accuracy and Precision.* Retrieved from Pennsylvania State University website: https://www.e-education.psu.edu/geog469/node/253

Santini, R. (2019, March 01). *Energy Industry Applications of GIS*. Retrieved from e-education.psu.edu: https://www.e-education.psu.edu/geog469/node/31

Sarfin, R. L. (2021, May 7). *5 Characteristics of Data Quality*. Retrieved from Precisely website: https://www.precisely.com/blog/data-quality/5-characteristics-of-data-quality

See, L., Mooney, P., Foody, G., Bastin, L., Comber, A., Estima, J., . . . Rutzinger, M. (2016). Crowdsourcing, Citizen Science or Volunteered Geographic Information? The Current State of Crowdsourced Geographic Information. *International Journal of Geo-Information, 5*(5), 55. https://doi.org/10.3390/ijgi5050055

Sekhula, P. P. (2013). The Law Governing the Acquisition and Use of Earth Observation Data in South Africa: A Need for Legal Harmonisation. *South African Journal of Geomatics, 2*(2).

Senaratne, H., Mobasheri, A., Ali, A. L., Capineri, C., & Haklay, M. (. (2016). A Review of Volunteered Geographic Information Quality Assessment Methods. *International Journal of Geographical Information Science, 31*(1), 139-167. https://doi.org/https://doi.org/10.1080/13658816.2016.1189556

Shamoo, A., & Resnik, D. (2015). *Responsible Conduct of Research.* New York: Oxford University Press.

Shin, M. E., Campbell, J., & Burkhart, N. (2018). *Essentials of Geographic Information Systems.* Massachusetts: Flat World Knowledge.

Spencer, J., & Wilkes, B. (2019). *Assessing Spatial Data Quality Using Five Data Anomalies.* North Carolina: MEASURE Evaluation.

Stage, D. (2009). *Authority and Authoritative Data: A Clarification of Terms and Concepts.* IAAO organisation.

Stieglitz, S., Milad, M., Fromm, J., & Melzer, S. (2018). The Adoption of Social Media Analytics for Crisis Management – Challenges and Opportunities. *Twenty-Sixth European Conference on Information Systems(ECIS2018).* Portsmouth: AIS Electronic Library. Retrieved from https://aisel.aisnet.org/ecis2018_rp/4

Thompson, K. (2015, May 18). *Positivism and Interpretivism in Social Research*. Retrieved from Revise Sociology website: https://revisesociology.com/2015/05/18/positivism-interpretivism-sociology/

Tusker, H. B., Eleanor, B., Joshua, W., & Róisín, R. (2018). The Good, the Bad, and the Uncertain: Contributions of Volunteered Geographic Information to Community Disaster Resilience . *Frontiers in Earth Science* , 183.

UNDP. (2020, November 4). *Even the Pandemic Couldn't Slow a Mapathon Down*. Retrieved from UNDP website: https://www.undp.org/malawi/news/even-pandemic-couldn%E2%80%99t-slow-mapathon-down

United Nations Economic Commission for Africa. (2017). *Volunteered Geographic Information in Africa, United Nations Economic Commission for Africa.* Addis Ababa: UNECA.

Veregin, H. (1999). Data Quality Parameters. In P. A. Longley, M. F. Goodchild, D. J. Maguire, & D. W. Rhind, *Geographical Information Systems* (pp. 177-189). New York: John Wiley & Sons, Inc.

Vosoughi, S., Roy, D., & Aral, S. (2018). The Spread of True And False News Online. *Science, 359*(6380), 1146–1151. https://doi.org/https://doi.org/10.1126/science.aap9559

Wassén, O. (2019, August 26). *11 Benefits of Achieving High Data Quality*. Retrieved from Nodegraph website: https://www.nodegraph.se/11-benefits-of-achieving-high-data-quality/

Wisdom, J., & Creswell, J. W. (2013). *Mixed Methods: Integrating Quantitative and Qualitative Data Collection and Analysis While Studying Patient-Centered Medical Home Models.* Rockville: AHRQ Publication.

World Bank. (2014). *Open Data for Resilience Initiative Field Guide.* Washington, DC: World Bank, License: Creative Commons Attribution CC BY 3.0.

World Bank Group. (2022, March). *Population, total - Malawi*. Retrieved from World Bank group website: https://data.worldbank.org/indicator/SP.POP.TOTL?locations=MW

Yan, Y., Feng, C.-C., Huang, W., Fan, H., Wang, Y.-C., & Zipf, A. (2020). Volunteered Geographic Information Research in the First Decade: A Narrative Review of Selected Journal Articles in GIScience. *International Journal of Geographical*

Information Science, 34(9), 1765-1791. https://doi.org/https://doi.org/10.1080/13658816.2020.1730848

Yilma, A. D. (2016). Volunteered Geographic Information in Africa. *ISPRS Geospatial Week* (pp. 1615-1620). Enschede: UNCEA.

Young, J., Lynch, R., Jowaisas, C., Norlander, B., Achampong, S., & Sam, J. (2020). Volunteer Geographic Information in the Global South: Barriers to Local Implementation of Mapping Projects Across Africa. *GeoJournal*, 2227–2243.

Zola, A., & Fontecchio, M. (2021, June 2). *Spatial Data*. Retrieved from Techtarget: https://www.techtarget.com/searchdatamanagement/definition/spatial-data

Žukauskas, P., Vveinhardt, J., & Andriukaitienė, R. (2018, April 18). *Philosophy and Paradigm of Scientific Research, Management Culture and Corporate Social Responsibility*. Retrieved from Intechopen website: https://www.intechopen.com/books/management-culture-and-corporate-social-responsibility/philosophy-and-paradigm-of-scientific-research

APPENDICES

Appendix A: Researcher reference/introductory letter

VICE-CHANCELLOR
Prof. Samson M.I. Sajidu,
BSc Mlw, MPhil Cantab, PhD Mlw

Connect with Excellence

UNIVERSITY OF MALAWI
P.O. Box 280, Zomba, Malawi
TEL: (265) 1 524 222
FAX: (265) 1 524 046
EMAIL: vc@unima.ac.mw

TO WHOM IT MAY CONCERN
9th August, 2022

REFERENCE FOR MR. GIOVANNIE MAKONDI (MSC-INF-14-19)

As per the subject matter, Mr. Giovannie Makondi, registration number MSC-INF-14-19 is a student at the University of Malawi pursuing an MSc in Informatics which is offered jointly between the Computer Science department at University of Malawi (UNIMA) and the CIT department at Malawi University of Business and Applied Sciences (MUBAS).

He seeks to collect data for her research at your organisation. The collected data will only be used for academic purposes. At the end of his study, Mr. Giovannie Makondi will share the results of his findings. Any assistance rendered to him for the same will be greatly appreciated.

Yours faithfully,

Lawrence Fatsani Byson
Programme Coordinator, MSc Informatics
lbyson@unima.ac.mw – +265 881 051 359

Appendix B: Primary email contact

Subject: Participant Recruitment Exercise for Research Study – Academic

Good day [Participant's Name],

Reference is made to the telephone conversation we had a few days ago. Once again, I am Giovannie Makondi, a Master of Science in Informatics student at the University of Malawi, Department of Computer Science.

I am doing an academic research study on **"Volunteered Geographic Information (VGI), an assessment of data quality in Southern Africa: A case study of Malawi",** currently in the data collection phase. This study brings special attention to volunteered spatial data being hosted on the Malawi Spatial Data Platform -MASDAP, OpenStreetMap - OSM, Malawi Geo-tagged images across the internet, i.e., Tourism hotspots, google maps/guides and their respective users.

The study has adopted a mixed methodology approach in the quality assessment of volunteered (citizen-mapped) spatial data on the accuracy dimension compared to authoritative data sources within the same area coverage. The adopted method calls for interviewing participants and prescribing a brief survey questionnaire. A colleague recommended you based on your GIS and data management expertise and involvement in VGI in Malawi, among other projects. I write to invite you as a participant in this study officially, and it would be a great honour to have a contribution from someone with rich experience in the field of GIS and data management like yourself.

As a participant, you are asked to dedicate at most 30 minutes of your time for an oral interview with the researcher, which, with your permission, will be recorded to simplify transcription and, at most, 10 minutes on a survey questionnaire. All ethical considerations, including integrity, non-disclosure of participants' identity, sharing of results and respect for values, are taken seriously, and any questions or doubts will be cleared before the interview.

You may wish to know that your participation is voluntary and can be withdrawn at any given time without consequences. I would appreciate an opportunity to meet you at your earliest convenience.

Looking forward to hearing from you.

Sincere,

Giovannie Makondi *BSc MIS*
MSc in Informatics Student - UNIMA

Appendix C: AHP Survey Questionnaire for data quality dimensions ranking by VGI expert users.

A questionnaire for ranking Spatial Data (VGI) Quality Dimensions

Dear participant, thank you for taking the time to complete this questionnaire. The data collected on this form will be treated with the highest level of confidentiality.

For any enquires about this survey, kindly contact:

Name: Mr. Giovannie Makondi
Contact: +265884777686 / +265994371111
Email: msc-inf-14-19@unima.ac.mw / gmakondi@gmail.com

Section A: Demography

This section is required for classification purposes. Please tick in the appropriate space below:

1. **Gender** ( ) Male ( ) Female

2. **Age group** ( ) 20 years or below ( ) 21 – 30 years
( ) 31 – 40 years ( ) 41 – 50 years
( ) Above 50 years

3. **Nationality** ( ) Malawian ( ) International (specify)
____

4. Age group
( ) 20 years or below
( ) 21 – 30 years
( ) 31 – 40 years
( ) 41 – 50 years
( ) Above 50 years

5. Highest level of education
( ) MSCE or equivalent
( ) Diploma
( ) Bachelor's degree
( ) Master's Degree
( ) PhD
( ) Other (please specify)

____

6. Experience in GIS
( ) Less than a year
( ) 1 – 5 years
( ) 6 – 10 years
( ) 11 – 15 years
( ) Above 15 years

The list of dimensions of data quality in Volunteered Geographic Information

Data quality dimensions	Accuracy Validity Timeliness Completeness Consistency Uniqueness

Data quality dimensions defined

Accuracy	How well a piece of information reflects reality?
Validity	Is the information in a usable format that follows business rules?
Timeliness	Is the information available when the user needs it?

Completeness	Does the information satisfy the expectations of comprehensiveness (Is it rich in detail)?
Consistency	Does the information match a similar instance stored elsewhere?
Uniqueness	Is there only one instance of such information in the database (The data is not duplicated)?

Guidelines to answer the questionnaire.

Strength of importance	Classification	Explanation
1	Equal importance	Judgement entails equal importance between criteria
3	Moderate importance of one judgement over the other	The judgement shows a slight preference for one criterion over the other
5	Strong importance	Judgement strongly favours one criterion over the other
7	Extraordinarily strong importance	The judgement shows very strong favour of one criterion over the other
9	Extremely important	The judgement shows that one criterion is extremely preferred over the other
2,4,6,8	Intermediate values between two adjacent judgements	Choice shows compromise because an absolute verdict cannot be given.

Example: (This example has been provided to explain the structure of the questionnaire) For each of the statements below, please **compare** the **relative importance** of two factors concerning the goal of "choosing the best phone handset".

CHOOSE and **CIRCLE** only one number per row by using the following scale.

1 = EQUAL, **3** = MODERATE, **5** = STRONG, **7** = VERY STRONG, **9** = EXTREMELY IMPORTANT

Storage space	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Colour

In the case above, it is assumed that the questionnaire respondent has perceived that **"Storage space"** is **extremely important** than **"colour";** hence **9** has been circled on the side of **Storage space.**

On the other hand, if the respondent perceives that "**colour"** is of strong importance than storage space, then 5 should be circled on the side of **"colour"** as shown below:

Storage space	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Colour

Section B: Spatial Data Quality

For each of the statements below, please **COMPARE** the **RELATIVE IMPORTANCE** of two factors to the **goal**: prioritising data quality dimensions in Volunteered Geographic Information (a form of spatial data) in Malawi. Choose and **CIRCLE** only one number per row by using the following scale.

1 = EQUAL, **3** = MODERATE, **5** = STRONG, **7** = VERY STRONG, **9** = EXTREMELY IMPORTANT

Data quality in Volunteered Geographic Information

Accuracy	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Validity
Accuracy	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Timeliness
Accuracy	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Completeness
Accuracy	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Consistency
Accuracy	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Uniqueness
Validity	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Timeliness
Validity	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Completeness
Validity	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Consistency
Validity	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Uniqueness
Timeliness	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Completeness
Timeliness	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Consistency
Timeliness	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Uniqueness
Completeness	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Consistency
Completeness	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Uniqueness
Consistency	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Uniqueness

Thank you very much for your participation

Appendix D: Semi-structured interview guide – Interview one

Interview 1: Opening remarks

Thank you for taking the time to sit with me and participate in this study. I am Giovannie Makondi, a Master of Science in Informatics student at the University of Malawi, Department of Computer Science. I am conducting a study titled Volunteered Geographic Information, an assessment of Data Quality in Southern Africa: A case study of Malawi. As discussed, when I first contacted you over the telephone, you have been chosen to participate in this study due to your expertise in GIS, Data management and particularly VGI practice as an expert user of VGI. As communicated earlier, there is a need to record this interview for the entire period; kindly say yes to start recording. Thank you. As we start this interview, please state your alias and that you have agreed to be recorded. Thank you. As a reminder, remember that your participation is voluntary and unconditional. You are free to skip questions you are uncomfortable answering, and you can withdraw your participation from this study anytime without any consequences. Thank you. How are you?

Estimated interview period: 15 – 30 minutes.

Key interview questions:

1) Could you briefly explain your journey into the GIS profession,

   Follow-ups: How you became a professional? Did you see that coming? How exciting was it? Tell me more about that.

2) How did you get into VGI practice?

   Follow-ups: Was it interesting? What were your expectations? Can you expand on that? Tell me more about that.

3) What was your first impression of VGI quality before you familiarised yourself with VGI?

   Follow-ups: How did that make you feel? How did you move on from that? Tell me more about your motivation. How did that make you feel?

4) Having been in practice yourself, what is the impression and observation of VGI quality now?

   Follow-ups: Are you sure about that? Are you convinced? Are you satisfied?

5) What would you say about VGI and VGI quality now and in the future?

   Follow-ups: Do you think there is potential in VGI quality? What do you think needs to be done? Any major stakeholders you can think of?

6) Anything else you feel like adding that we have not discussed or that I have not brought up?

Interview one closing remarks.

Thank you very much for taking the time to be part of this conversation and sharing your experience in this field. Before we close and stop recording, do you have any questions or comments? Within a fortnight, you will have an opportunity to review a summary of your story once I have composed the entire narrative, mainly for you to verify the correctness of the narrative and weigh in on any emerging issues; we can call it a follow-up interview. I do not want to tell the wrong story. Once again, thank you for taking the time to participate in this study and for allowing me to invade your environment; thank you and thank you!

Appendix E: Semi-structured interview guide – Interview two

Interview 2: opening remarks

Thank you for taking the time to sit down with me as a continuation of your participation in this study. To ensure that your story is not compromised or misrepresented and captured correctly, I shared the narrative reconstruction you confirmed to have gone through. Can you affirm that?

This follow-up interview will take about 10 minutes of your time.

1. What do you make of the results of the AHP analysis? Accuracy is a dominant dimension of data quality.

   Follow-ups: Could it be different under some circumstances? Does it reflect your views? Tell me more about that.

2. The findings of the quality test for VGI report an NSSDA accuracy of about 47 metres; what is your take on that?

   Follow-ups: Are you sure? What applications of this level of accuracy can you recommend? Can you elaborate on that? Any further suggestions?

3. Having gone through your reconstructed narrative, is there anything you want to add or retract?

   Follow-ups: Tell me more about that. Are you sure?

4. How has your experience been throughout this journey as a participant?

   Follow-ups: Tell me more about that. Would you do it again? Any recommendations?
5. Anything you would like to add? An observation or comment?

Interview two closing remarks

Thank you very much for offering your time and participating in this study. Your time and commitment are greatly appreciated. Stay productive. Wishing you all the best in your career and future endeavours.

Appendix F: Individual Responses from the AHP survey questionnaire

R1	Accuracy	Validity	Timeliness	Completeness	Consistency	Uniqueness
Accuracy	**1**	0.20	0.20	0.14	0.11	0.11
Validity	5.00	**1**	0.20	0.33	0.14	0.14
Timeliness	5.00	5.00	**1**	0.33	0.20	5.00
Completeness	7.00	3.00	3.00	**1**	0.20	0.20
Consistency	9.00	7.00	5.00	5.00	**1**	7.00
Uniqueness	9.00	7.00	0.20	5.00	0.14	**1**

R2	Accuracy	Validity	Timeliness	Completeness	Consistency	Uniqueness
Accuracy	**1**	5.00	0.20	0.20	0.11	0.14
Validity	7.00	**1**	0.20	0.33	0.11	0.20
Timeliness	5.00	5.00	**1**	1.00	0.20	5.00
Completeness	5.00	3.00	1.00	**1**	0.33	3.00
Consistency	9.00	9.00	5.00	3.00	**1**	3.00
Uniqueness	7.00	5.00	0.20	0.33	0.33	**1**

R3	Accuracy	Validity	Timeliness	Completeness	Consistency	Uniqueness
Accuracy	**1**	0.20	0.33	0.20	0.11	0.14
Validity	5.00	**1**	0.33	0.33	0.11	0.20
Timeliness	3.00	3.00	**1**	0.33	0.20	7.00
Completeness	5.00	3.00	3.00	**1**	0.14	3.00
Consistency	9.00	9.00	5.00	7.00	**1**	9.00
Uniqueness	7.00	5.00	0.14	0.33	0.11	**1**

R4	Accuracy	Validity	Timeliness	Completeness	Consistency	Uniqueness
Accuracy	**1**	0.33	0.20	0.14	0.11	0.14
Validity	3.00	**1**	0.20	0.33	0.11	0.20
Timeliness	5.00	5.00	**1**	0.14	0.20	5.00
Completeness	7.00	3.00	7.00	**1**	0.20	3.00
Consistency	9.00	9.00	5.00	5.00	**1**	9.00
Uniqueness	7.00	5.00	0.20	0.33	0.11	**1**

R5	Accuracy	Validity	Timeliness	Completeness	Consistency	Uniqueness
Accuracy	**1**	0.14	0.20	0.33	0.14	0.20
Validity	7.00	**1**	0.20	0.33	0.14	0.14
Timeliness	5.00	5.00	**1**	1.00	0.20	0.33
Completeness	3.00	3.00	1.00	**1**	0.14	1.00
Consistency	7.00	7.00	5.00	7.00	**1**	5.00
Uniqueness	5.00	7.00	3.00	1.00	0.20	**1**

Appendix G: NSSDA Horizontal Accuracy computations for MW-OSM-PS vs MW-MoE-PS, RMSEx ≠ RMSEy

No	Facility Name	District	Region	MoEP Latitude	MoEP Longitude	OSM Latitude	OSM Longitude	Diff in Long	Diff in long sq.	Diff in Lat	Diff in Lat sq.	Diff in La sq. + Diff in Long sq.
1	Senjere Primary School	Zomba	South	-15.44276	35.30093	-15.44285	35.30126	-0.00033	0.00000	0.00009	0.00000	0.00000
2	Kanjeza Primary School	Blantyre	South	-15.81273	35.05116	-15.81271	35.05127	-0.00011	0.00000	-0.00002	0.00000	0.00000
3	Nyamithuthu Primary School	Nsanje	South	-16.65710	35.20779	-16.65709	35.2077	0.00009	0.00000	-0.00001	0.00000	0.00000
4	Mwanawanjovu primary School	Chikwawa	South	-16.51253	35.03526	-16.51252	35.03525	0.00001	0.00000	-0.00001	0.00000	0.00000
5	Namiyala Primary School	Nsanje	South	-16.48127	35.18693	-16.48124	35.1868	0.00013	0.00000	-0.00003	0.00000	0.00000
6	Chaona Primary School	Mangochi	South	-14.47013	35.15851	-14.47014	35.15831	0.00020	0.00000	0.00001	0.00000	0.00000
7	Kanchito Primary School	Dedza	Central	-14.36282	34.40246	-14.36289	34.40271	-0.00025	0.00000	0.00007	0.00000	0.00000
8	Moonekera Primary School	Dedza	Central	-14.37342	34.39817	-14.37363	34.3982	-0.00003	0.00000	0.00021	0.00000	0.00000
9	Magaleta Primary School	Dedza	Central	-14.41058	34.40051	-14.41055	34.4006	-0.00009	0.00000	-0.00003	0.00000	0.00000
10	Matawa Primary School	Phalombe	South	-15.66813	35.72413	-15.66806	35.72438	-0.00025	0.00000	-0.00007	0.00000	0.00000
11	Biwi LEA Primary School	Lilongwe	Central	-14.00350	33.79090	-14.00331	33.79096	-0.00006	0.00000	-0.00019	0.00000	0.00000
12	Monjo Primary School	Phalombe	South	-15.72814	35.73369	-15.72823	35.73343	0.00026	0.00000	0.00009	0.00000	0.00000
13	Mkulula Primary School	Lilongwe	Central	-13.76760	33.81592	-13.7679	33.81594	-0.00002	0.00000	0.00030	0.00000	0.00000
14	Lilongwe Demonstration Primary School	Lilongwe	Central	-13.89216	33.77626	-13.89219	33.77646	-0.00020	0.00000	0.00003	0.00000	0.00000
15	Muzu Primary School	Lilongwe	Central	-13.94627	33.71678	-13.94619	33.71656	0.00022	0.00000	-0.00008	0.00000	0.00000
16	Pheleni Primary School	Lilongwe	Central	-13.92638	33.71312	-13.92651	33.71317	-0.00005	0.00000	0.00013	0.00000	0.00000
17	Chatuwa Primary School	Lilongwe	Central	-13.93528	33.77267	-13.93533	33.77255	0.00012	0.00000	0.00005	0.00000	0.00000
18	Chejika Private Primary School	Lilongwe	Central	-13.94205	33.77786	-13.94201	33.77784	0.00002	0.00000	-0.00004	0.00000	0.00000
19	Kalonga LEA Primary School	Lilongwe	Central	-13.93888	33.75677	-13.93892	33.75692	-0.00015	0.00000	0.00004	0.00000	0.00000
20	Gumbu Primary School	Ntcheu	South	-14.82659	34.63547	-14.82668	34.63544	0.00003	0.00000	0.00009	0.00000	0.00000
21	Bangala LEA Primary School	Ntcheu	South	-14.84596	34.68626	-14.84591	34.68631	-0.00005	0.00000	-0.00005	0.00000	0.00000
22	Mzinga Primary School	Karonga	North	-9.85006	33.86202	-9.84984	33.86191	0.00011	0.00000	-0.00022	0.00000	0.00000
23	Malo Private Primary School	Mzimba	North	-9.82736	33.89087	-9.82746	33.8911	-0.00023	0.00000	0.00010	0.00000	0.00000
24	Lulindo Primary School	Karonga	North	-9.93707	33.93299	-9.9369	33.9329	0.00009	0.00000	-0.00017	0.00000	0.00000
25	Bwiba Primary School	Karonga	North	-9.94818	33.89040	-9.94841	33.89025	0.00015	0.00000	0.00023	0.00000	0.00000

26	Ipyana Model Primary School	Karonga	North	-9.97175	33.91591	-9.97176	33.91615	-0.00024	0.00000	0.00001	0.00000	0.00000
27	Livingstonia Primary School	Rumphi	North	-10.59883	34.10691	-10.59892	34.10702	-0.00011	0.00000	0.00009	0.00000	0.00000
28	Beehive Private Primary School	Mzimba	North	-11.47648	34.00395	-11.47642	34.00394	0.00001	0.00000	-0.00006	0.00000	0.00000
29	Kanyerere Primary School	Rumphi	North	-10.94944	33.75950	-10.94938	33.75971	-0.00021	0.00000	-0.00006	0.00000	0.00000
30	Ngala Primary School	Lilongwe	Central	-14.15183	33.88226	-14.15195	33.88206	0.00020	0.00000	0.00012	0.00000	0.00000
31	Karonga School for the deaf	Karonga	North	-9.94872	33.88254	-9.94864	33.88274	-0.00020	0.00000	-0.00008	0.00000	0.00000
32	Mchacha Primary School	Chikwawa	South	-16.23464	35.03433	-16.23464	35.03458	-0.00025	0.00000	0.00000	0.00000	0.00000
33	Chikonje Primary School	Nsanje	South	-16.49559	35.20688	-16.49557	35.20678	0.00010	0.00000	-0.00002	0.00000	0.00000
34	Thangadzi 1 Primary School	Nsanje	South	-16.55125	35.11730	-16.55121	35.11732	-0.00002	0.00000	-0.00004	0.00000	0.00000
35	Kamuzu LEA Primary School	Kasungu	Central	-13.09262	33.53337	-13.09265	33.53366	-0.00029	0.00000	0.00003	0.00000	0.00000
36	Chitsime Primary School	Blantyre	South	-15.76801	35.03668	-15.76808	35.0369	-0.00022	0.00000	0.00007	0.00000	0.00000
37	Mwangothaya Primary School	Mulanje	South	-15.89236	35.40513	-15.89241	35.40494	0.00019	0.00000	0.00005	0.00000	0.00000
38	Nthola Primary School	Karonga	North	-9.99375	33.92155	-9.99381	33.92124	0.00031	0.00000	0.00006	0.00000	0.00000
39	Chisambo Primary School	Mulanje	South	-16.04753	35.71156	-16.04738	35.7113	0.00026	0.00000	-0.00015	0.00000	0.00000
40	Kings Foundation Primary School	Ntcheu	South	-14.80382	34.62074	-14.80368	34.62047	0.00027	0.00000	-0.00014	0.00000	0.00000
41	Mtonya Primary School	Dedza	Central	-14.39870	34.43634	-14.39877	34.4367	-0.00036	0.00000	0.00007	0.00000	0.00000
42	Nazombe Primary School	Phalombe	South	-15.76949	35.76914	-15.76961	35.76878	0.00036	0.00000	0.00012	0.00000	0.00000
43	Mponda Primary School	Balaka	South	-14.98591	34.94446	-14.98561	34.94416	0.00030	0.00000	-0.00030	0.00000	0.00000
44	Mtsiliza Primary School	Lilongwe	Central	-13.95216	33.73665	-13.95222	33.737	-0.00035	0.00000	0.00006	0.00000	0.00000
45	Chilobwe Primary School	Lilongwe	Central	-14.54563	34.51497	-14.54556	34.51533	-0.00036	0.00000	-0.00007	0.00000	0.00000
46	Karonga Demo Primary School	Karonga	North	-9.97185	33.90046	-9.97178	33.9001	0.00036	0.00000	-0.00007	0.00000	0.00000
47	Mphungu Primary School	Lilongwe	Central	-13.95564	33.81147	-13.95595	33.81122	0.00025	0.00000	0.00031	0.00000	0.00000
48	Kazomba Primary School	Mzimba	North	-11.95474	33.61027	-11.95511	33.61046	-0.00019	0.00000	0.00037	0.00000	0.00000
49	Blantyre Girls Primary School	Blantyre	South	-15.78122	35.02455	-15.78089	35.02443	0.00012	0.00000	-0.00033	0.00000	0.00000
50	Mzuzu SOS Primary School	Mzimba	North	-11.47163	34.00456	-11.47148	34.00444	0.00012	0.00000	-0.00015	0.00000	0.00000
								SUM	0.00000		0.00000	
								AVERAGE	0.00000		0.00000	
								RMSE	0.00021		0.00014	

										KM	**M**	
								RMSEx/RMSEy	0.66806			
								CSE 39.35%	0.00017	0.01921	19.20786	
								ACCr 95%	0.00042	0.04702	47.01508	