2015

An Analysis of Quality for Volunteered Geographic Information

Joel C H Meier
Wilfrid Laurier University, joelmeier@gmail.com

Recommended Citation

Meier, Joel C H, "An Analysis of Quality for Volunteered Geographic Information" (2015). *Theses and Dissertations (Comprehensive)*. 1712.
https://scholars.wlu.ca/etd/1712

An Analysis of Quality for Volunteered Geographic Information

by

Joel Clarence Howard Meier

B.A., Geography University of Guelph, 2010

P.G. Dip, Centre of Geographic Sciences, 2011

THESIS

Submitted to the Faculty of Arts Geography and Environmental Studies

in partial fulfillment of the requirements for

Master of Science in Geomatics

Wilfrid Laurier University

Abstract

Volunteered geographic information is a new and growing source of information for generating accurate maps which display various landscapes and populations. This thesis discusses the definitions, history and theory of VGI with an eye toward its practical applications for geographic and environmental research. A major issue with the practical application of VGI is uncertainty about how to measure and characterize data quality. This thesis investigates issues of data quality for VGI by using the International Standards Organization framework for data quality assessment for geographic information. The individual data quality elements for geographic information are defined and measures appropriate for VGI are proposed. Three case studies are detailed and the results conclude that the quality of volunteered data depends on both context and content.

Acknowledgments

I started this journey what feels like a very long time ago not entirely sure where it was taking me. It is with great gratitude that I write my acknowledgements section in order to thank and pay tribute to the people who helped me along this journey of discovery. I learned so much about myself and I am very happy it is finally over. I am very fortunate for the opportunities I had and the life I now live. Which leads me to first and foremost, I want to thank my parents for their support throughout the process, always encouraging me to strive for more, my advisor Colin, for his patience with me, my colleagues and friends who started the journey with me and the ones I met along the way. Finally I'd like to thank the reader, because if you're reading this then you have at least read one page of my thesis and thank you for that.

TABLE OF CONTENTS

List of Tables

List of Figures

List of Abbreviations

ESRI Environmental Systems Research Institute

GIS Geographic Information Systems

VGD Volunteered Geographic Data

VGI-S Volunteered Geographic Information Systems

VGI Volunteered Geographic Information

KML Keyhole Markup Language

GPS Global Positioning System

HSDA Hardware, Software, Data and Applications

GPX GPS Exchange Format

XML Extensible Markup Language

OSM Open Street map

RMS Root Mean Squared

HMGS Hellenic Military Geographical Service

IMAP and Input, Management, Analysis and Presentation

ISO International Standards Organization

1. Introduction

Volunteered geographical information (VGI) has emerged as a new paradigm of data production in geography due to advances in technology that have facilitated communication and collaboration between people on a global scale. VGI can be defined as “user generated content within a spatial context,” (Goodchild 2007, 212). The definition provided by Goodchild for VGI is purposefully vague because VGI can encompass so many different types of geographic information. Because of this, VGI can be considered to fall into the category of non-expert information common to Web 2.0 activities such as wikis and crowdsourcing. NeoGeography has emerged as an alternative to traditional academic geography or paleogeography; VGI is a new science in which there is “a blurring of the traditional roles of subject, producer, communicator and consumer of geographic information” (Goodchild 2009, 82). This commingling of source, resource, creator and user leads to a new cartography, where collaboration and democracy are integrated as an essential aspect of NeoGeography (Turner 2006).

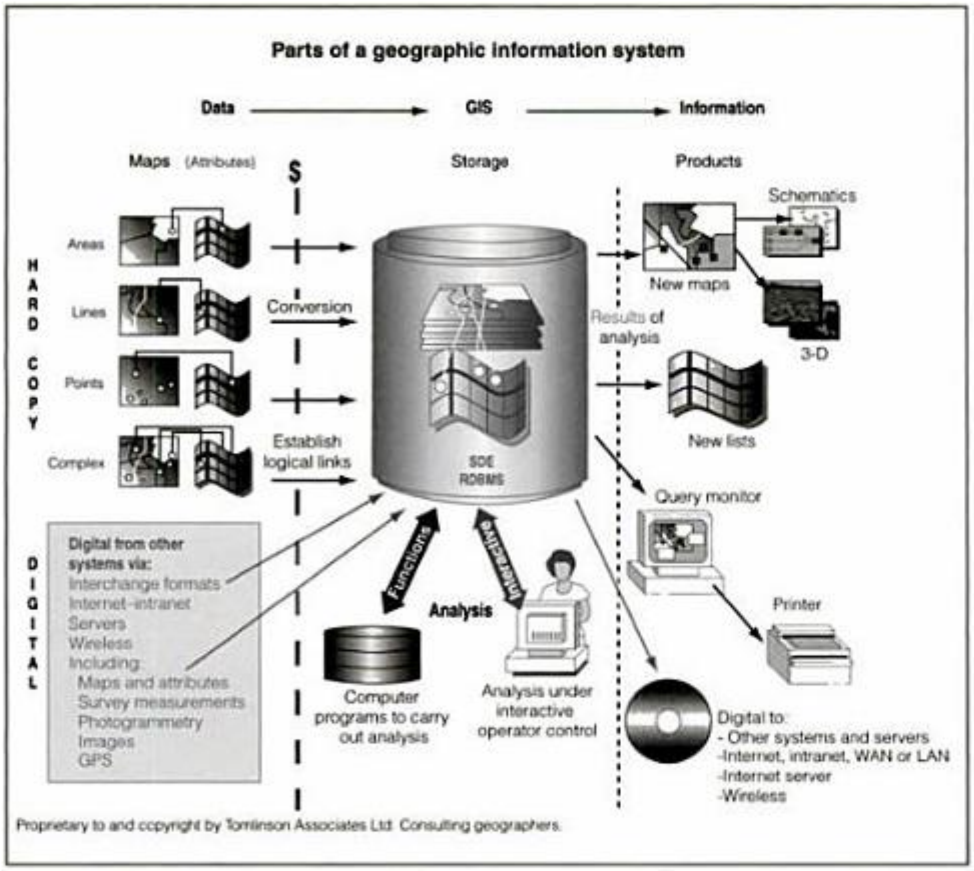
1.1 Overview

While the term VGI is widely used in academic literature, some have argued that the name volunteered geographic information is semantically misleading, because ‘information’ implies meaningful analysis which is built upon data points and their presentation or interpretation, thus it would be an improper label for a primary data source (Rinner and Fast 2013). Essentially, geographic information is information with a location component, which differs from geographic data because it may incorporate multiple sources of data as well as draw or infer conclusions about that data based on presentation. Translating data into information requires an understanding of what is represented by the data, which includes an assessment of data quality. Furthermore, a geographic information system incorporates hardware, software, applications, and data, whether user-generated or from other sources (Rinner and Fast 2013).

Volunteered Geographic Information Systems; “VGI-S” as outlined by Rinner and Fast (2013) goes beyond data and incorporates the input of the researcher who assembles data into meaningful conclusions, VGI-S employ hardware, software, data and application components, “HSDA”, and input, management, analysis and presentation “IMAP” functions, to transform web-based user-generated geographic data into information” (Rinner and Fast 2013). In other words, Volunteered Geographic Data (VGD) differs from VGI and VGI-S in that it provides raw data (points, polygons, or lines typically) which are parsed and analyzed through the VGI-S in order to produce VGI. The process by which different types of data, including VGD, can be rendered into information by using a GIS system (Rinner and Fast 2013; Tomlinson 2003).

However, the precursor to this process is an assessment of the underlying data quality of the VGD, which in part, will be determined by the data, and in part, by the intended use of the information product. Figure 1 also illustrates the role computers and humans play in performing analysis through functions and interactions with the stored data (Rinner and Fast 2013; Tomlinson 2003)

Figure 1 Parts of a Geographic Information System (Tomlinson 2003)



Source:
Tomlinson
(2003)

There has been extensive research into the quality of traditional geographic information, which has led to the creation of data standards as set out by the International Organization for Standardization, including ISO/IEC IS 19113:2000 2002; ISO/IEC IS 19114:2003 2009, ISO 19138:2006 and ISO 19157:2013. ISO 9000 is a family of standards, related to quality management systems. These specifications define quality as the degree to which a set of inherent characteristics fulfills requirements. Content quality is related to whether this content is useful or not for a user’s purposes. In user-generated content, data quality, a qualitative concept, is also associated with the user’s trust in the content (a subjective concept); this leads to a connection between content quality and the original provider’s authority (Brando and Bucher 2010). For example, data contributed to a VGI system by an employee of a national mapping agency may be perceived to have higher quality than that of an ordinary citizen.

Many researchers believe that VGI offers significant potential to bring mapping and spatial data collection to citizens e.g., citizens as sensors and provide researchers with new sources of data for previously hard-to-survey social and environmental processes(Goodchild 2007). Because VGI is usually inexpensive and accessible to many more people than solely professional cartographers or national mapping agencies, VGI has a huge potential to engage citizens in place-based issues and provide significant, timely, and cost-effective source for geographer’s and other spatially related fields of research and management (De Longueville, Smith and Luraschi 2009, 73). These purported benefits have served to increase the interest in assessing both the potential applications and current limitations of VGI.

However, the benefits of VGI as a replacement or supplement to traditional GI are also a source of debate and point of contention amongst geographers and cartographers (Heipke 2010). For example, while some research has demonstrated that VGI can be of high quality similar to that of traditional GI (Haklay 2010), the distribution of submitted data follows areas of easy accessibility (Haklay and Muki 2013) or interest wanes after initial enthusiasm in a VGI project and that can have a negative effect on the quality

(Coleman 2009). Thus it is necessary to provide tools that can help to distinguish between different sources of VGD and different intended uses of derived information products which can therefore provide a rationale for allocating financial (and other) resources to the uptake of VGI in both research and practical applications. Even in cases where research is conducted, the funding that is required for implementation can serve as another disincentive for supporting VGI applications. Thus, the goal of this paper is to use the existing framework of the ISO standards for geographic information quality and relate them specifically to VGI in order to justify further investment in VGI. By developing a series of metrics that can provide tangible verification of the claims regarding the merits of VGI, the quality of VGI data can be evaluated more effectively. After VGI metrics have been introduced, three case studies will illustrate the practical uses of user provided data for the geographical sciences.

To develop explanatory metrics, the methods of determining the quality of the information obtained through VGI must be thoroughly evaluated and defined. High quality data is often associated with positional accuracy. Positional accuracy is used to "describe the quality of geo-data collected and produced with a commissioned effort, which entails a specific and uniform method to gather and process the data" (Excel, Dias and Fruijiter 2010, 213). Furthermore, it is also important to keep in mind that positional accuracy itself is affected by a wide variety of factors. As Helbich and Amelunxen note, "the positional accuracy of collected data is affected by different influences, [such as] the technological bias like the accuracy of the GPS-receiver used, different data acquisition techniques (e.g. digitizing), or subjective knowledge about the data gathering process"

(Helbich, Amelunxen and Neis 2012, 28). As the literature demonstrates, there is a technological and interpretive component that impacts the quality of gathered and processed data. Namely, the process by which users acquire and catalog geographic information can affect the quality and usefulness of the data.

The primary shortcoming of these methods of assessment is that they rely on a single dimension to assess the quality of the data submitted. However, there are many aspects of data quality within VGI that extend beyond mere positional accuracy. Expanding upon accuracy-based assessment metrics, there are alternative accuracy determiners, including lineage, attribute accuracy, logical consistency, completeness, semantic accuracy, usage, purpose, constraints, temporal quality, variation in quality, meta-quality, and resolution (Excel Dias, and Fruijiter 2010, 213). These alternative metrics can also be utilized in lieu of or in compliment to metrics that contribute to accuracy determiners. By taking multiple metrics into account, a more complete view of the accuracy and relevancy of geographical data can be achieved.

It also must be noted that, when it comes to the modern field of cartography, concerns about accuracy are commonplace. This can be attributed to the prevalence of inadequately educated and poorly trained individuals who can nonetheless provide information about locations and mapping. Haklay et al. (2010) argue that the issue of spatial data quality is a clear challenge in the area of volunteered geographic information. The data that are contributed to VGI projects do not comply with standard spatial data quality assurance procedures, and the contributors operate without central coordination and strict data collection frameworks. In consideration of these concerns, additional precautions must be taken to ensure the veracity of the data that is collected. Furthermore, any employed framework should consider the methodology of data collection. (Haklay et al. 2010)

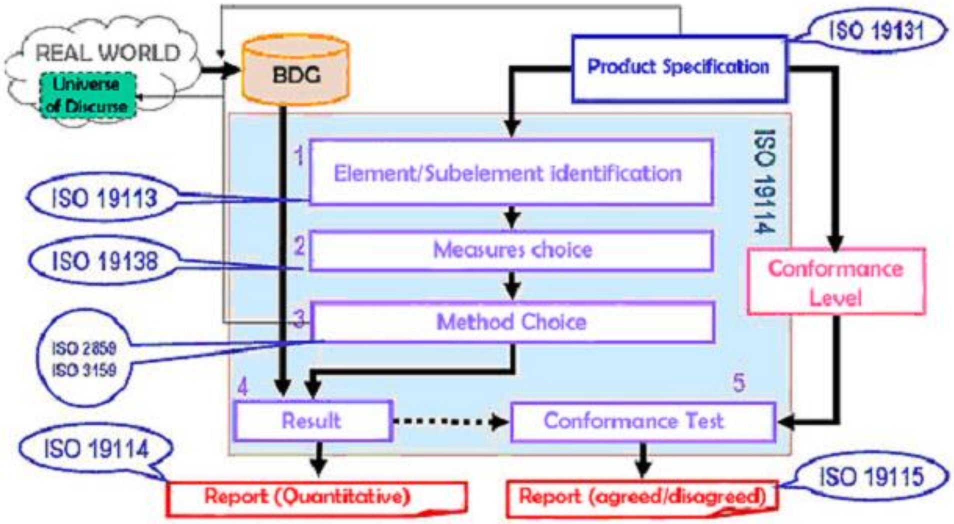
Additionally, factors like user ratings of restaurants or the rating system for hiking trails encompass subjective opinions that are difficult to verify empirically as they may not be comparably quantifiable (De Longueville, Smith and Luraschi 2009). For the purposes of this paper, VGI is defined as user-created positional data with descriptor attributes. This definition of VGI is more in line with traditional geographic information representations that include point, polygon or line spatial data, in addition to attributes that do not include personal opinions. These attributes may include names or points of interest which could be assessed for nomenclature consistency and positional accuracy (Sharma 2011).

Assessing the nomenclature consistency allows for a simplified appraisal of the attribute data. (Burrough, van Rijn and Rikken 1996). The attribute data which is an outlier, would illustrate the need for further analysis to see if the outlier is important. Little research has been conducted in terms of measuring and quantifying quality for VGI (Brando and Bucher 2010).

International standards are important in order to compare and evaluate the quality of data and its subsequent analyses. There has been a significant amount of research into the quality of GI, much of which centers on the ISO 19113 and ISO 19114 (ISO/IEC IS 19113:2000 2002; ISO/IEC IS 19114:2003 2009) quality principals for GI outlined in Table 1. The latest ISO standard for the quality of geographic information is ISO

19157:2013. The flowchart for ISO compliance including ISO 19114 and other relevant data collection standards is outlined in Figure 2.

Figure 2 Flowchart for ISO Standard Compliance (Ariza n.d.; Ceballos and Reyes-Gatica n.d.)



Geographic data is an ongoing area of ISO standards development, as technologies and the contexts in which data is used evolve. Busch and Willrich (2002), Caspary, Wilhelm, and Joos (2002) and Poser and Dransch (2010) all emphasize the use of the ISO elements and subelements of data quality for assessing the quality of geographic information. ISO standards help to eliminate or mitigate the many factors that favour errors in the production and handling of geographical information such as: experts’ and users’ lack of knowledge, poor command of technical aspects, ignorance about databases genealogy, etc (Ceballos and Reyes-Gatica n.d.).

1.2 Research Goals and Objectives

This research is intended to help fill the gap between current research on quality of GI, user submitted data, and VGI, and the concept of fitness-for-use (Grira, Bedard and Roche 2010; Sui, Elwood and Goodchild 2013). The primary research goal is to understand related components of data quality for VGI and how to measure these components. Using existing data and three case studies, this research will quantify and predict VGI quality and determine the fitness-for-use for different applied contexts related to social and environmental research. Table 1 helps to delineate the different elements and metrics of data quality which can be taken into account when evaluating the overall accuracy of VGI data. In simple terms, maps created by means of user-generated data can be compared to maps created by professional cartographers in order to determine accuracy.

Table 1 Elements, Subelements and Evaluation Methods for GI

Data quality elements	Data quality subelements	Evaluation method
Completeness	Omission	Compare count of items in dataset against count of items in reference data.
	Commission	
Logical consistency	Conceptual consistency	Count the number of features and relationships which violate the conceptual schema.
	Domain consistency	Compare attributes against acceptable domains. Count the violations.
	Format consistency	Compare the record structure for all items to field definitions. (Boolean result)
	Topological consistency	Check the boundaries for closure or duplicates. Count the number of inconsistencies.

Positional accuracy	Absolute or external accuracy	For each node measure the error distance between absolute coordinate values of the node in the dataset and those in the reference data. Compute RMS from error distance.
	Relative or internal accuracy	Measure the relative error distance between relative coordinates values, compute RMS from error distance.
	Gridded data position accuracy	Measure the error distance between a gridded point and the reference data, calculate RMS from error distance.
Temporal accuracy	Accuracy of time measurement	Measure the difference between occurrence in the dataset and in the reference data. Compute RMS from time difference.
	Temporal consistency	Confirm the temporal order (Boolean result)
	Temporal validity	Check to confirm the date of acquisition is true (Boolean result)
Thematic accuracy	Classification correctness	Compare classified item against true class using Kappa coefficient.
	Non-quantitative attribute correctness	Compare non-quantitative attributes against those in the universe.
	Quantitative attribute accuracy	Measure the difference between quantitative attributes and those in the universe, compute RMS from the difference.

Sources: (ISO/IEC ISO 19113:2000 2002; ISO/IEC IS 19114:2003 2009.)

2. Literature Review

VGI in the form of downloaded digital data is a newer data source than traditional VGI (Goodchild 2007). An example of previous forms of VGI, such as the Christmas

Bird Count (CBC), would engage volunteers to go out and collect avian data for the National Audubon Society, which would subsequently publish the results in hard copy. Being over 100 years old, the Wiersma shows there has been significant criticism regarding lack of accuracy and quality with the CBC (Wiersma 2010, 1-9). New data collection methods may involve mobile devices, including GPS-enabled phones, which allow geotagging of images, tweets, and other data points which may be utilized as VGD. Newson and Noble (2003) discuss how ArcGIS is a valuable tool for visualizing bird species by combining multiple sets of data, some of which include VGI.

With advances in technology, VGI projects, including those such as the CBC, have adopted new means for collecting and distributing data. Nunez-Redo et al. (2011) explain how updated VGI projects that use modern technology, such as internet tools, "allow active user participation that is becoming a massive source of dynamic geospatial resources." Davis (1996, 421) also explores how the connectivity provided by the internet can "greatly enhance GIS productivity, particularly for users largely out of the world mainstream of activities, such as much of the Third world." When expert data is unavailable or out of date, volunteered geographic data can be an effective way to generate useful geographic information.

Furthermore, one only has to look at the popularity of technologies such as Google Maps and Mapquest to see the widespread usefulness of these technologies and the willingness of the general public to adopt them. Businesses and consumers rely on the ability to get directions, map out clients or prospective sales, and track couriers and messengers (Wang and Strong 1996). Open-source mapping efforts such as

OpenStreetMap demonstrate that there are also alternatives for consumers and creators of mapping data.

Fortunately, for those who rely on these technologies to make their day-to-day lives and tasks easier, collecting VGI is no longer a long or difficult process. In fact, due to the ease of collecting VGI, there has been an increase in VGI availability (Wiersma 2010). As downloaded raw data, user-created GI is an untested source and currently is not fulfilling its potential as a new source of data available to researchers (Sui and Delyser 2012).

Goodchild (2007; 2009) finds many problems with user-generated data. There are numerous issues with data which can be submitted by anyone, including the accuracy of such data, the potential for such data to violate privacy and raise privacy related concerns, and the negative effect that a reliance on user-generated data could have on more conventional, proven, and widely accepted methods of gaining such information. Goodchild (2007, 212) explains that "anyone with an internet connection can select an area on the Earth's surface and provide it with a description, including links to other sources." This is where the problem of open access arises, because many submitters of geographic data are not trained as professional cartographers.

If any person who can get on the internet can offer up some information, there is no way to determine its accuracy other than through validating it by knowledgeable, professional cartographers. That in and of itself takes a great deal of time and would lead to cartographers, who could be spending time on more valuable pursuits, wasting a lot of time and energy checking information that is not in the least bit factual. Hopefully, cartographers would be able to spot non-factual information fairly quickly and dismiss it, but that still requires a lot of time that cannot be replaced. Furthermore, one has to wonder why society is currently fascinated with letting the citizen scientist contribute to fields they possibly know nothing about. For example, most laymen are unable to contribute meaningfully to the field of physics; thus the trend toward user-generated data is, for the aforementioned reasons, not one that certain industry professionals and researchers are particularly fond of (Goodchild, 2009).

However, this type of collaboration is becoming increasingly popular in the field of geography among other research disciplines which have an appetite for vast amounts of data. Internet collaboration which facilitates information sharing and user-created content is referred to as Web 2.0 (Anderson 2008). Web 2.0 allows users to upload and display whatever type of information they so choose, including spatial information (DiNucci 1999). Web 2.0 is a phenomenon that has started to become the norm in the past decade. One mainstream example of a Web 2.0 application is Wikis, which were originally adopted by large companies so that many users could collaborate on a single project. Internet collaboration became popular in the geographic community when users started to build online communities to display spatial information visually(DiNucci 1999).

As in the software world in general, open-source products differ in significant ways from their proprietary counterparts. They are generally free to use, and their code and datasets are much more transparent than commercial offerings. Companies like ESRI may charge steep amounts to users for GIS solutions which are closed-source. Though they may offer greater support and training than open-source solutions which are usually supported by communities of volunteers, software suites such as ArcGIS lack the freedom and transparency of open-source mapping tools. There are a number of open source GIS tools available, but they are generally not as robust or well-supported as ESRI's ArcGIS. For example, Chen et al. (2010) describe how the coding languages Python and Eclipse have been used to create completely open-source GIS packages, rather than simply in the creation of scripts to be used for extending the capabilities of ArcGIS. Butler (2004) also describes the vast capabilities Python scripts provide for expanding ArcGIS beyond its intended capabilities.

Despite the controversies regarding the use of closed source software for research purposes, ArcGIS is by far the most widely used GIS product: according to a survey "of nearly 40,000 GIS professionals found that ArcGIS is the dominant GIS platform, with 78% of respondents reporting that they used ArcGIS or related ESRI products and only 27% reporting that they used the next most popular GIS product" (Roberts 2010). Additionally, "ESRI software is used in more than 350,000 organizations worldwide including each of the 200 largest cities in the United States, most national governments, more than two-thirds of Fortune 500 companies, and more than 7,000 colleges and universities"(ESRI 2014). Open source mapping projects which compete with corporate offerings are, however, becoming more popular. The web hosts a number of GIS products which may be commercial, non-commercial, closed-source, open-source, or a mixture of these types.

Table 2 Examples of VGI

Website	Description
Panoramio	Geolocation-oriented photo sharing website which interacts with Google Earth. Panoramio utilizes the interaction of humans to determine the accuracy and value of each photo uploaded. This is related to the ISO standards, every Panoramio photo is a candidate for transfer to the Google Earth Panoramio layer
Ushahidi	Collects eyewitness reports of violence sent in by email and text-message and places them on Google maps.
Google Map maker	Google Map Maker is a service launched by Google in June 2008, designed to expand the breadth of the service currently offered by Google Maps. In some countries mapping data is unavailable, and so to combat this problem Google has decided to open up Google Maps to a collaborative community effort in certain territories.
OpenStreetMap	OpenStreetMap (OSM) is a collaborative project to create a free editable map of the world. Compared to Google Map Maker, OSM has data which is freely "available to download with an open license giving everyone the freedom to reuse, redistribute and build applications with it" (
EveryTrail	EveryTrail is a global web2.0 platform for geo-tagged user-generated travel

	content that Is changing the way millions of people share travel experiences and plan trips.
EBird	A real-time, online checklist program, eBird has revolutionized the way that the birding community reports and accesses information about birds.
TrailFu	TrailFu provides mountain biking trail VGD for the United States, Canada, and other countries in the form of KML and GPX XML files which are created from user logs.

In 2004, for example, OpenStreetMaps (OSM) was launched with the intent to create a free editable map of the world. This concept garnered attention, and, as of June 2014, there were 1,654,095 users and 4,026,124,247 uploaded GPS points (OpenStreetMap 2014). This open mapping platform is similar in nature to Wikipedia, which is a powerful open encyclopedia that allows users to add information on any topic. As is to be expected, some of that information is accurate, and some of that information is deliberately or unintentionally spurious.

There have been several previous studies such as, Ather 2009; Kounadi 2009; Zilske, Neumann, and Nagel 2011 on the quality of Open Street Map data and most have proven Linus’s Law which states "given enough eyeballs, all bugs are shallow"; meaning that with enough collaboration between users that any error by a single user will be picked up and corrected by another user (Ather 2009). As Kounadi (2009, 1) notes, “there is a concern about [the data’s] quality because the volunteers that contribute the data lack the sufficient cartographic training and the quality cannot be guaranteed. Kounadi (2009,

1) finds that, using buffer analysis, the user generated data is quite accurate when compared with official cartographic data from the Hellenic Military Geographical Service (HMGS), the official cartographic service in Greece. As the length completeness was 88%, the name accuracy more than 87% and the average percentage of overlap between the two datasets was more than 89%. There were some significant differences, however, as the name completeness and the type accuracy had the lowest results; 26% and 33% respectively (Kounadi 2009, 1). In terms of total length completeness, the OSM road network was 12% (81,974 meters) less than the HMGS network (Kounadi 2009, 53).

These differences illustrate the disparities between certain ISO standard requirements, such as a high area of overlap in length completeness, but a poor sense of thematic data in user submitted data. In the grid cell covering Acropolis the only area where OSM data was more complete than HMGS data was, where "in order to protect the landscape there are many pedestrian roads; vehicle access is not permitted" and a "more detailed observation of this grid square reveals that the extra road segments in the OSM dataset represent footway, track and pedestrian road types whereas these types of road are not represented at all" in the official figures (Kounadi 2009, 53). As with OSM, Google Map Maker, and other GIS products, users can provide data which experts might be ill-equipped to provide due to technological or other constraints.

Kounadi (2009, 31) ran into two problems resulting from the lack of name and type accuracy demonstrated by user-uploaded data: firstly, when a road changes type and, secondly when some roads have two different road names though they may appear to be one section of the same road. These two problematic accuracy issues demonstrate the lack of users' specific methodology when editing the OSM lines and could have distorted the results of the analysis if not labeled correctly (Kounadi 2009, 31). Figure 3 Overlap between OSM and ordinance data.shows some of the strengths and limitations of VGD when compared to ordinance data: smaller roads and linkages which curve tend to have more inaccurate data, reflecting possible confusion on the part of OSM submitters as can been seen by the difference in green and blue linear data around the main intersection of the figure. The comparison of these types of data is a matter of perspective and involves assumptions regarding the supremacy of official data, particularly in the thematic areas of type and nomenclature.

Figure 3 Overlap between OSM and ordinance data.

One could also argue that user-generated data has the potential to be more accurate due to temporal considerations, as users can submit data to websites the instant it is captured, and this recent data may reflect changes in nomenclature or structural facets that have yet to be described by traditional or official maps. Kounadi (2009, 59) also points out how slightly tweaking the algorithm used to compare the datasets can result in increased levels of accuracy. For instance, one road mapped by OSM fell “exactly at the edge of the HMGS buffer meaning that if the buffer was slightly larger, 0.5 meters, the specific road would have a higher percentage overlap Kounadi (2009, 59).” As shown in the case studies which will be discussed later on, the buffer zone is an important, but subjective tool which can be used to compare the relative agreement of different sources of VGD and official datasets.

An eight stage analysis developed by Koukoletsos, Haklay and Ellul (2011) assesses both completeness and accuracy by measuring the percentage of buffer overlap between proprietary and OSM data. Generally speaking, proprietary sources tend to be much more complete than VGI-based efforts such as OSM, making the issue of completeness testing an important factor for the acceptance of VGI (Koukoletsos, Haklay and Ellul 2011).

A similar analysis of OSM road networks in London England found that 57% of user-created roads had between 85% and 100% overlap between ordinance survey datasets and OSM data (Ather 2009)**.** In addition, there was a positive trend between the number of users and the completeness of the dataset. With OSM it is possible to create very accurate and precise data using only simple off the shelf GPS devices (Ather 2009). Other studies that have focused on OSM have found similar results. If there are more than 15 users per square kilometer the positional accuracy is much higher according to Haklay et al. (2010)**.**

Users can contribute any information they wish to OSM using its technical infrastructure, which was initially created and is continually maintained by a group of volunteers. Aside from the volunteers, OSM employs software developers who seek to create usable software that will make OSM accessible to even more people. Some critics have questioned whether open access databases such as Wikipedia and OSM can provide accurate, useful data, and whether there is a moral obligation to attempt to maintain objectivity and factuality in a world where each individual has the potential for content creation. (Casanovas and Poblet 2012)

Casanovas and Poblet (2012) also speculate that the more people who have access to OSM, the higher the potential for more accurate information, the opposite can also be true. More people being able to access OSM means more possible mistakes and non-factual data entered into the VGI-S, which the OSM team will then have to edit or discard. If that non-factual information is accidentally published, all it would take would be one teacher or author to disseminate this incorrect information and, from there, for misinformation to spread beyond control, creating a chain of inaccuracy that is hard to follow and correct.

This issue has been demonstrated in real-life, as noted by Randall when they state "this kind of feedback loop—wherein an error that appears on Wikipedia then trickles to sources that Wikipedia considers authoritative, which are in turn used as evidence for the original falsehood—is a documented phenomenon" (Randall 2014). Obviously, there are problems inherent in the OSM technology and process which are also present in the technology of similar sites, but those problems have not slowed the growth and popularity of such technologies. In fact, OSM has become widely used since 2010 (Randall 2014).

OSM initially came to the attention of the media during the Haitian earthquake of January 2010 (Zook et al. 2010). Many relief agencies were frustrated with the lack of accurate and up to date information and turned to VGI for help in locating areas in Haiti, such as roads or buildings, which would allow rescuers to better identify areas of need and better allocate resources (Zook et al. 2010). While this certainly seems to speak of the benefits of VGI, the truth is that one must still consider the possible downsides. If that information is inaccurate, scarce monetary and temporal resources would be wasted trying to locate these victims using incomplete or erroneous information. Despite this uncertainties, some researchers have found that VGI is quite useful during crises, providing relatively accurate spatial and temporal information (De Longueville, Smith, and Luraschi 2009; Poser and Dransch 2010; Okolloh 2009). Perhaps because few people consider these possible negative implications as it is possible that even bad data triumphs over having no data. The role of VGI in aiding response operations during the Haitian earthquake has been widely viewed as a success. The issue of fitness-for-use would appear to be applicable in these types of situations, where data which can be considered accurate overall as more data points stream in, even if there are numerous erroneous or misleading entries. The general picture provided by VGD gives users of GIS an excellent view of the current, ground-level situation in a crisis for which there is simply no official data available.

However, if data quality for VGI were more quantifiable, this apparent success could be replicated for future crises and contribute to new areas of research (Zook et al. 2010). The bottom line is that OSM and similar technologies are often viewed as more beneficial than they possibly are and perhaps are not viewed as being as potentially dangerous. However, not enough research has been conducted as of yet to quantify the usefulness or danger of OSM technologies, and the fact that such dependence is being placed on potentially unfounded resources is alarming.

Two facets of quality information that OSM and similar technologies lack are reliability and inherent trustworthiness. With traditional maps, there is implied trust in the creator of the data, whether it is a private company or a government organization (Poser and Dransch 2010). Perhaps that fact, in and of itself, is a problem. While it can be assumed that all maps created by legitimate companies with seasoned professionals working for them are trustworthy, mistakes can and do happen. Goodchild touches upon this problem, stating that

> Mapping by national mapping agencies has involved high levels of expertise, in the form of cartographic skills, knowledge of the operation of machinery, and familiarity with the subject matter. The information produced by mapping agencies was and is credible first and foremost because its employees are perceived as experts by the user community, who are in turn willing to believe that map making requires expertise. But these perceptions and beliefs have been shaken by the fundamental changes that have occurred in geographic information technologies in the past few decades (Goodchild 2007, 240).

Even considering these facts, many geographers assume that printed, tangible maps are likely to be more accurate than their online counterparts. There is a significant counterargument to this concept. Many researchers contend that it is old-fashioned to assume that these online maps, which they claim have also been edited and studied to make sure they are as accurate as possible, are just as reliable as their physical, printed counterparts; and that the notion that just because they exist online instead of in print is outdated and invalid. (Goodchild 2007)

Regardless of whether printed maps created by professional cartographers are more accurate than their online counterparts, the fact remains that with VGI there is no single person or organization accountable for the creation of the data. With so many people working behind the scenes and creating the information presented to the audience, there is even more room for error that does not really exist in the traditional map-making and map-printing industry for example thematic data accuracy. One could still argue though, that there are ways to judge the accuracy of these non-physical maps, just as there are ways to judge the accuracy of physical maps.

Using reference data to assess accuracy is one means for evaluating quality of GI, but if the reference data is unavailable, there is a diminished capacity for measuring quality effectively (Tomlinson 2003, 89). With VGI, the volume of data may act as a surrogate for high-quality reference data in some circumstances. One study used Twitter messages in order to acquire spatio-temporal data on forest fires, concluding that social media may provide promising sources from which to rapidly collect event-related spatial information (De Longueville, Smith and Luraschi 2009).

Likewise, Sheppard (2012) explains that “Twitter allows tweets to be geo-tagged, and a number of projects are examining Twitter as a potential source of crisis information.” Obviously, many researchers consider Twitter to be a useful and potentially valid source of geographical and disaster information. While Twitter and other social media and networking sites, such as Facebook, can be potentially utilized as untapped resources of sorts, one must bear in mind that there is, as is the case with OSM and similar technologies, immense capacity for error. (Sheppard 2012) If Twitter posts led to emergency resources being sent out to where they were not needed, a lot of money and time could be wasted in the process. Even more tragically, truly injured people who actually could have used the help being sent to a false alarm could suffer. Furthermore, there are possibilities of terrorism that could occur through the misuse of social media. Terrorists, for example, could hijack social media and use it to spread false information about where an attack was occurring, as a diversion to the real attack. There will also always be pranksters and conspirators who could waste emergency resources. As in other areas of VGI research, the question of whether a given data source can be trusted is of utmost importance (Poblet 2010).

Until there is a way to control and assess the validity of these social media posts, relying on them to provide useful and accurate information in times of crisis is extremely risky and not advisable, at least not until more thorough research has been done and methods are in place to control detrimental situations from arising. With social media playing such a large part in VGI recently, there have been some contrasting examples of VGI or citizen science which have existed for some time.

Wiersma (2010) examines the accuracy of the Cornell-based eBird project and the CBC in which citizen scientists are utilized in order to allow researchers to study long-term health of birds across North America. The CBC does not put any restriction on the background education of users who are able to contribute which can cause controversy with the quality of the VGI (Butcher et al. 1990). With such varying backgrounds in geography and mapping contributing to the creation of VGI, it is unlikely that all VGI from the same source has similar accuracy. The background experience level of citizen sensors were broken down into several categories by Wiersma, (2010) ranging from a neophyte who has no formal background in the subject but possesses the interest, time, and willingness to offer an opinion, to expert authorities, who are educated, studied and practiced in a subject to the point where they possess an established record of providing high quality products.

In addition to the varying degrees of education of data creators, users have access to different types of location sensors. Some data may come from companies using survey-grade GPS instruments, which have spatial accuracies down to the centimeter range, while other data might come from users with GPS enabled smart phones with accuracies in excess of 10 meters (Brando and Bucher 2010). Knowing how inaccurate the technology which was used to report VGI can inform the buffering analysis and clean-up of data which can be used to evaluate the information's accuracy and increased usability.

3. Methods

3.1 Methods overview

3.1.1 Metrics of Logical Consistency

Logical consistency referring back to Table 1, is looking into how the data relates to itself. For example, are all polygons closed or do they have dangling arcs. Checking for duplicates and missing information, ensuring that the format of the data is consistent. For the purpose of this thesis, items such as duplicates were embraced as they would be able to help provide a mean for the data that would otherwise not be available without other data to reference against. Many of the tests for conceptual consistency were not conducted as they would limit the variability associated with VGI that is needed in order to analyze the quality.

3.1.2 Metrics of Completeness

Completeness is analyzed by looked at the total number of nodes created when converting from linear shapefiles to points. Using Equation 1, it is possible to identify only the key nodes in most types of linear shape files by looking only at nodes which are either start or end nodes, also known as dangling arcs, or nodes with a turning angle which is greater than a certain threshold.

Equation 1 Turning angle formula

$$\theta = arccos\left(\frac{a \cdot b}{|a||b|}\right)$$

Using a threshold starting at twenty degrees a measure was conducted connecting the nodes creating a median line network to compare against both reference data and the original test data. The angle threshold was increased in five degree increments until a point at which there was a 90% similarity between the connected nodes and the original data. This point was determined to be at thirty five degrees. Once the threshold was determined, it was possible to continue and run the process on the full data set to compare and allow for a measure of completeness. The original data and the reference data both converted to points with a threshold of thirty five degrees to determine which nodes would be used, the total was then summed in order to compare the total number of nodes in both data sets.

3.1.3 Metrics of Positional Accuracy

The metrics for positional accuracy were created in the Python programing language using a variety of additional user created libraries including Numpy, Shapely, and Shapefile.py. Using these tools and a work flow which addresses the sub elements of data quality as outlined in ISO 19113 and ISO 19114 in Table 1, a Python script was created which will sort through a dataset and compare it to available reference data. The first step for the script is to take whatever Shapefile type the data comes in, whether it is point line or polygon, and convert it into a point file as this type of file is easier to compare against other datasets. Etherington (2011) explains how Python scripts can be useful when examining genetic and environmental data. Because ArcGIS is able to incorporate Python scripts into its map-making capabilities, it can be highly useful for giving researchers the "ability to quantify the connectivity of a landscape for a given species as represented in the genetics of a population is of great interest to ecologists concerned with a wide range of issues such as habitat fragmentation, invasive species, or wildlife diseases" (Etherington 2011).

This is especially applicable in case study 3, which examines the populations of different birds as reported by the public. The concept and processes of creating buffers are important to understanding and digesting map data (Couch 2011). Tomlinson (2003, 216) defines buffers as "the ability to generate zones of specified width around point, line, or area features," when these areas are generated around "point and area features," they are called buffers, while "zones of interest around line feature may be called…corridors." Figure 4 and Figure 5 demonstrate a line buffer, or corridor, and a point buffer, respectively.

Figure 4 Line Buffer (Couch 2011)

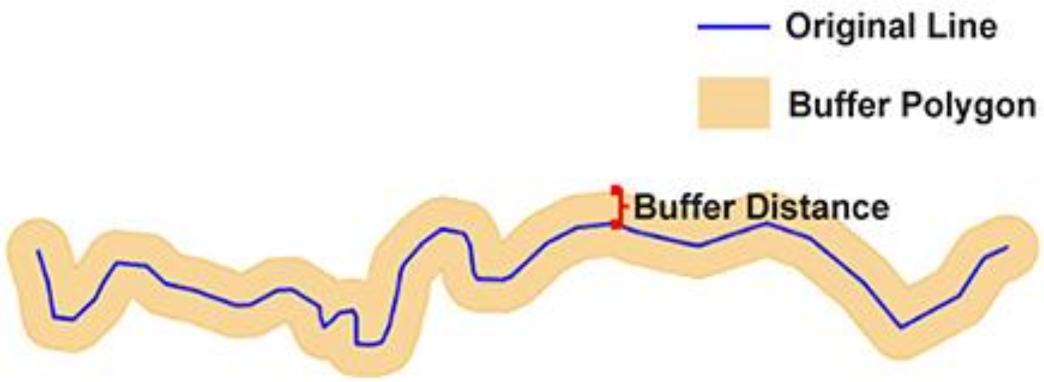
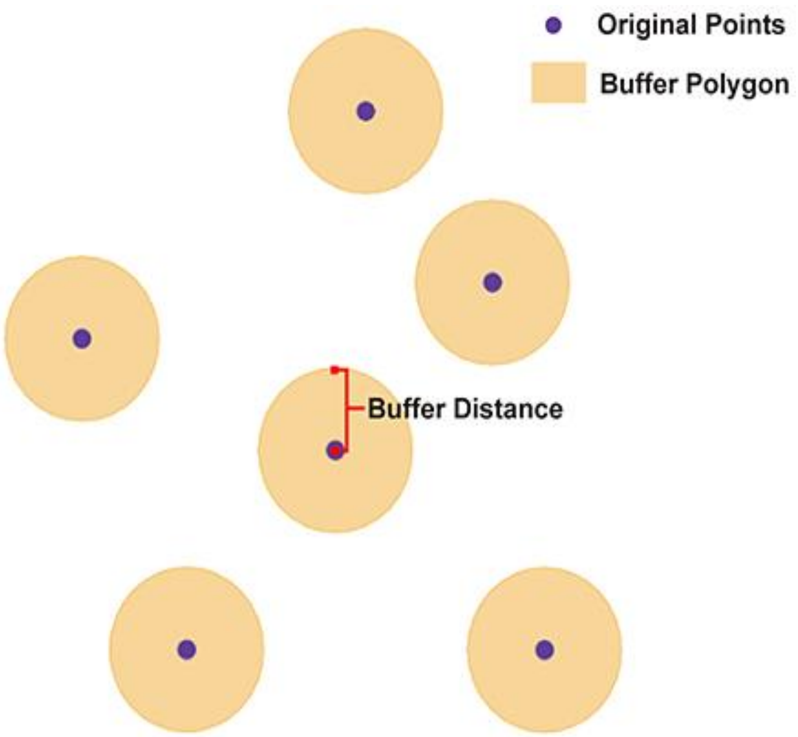


Figure 5 Point Buffers (Couch 2011)



Buffer analysis can be used to compare two sets of geographic data, one user-generated and one created officially, by mathematically determining the amount of overlap between the lines, points, or buffer zones. Tomlinson (2003, 81) also provides a useful chart Table 3 for determining the possible positional accuracy of different scales and percentage of likely errors at different map scales, meaning that maps with a larger scale can be assumed to contain less error than a comparable map with a smaller scale. Combining the general error rate and the discrepancies revealed through buffer analysis can lead to an effective evaluation of VGI quality.

Table 3 Expected Error (Tomlinson 2003, 81)

Map scale for a given area and error tolerance	
minimum area (ha)	% error in area measurement

	1	3	5	8	10
0.01	1:100	1:300	1:500	1:800	1:1000
0.1	1:300	1:900	1:1500	1:2400	1:300
1	1:1000	1:3000	1:500	1:8000	1:10000
10	1:300	1:9000	1:15000	1:24000	1:30000
100	1:10000	1:30000	1:50000	1:80000	1:100000
1000	1:30000	1:90000	1:150000	1:240000	1:300000

Map scale for a given area and error tolerance					
	% error in area measurement				
minimum area (ha)	1:1000	1:5000	1:10000	1:50000	1:100000
0.01	10.0	50.0		Invalid	
0.1	3.3	16.6	33.3		
1	1.0	5.0	10.0	50.0	
10		1.6	3.3	16.6	33.3
100	Insignificant		1.0	5.0	10.0
1000				1.6	3.3

Von Lindenburg (2014) discusses the use of calculus to determine distance given GPS data, namely the Haversine and Trapezoidal rule methods, and compares the results of these calculations to those obtained by a smart phone application designed to use GPS to track trail length. Von Lindenburg found that the smart phone likely uses a method similar to the Haversine formula for determining the distance based on the latitude and longitude of points located on a sphere (von Lindenburg 2014). The complicated nature of VGI accuracy is accounted for when the author states that both methods yield relevant and concordant distance information, "in good agreement with the distance given by smartphone application, considering the amount of variables that can affect the accuracy of the data" (von Lindenburg 2014).

3.1.4 Metrics of Temporal Accuracy

Temporal accuracy was measured in the ISO standard by ensuring that there was a logical linear order followed so that date X would be before date X+1. The data sets used in this thesis did have date data associated with them that would lend itself well the analysis. The analysis of the date for correctness was not applicable as the date did not provide any bearing on the information that the data was purveying. While Haklay et al. (2010**)** suggest a linear relationship between the decline in accuracy of the data and the age of data, the data provided for all three case studies was the most recent available; an analysis of historical data may be something that would be cause for future study.

3.1.5 Metrics of Thematic Accuracy

In order to measure thematic accuracy, the ISO standard suggests several possible causes for disagreement between data thematically. For the purpose of this study, the possibility of unique or outlying data was determined as the greatest possibility for error as illustrated by (Raggatt 1989). Due to this a measure, of the number of unique values and unique contributors were used as criteria in an evaluation of possible causes of error. A statistical approach using R was adopted in order to sift through the large data tables provided for this measure of accuracy.

In the following case studies, the ISO standards will form the backbone of the quality analysis, showing the ways that ISO standardization efforts can be applied to VGI, as well as official GI. OpenStreetMap is an open source alternative to Google Maps and Earth. Users are able to submit their own information concerning the layout of roads.

Case study 1 OpenStreetMap discusses the accuracy and quality of OpenStreetMap data and information. Case study 2 looks at bike trail maps. Finally, case study 3 provides a unique analysis of bird observations.

3.2 Case Studies

3.2.1 Open Street Map

OpenStreetMap (OSM) was selected as a case study because it is hailed as the largest user editable map of the world. With millions of users contributing data worldwide, this data source has gained notoriety and many people have called in to question the data quality (Hackley, 2010, Coleman 2013). The data provided by OSM was mainly analyzed for completeness and positional accuracy. It was compared against the survey data provided by the Canadian Government in The National Road Network.

3.2.2 Trailfu.com

The trailfu.com dataset is very similar to the OpenStreetMap data in regards to the fact that it is user created and provided freely on the internet for anyone to download. In addition, it is also a dataset that was analyzed for positional accuracy and completeness as the downloaded data contained no thematic information such as trail names or landmarks. The most important difference between trailfu.com data and that of OSM is that there is no reference data available for the particular area that was chosen. There was, however, many trails that varied slightly of the same area. This allowed for a slightly different analysis to be conducted using the mean data as a reference set.

3.2.3 Bird Atlas of Canada

The final data set was one that relied less on positional accuracy and instead looked at the thematic accuracy of the data. The data from the Bird Atlas of Canada in recorded as observations in a survey block. While there is an element of positional accuracy, the focus of the data was the thematic accuracy which was the quantity and quality of the observations of birds. Like the trailfu.com data there was no reference data in order to measure the observations against.

4. Case study analysis

4.1 Case Study 1 OpenStreetMap

4.1.1 Study Area

The OSM data used comprises the majority of the City of Waterloo, Ontario. This area was chosen as numerous users actively contribute to the mapping information for this area and I am familiar with the area. This ensures that a sufficiently large dataset of volunteered data is available which will provide a good base for analysis. In addition to providing more points of data, the study area is critical because space and place are two conceptions of location, the former formal, in terms of mathematic precision carried out by select experts, and the latter subtle and ambiguous because it is infused with human meaning (Roche and Feick 2012). The citizens of Waterloo, Ontario provide a source of information that emerges from public participation, resulting in more ample data.

Compared to other forms of crowdsourcing, OSM is limited by the location of its volunteers (Haklay 2009). Waterloo is a relatively urban area, and this is important because OSM is thought to have high accuracy in these contexts. If there are more than 15 users per square kilometer, OSM has been found to be highly accurate (Haklay et al. 2010). According to Linus' Law of software development, named after Linux creator Linus Torvalds, given enough eyeballs, all bugs are shallow, or in other words, a sufficient number of volunteers increase the ability to resolve errors (Haklay 2009).

Studies have used equations to compare polygon sizes between official and public sources of GIS data (Mooney, Corcoran and Winstanley 2010). The connection between the number of contributors and the quality of the data is non-linear; once there are fifteen contributors per square kilometer or greater, the accuracy increases and the error is often less than 6m (Haklay et al. 2010, p. 321). There appears to be a minimum threshold of observers required to gather the majority of VGI data points, as the first five contributors to an area seem to provide the biggest contribution in terms of positional accuracy improvement (Haklay et al. , p. 321).

Waterloo has a population of 98,780 and an area of 64.10 square km, resulting in an overall population density of 1520.7 per square kilometer (Statistics Canada 2014). By comparison in the US, 35% of all internet users have uploaded content, 26-35% data that they created themselves (Flanagin and Metzger 2008). If only 1% of all residents in Waterloo participated in publicly sourced mapping efforts, there would be 15.2 users per square kilometer, which is an acceptable rate to attain accuracy according to existing research (Haklay et al. 2010). Canada's national mapping agency, the Centre for

Topographic Information is well established, however, national and regional mapping agencies are increasingly reaping benefit from VGI (Devillers, Begin and Vandecasteele 2012). The region of Waterloo has its own GIS web portal, which allows users to view interactive maps of the Waterloo area with multiple informative layers (Region of Waterloo 2014). The government also provides a highly detailed official street map. The map depicts provincial, regional, municipal, private, and proposed roads, water features, drainages, mainline and spurs railways, as well as settlement and municipal boundaries (Region of Waterloo 2014). Certain aspects of this information would simply be unavailable or harder to ascertain through VGI such as the proposed roads, or the distinction between zones. Thus, due to its urban and populous nature, Waterloo is an excellent location for a case study examining the effectiveness, and potential inaccuracy, of VGI, particularly for mapping efforts like OSM.

4.1.2 Methodology

Data Preprocessing

There was very little pre-processing required for this dataset as the formatting of the data is consistent and mostly complete, with few spurious entries. The OSM data was downloaded in an OSM XML format but was easily converted to a shapefile using the ArcGIS editor for OpenStreetMap (ESRI 2014). Line and polygon shape files were converted to points by taking the vertex of either of the line segments and encoding that vertex to a point. The points which made up each line segment were given a number which corresponded to the original line. For example, if a shapefile has five lines within it, then each point created from each line will be assigned numbers zero through four, according to which line it derived from. Any attribute data from the parent file will be associated with all the points that were a derivative from that line. This process was the most efficient way to normalize the data before any metrics were used to analyze the data and would allow data to come from multiple sources in multiple formats.

Once the data has been normalized into shapefile points, the shapefile is read into an array that allows for the data to be stored as positional data in addition to attribute data. This is done for both the reference and test data.

Quality Metrics

The flow chart (Additional measures of accuracy were used for line shapefiles by creating equidistance buffers around a reference data set using the approach described by Hunter and Goodchild (1993). Figure 7 shows an example of how this was done ensuring that there is no overlap in line length by creating buffers that are hollow so that the line segments would only be clipped to one single buffer. Figure 8 is this buffer process being used on a reference dataset within the Waterloo study area. A buffer at 1 metre intervals was created and, subsequently, the total length of the test data was measured to see how much falls within each buffer length. Figure 9 shows the culmination of this analysis illustrating a trend which shows the majority of the data falling within the buffers closer to the reference data.

Figure 6a/6b) shows the process which the data then follows to come out with an assessment of the quality in relation to the reference data.

Additional measures of accuracy were used for line shapefiles by creating equidistance buffers around a reference data set using the approach described by Hunter and Goodchild (1993). Figure 7 shows an example of how this was done ensuring that there is no overlap in line length by creating buffers that are hollow so that the line segments would only be clipped to one single buffer. Figure 8 is this buffer process being used on a reference dataset within the Waterloo study area. A buffer at 1 metre intervals was created and, subsequently, the total length of the test data was measured to see how much falls within each buffer length. Figure 9 shows the culmination of this analysis illustrating a trend which shows the majority of the data falling within the buffers closer to the reference data.

Figure 6a Process Flow Chart

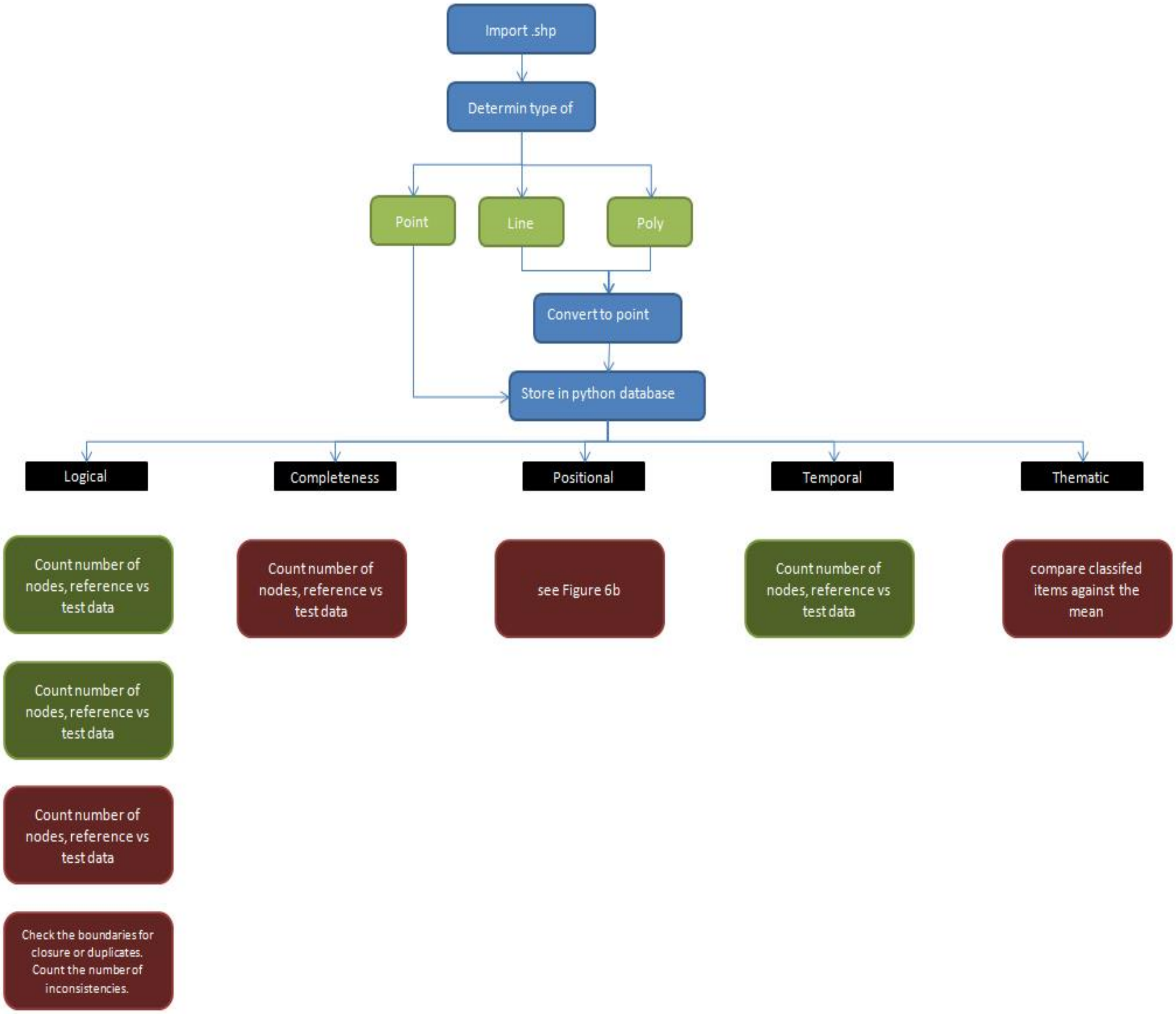

Import .shp
Determin type of
Point
Line
Poly
Convert to point
Store in python database
Logical
Completeness
Positional
Temporal
Thematic
Count number of nodes, reference vs test data
Count number of nodes, reference vs test data
see Figure 6b
Count number of nodes, reference vs test data
compare classifed items against the mean
Count number of nodes, reference vs test data
Count number of nodes, reference vs test data
Check the boundaries for closure or duplicates. Count the number of inconsistencies.

Figure 6b Positional Accuracy Flow Chart

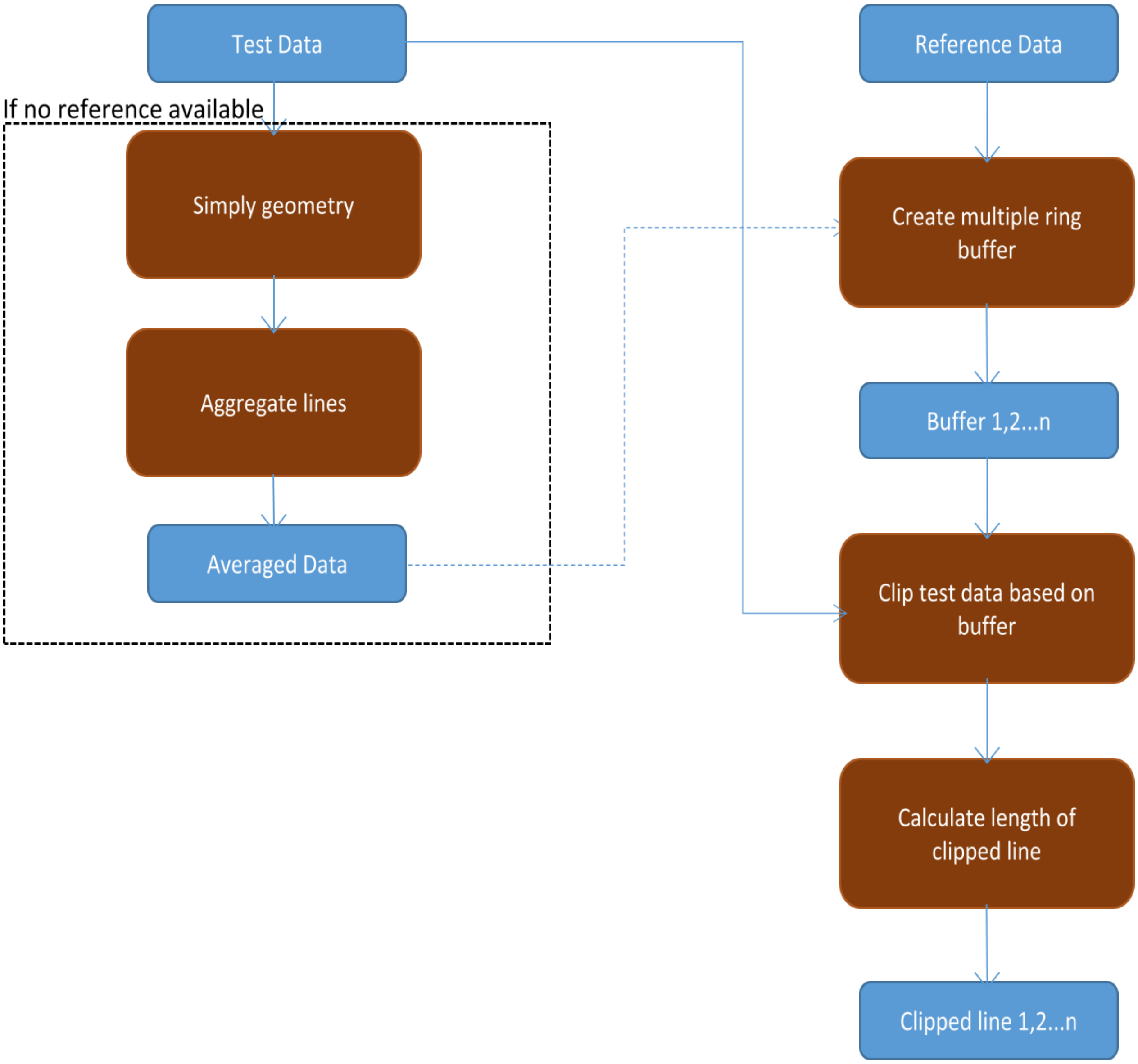


Figure 7 Buffer Process for Estimating Accuracy of Line Data.

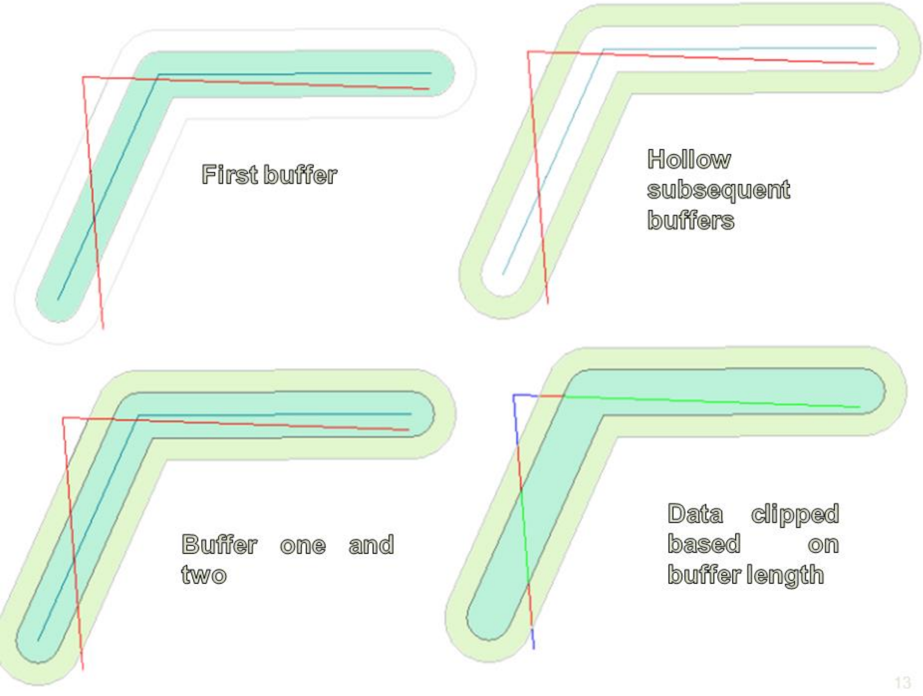
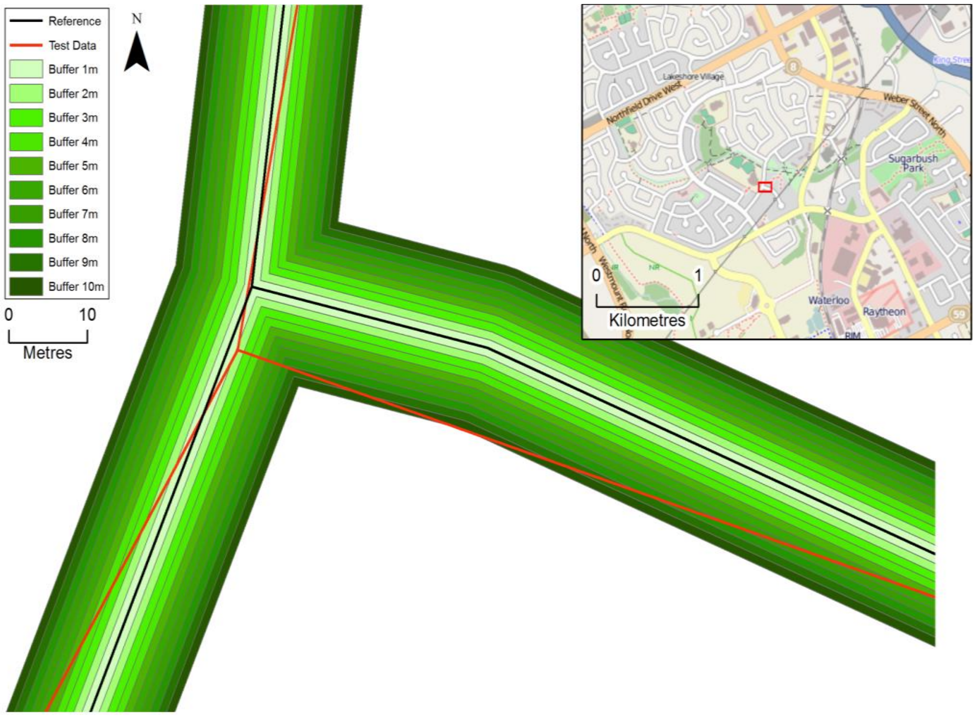


Figure 8 Practical application of buffer to OSM dataset.



In order to measure completeness, a count of the created points is taken and compared against the points created in the reference data. Assessing completeness is important because it provides an understanding of the reliability of the reported results and allows assessment of the usefulness of contributed data as a potential data source for use by mapping agencies and researchers (Jackson et al. 2013 24).

The thematic accuracy of the attribute data is more difficult to measure. For this sub-element the means to measure the accuracy is to create a percentage of correctly classified attributes. This percentage is calculated using a Python routine which tests to see if any attributes within the array of test data are also present in the reference data. The result of which is a count of which elements are unique to either one array or the other. A subsequent command can be optionally executed in which the program will list which elements are unique. The thematic accuracy is not dependent on the format of the data, thus, all data which is being checked using the routines is normalized to exclude possible outliers as a result of incorrectly formatted data which can be something as simple as capitalizing inconsistently.

A secondary routine, which is very similar to the thematic accuracy routine, is used to check logical consistency in which format is important, and thus any data fields in which the data is not formatted exactly like that of the reference data will be counted and will then be added to the total quality score of the data this process was found to be highly erroneous as the difference in the way the data sources formatted their attribute data.

The methods tested on the OSM data were less complicated to implement as there was reference data available, so metrics that were being tested could be evaluated more easily for their functionality. The buffer method outlined earlier was deemed most suitable and similar to previous studies on OSM accuracy.

4.1.3 Results

Positional accuracy results of roads in the OSM data using the line-buffering method is presented in Figure 9. Thematic accuracy analysis found that 28.5% of the line segments in the waterloo area fall within the one metre buffer around the reference data with the percentage falling quickly to 19.8% around the two metre buffer. On the other, end only 1.8% of the data fell within ten metre buffer meaning that the majority (64%) of all tested data fell within 3 meters of difference between the reference data and that of OSM.

Figure 9 Buffer Analysis Results for OSM, Length Of Line Segment Within Each Buffer Distance

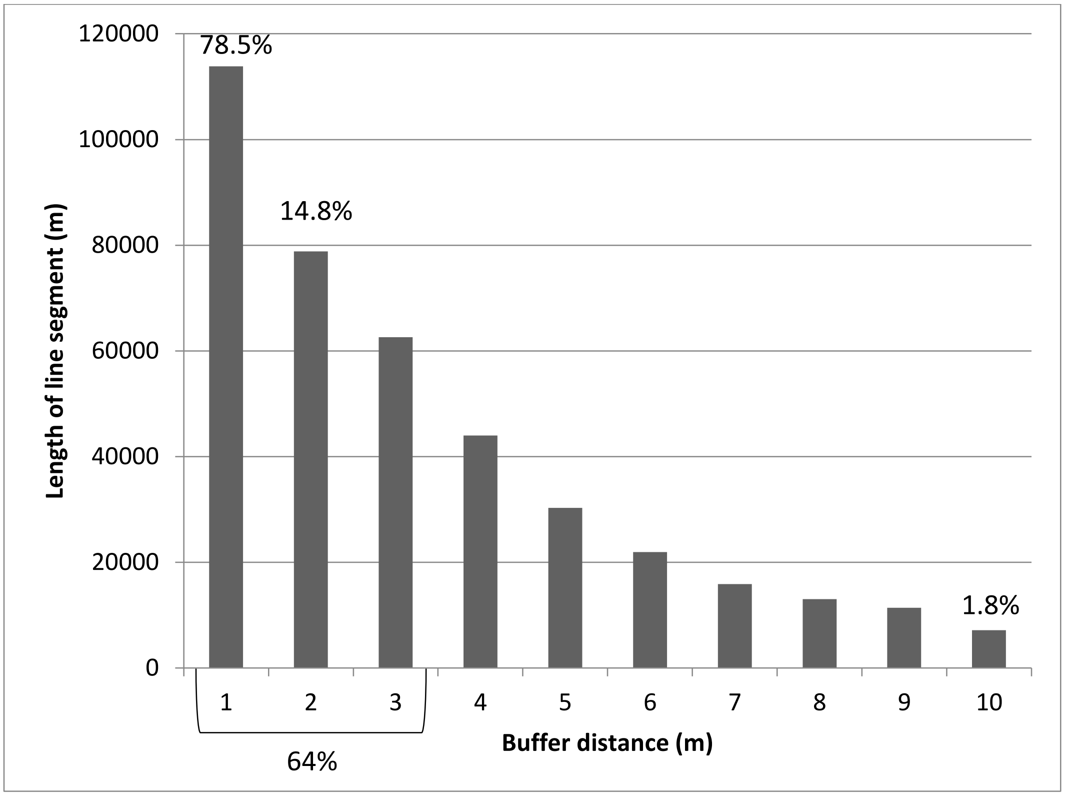


Figure 10 Smaller Details within OSM Map

Possible sources of error with the OSM data compared to the survey data is in the quantity and how objects are visualized and digitized. In the subset of OSM data used for analysis, there is a total length of lines of 404,808m while the OSM data for the same area is 652,159m, significantly more. This is due in part to what OSM considers roads. What can be seen in Figure 10, as red lines, which are representative of the OSM data, shows many private roadways and services roads. These are not included in the calculation for total length of OSM data. Conversely the blue lines, referring to the federal survey data, either generalize or do not include these roads at all. Another source of error in the length of lines is the way in which roads (roundabouts in particular for this situation) are digitized. An example of this concern is seen in Figure 11. where there are two OSM roads for each reference data road, the result of which leads to a drastic increase in road length in these areas.

The positional accuracy for these two particular examples in question provides a robust measure for positional accuracy which can be scaled according to the demands of the researcher. The manner in which the roads are digitized in the OSM data shows a much truer example of the actual roads, where each lane on a divided road in the OSM data gets its own individual line as opposed the survey data, where the center line of the road is mapped and deemed a suitable representation for the purpose of creating government road maps of the area. The temporal accuracy of the data is not always possible to assess. With the OSM data, there is a time stamp created when the data is created. For the sub elements of temporal accuracy, the accuracy of the time measurements cannot be compared against the reference data to check its accuracy. If the

Figure 11 Difference in Data Creation

roads are actually digitized when they say they are, however, the quantity of roads being included at a particular date will show how up-to-date the VGD is.

These findings demonstrate that new roads and altered roads are frequently updated more quickly through VGI than through traditional means of mapping road networks. This lends credence to the idea that the use and accuracy of VGD depends on the manner in which it is transformed into VGI. OSM also combines many different types of data and uses not only GPS tracks and points, but also free data from other sources such as aerial photography, Landsat 7, topologically integrated geographic encoding, and referenced data used by the Census Bureau. All these different forms of data add to accuracy of OSM because some of the data is already geographically referenced and can be used as reference data for the user generated content for legitimacy. This relates to the ISO standards quality element of completeness, and allows comparisons which regard the facets of the other standards temporality.

Most road networks are updated as soon as the roads' plans are created. One such example of this is shown in Figure 12 and Figure 13, where the imagery is from June 2007, and the completed Highway 26 finished in September 2012. This road was created from VGD by user *cartographer 29* who has been contributing to Google Map Maker (a similar project to OSM) for almost one year and has 1260 total contributions and a 94% success rate for approved edits. It is important to note, however, that though 94% is a fairly high success rate; it is still not perfect, further emphasizing the possible danger of relying on information that has been submitted by non-professionals within the cartography industry. Considering the idea that even a small error rate in an individual submitter of VGI could aggregate among a population of users and lead to larger mapping errors, it is important to thoroughly validate the data used in order to resolve this issue.

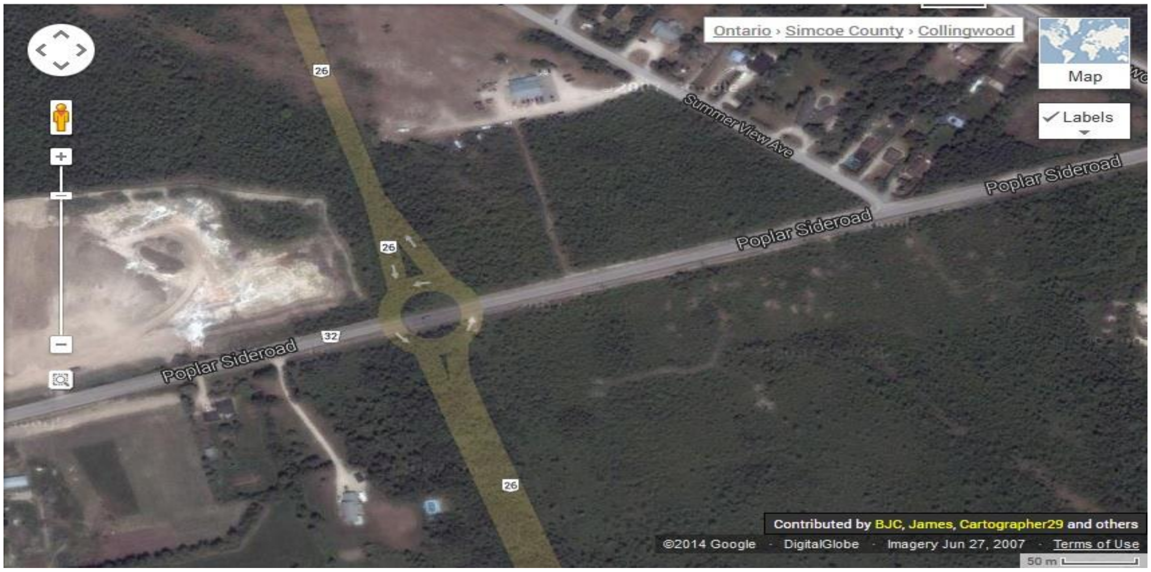
The same issues apply to the temporal consistency and validity. Satellite and aerial imagery is inevitably captured at different times, and stitched together through various methods in order to create a cohesive informational product. Additionally, as there are many users creating the data at any time, there is no consistency to the data. When a lot of inconsistent data is coming in, this creates a further problem for researchers and editors, who must then spend time differentiating between the correct and incorrect information submitted by a variety of users. In that sense, they must spend a great deal of time piecing together factual answers, something that again takes away from real work that could be being completed. The temporal accuracy data quality element would be more critical in data where the data is dependent on the temporal aspect, such as transit schedules or package tracking where a measure of the difference between the actual events can be associated with that of the recorded events. These events can then be analyzed to see if they appear in order and are precise. Figure 12 shows Google Map Maker's imagery and the contributions made by the users *BJC, James,* and *Cartographer29*, who have interesting metadata attached to their handles: member since Feb 2011 contributing 28380 total edits with 99% approved, member since May 2013 contributing 382 edits with 83% approved, and member since Jun 2012 contributing 1278 total edits and 100% approved, respectively. This shows how the temporal and completeness aspects of the relevant ISO standards can also be applied to VGD, and not just fully ISO-compliant expert provided data. By giving information about the history of each user and his or her edits, the ability to evaluate the data for quality and accuracy is greatly improved.

OSM would appear to validate the results of the past twenty years of utilizing VGI as it depicts how citizens can increasingly direct how the technology and data are applied to issues of local interest, for example, the streets they drive, walk, and inhabit (Roche and Feick 2012). Studies have found that OSM volunteers can attain competitive accuracy compared to proprietary data. In the UK, for example, users achieved 88% for A-Roads and 77% for B-Roads (Haklay 2009). Canada has achieved similar results with OSM mapping, particularly in urban areas (Tenney 2014).

OSM is a promising repository of user generated data that is currently being used by individuals, governments, and businesses in order to locate and understand geographic spaces and human-centered places. The lack of comprehensive metadata is one weakness in the OSM framework, but it can be overcome by emphasizing the usefulness of the available dataset, rather than sifting through metadata trying to ascertain accuracy.

Figure 12 Road Imagery of Highway 26 in June 2007

Figure 13 Google Map Maker Imagery of Highway 26 Roundabout



4.2 Case Study 2 Mountain Bike Trails

4.2.1 Study Area

The study area contains trail network data from a location just to the south of Collingwood, Ontario, in Pretty River Provincial Park. This data has no corresponding reference data available and appears to be quite disjointed. For spatial accuracy this data will provide a worst case scenario for creating an evaluation of quality with no reference data. This area is of particular interest as substantial a priori knowledge of the study site is available. Bike trails form an important method for humans that traverse the urban and rural landscapes, and "digital data repositories of bicycling related information in a transportation network, such as on-street bicycle lanes, off-street trails, lane width, or traffic volume, provide an important basis for a variety of bicycle transportation planning and analysis tasks, including latent demand analysis" (Hochmair, Zielstra, and Neis 2013 3). In other words, if highly accurate trail information can be obtained through VGI, it would improve the city and park system, and the ability to plan new pathways that cater to the local community's needs.

4.2.2 Software and Data Sources

The KML and GPX XML files with the track, waypoint, and attribute information were downloaded from trailfu.com and ArcGIS was used to further visualize the available data by importing these files as layers. Base maps provided by the South Western Ontario Orthography Project (SWOOP) provided high detail aerial images for contrast that data against.

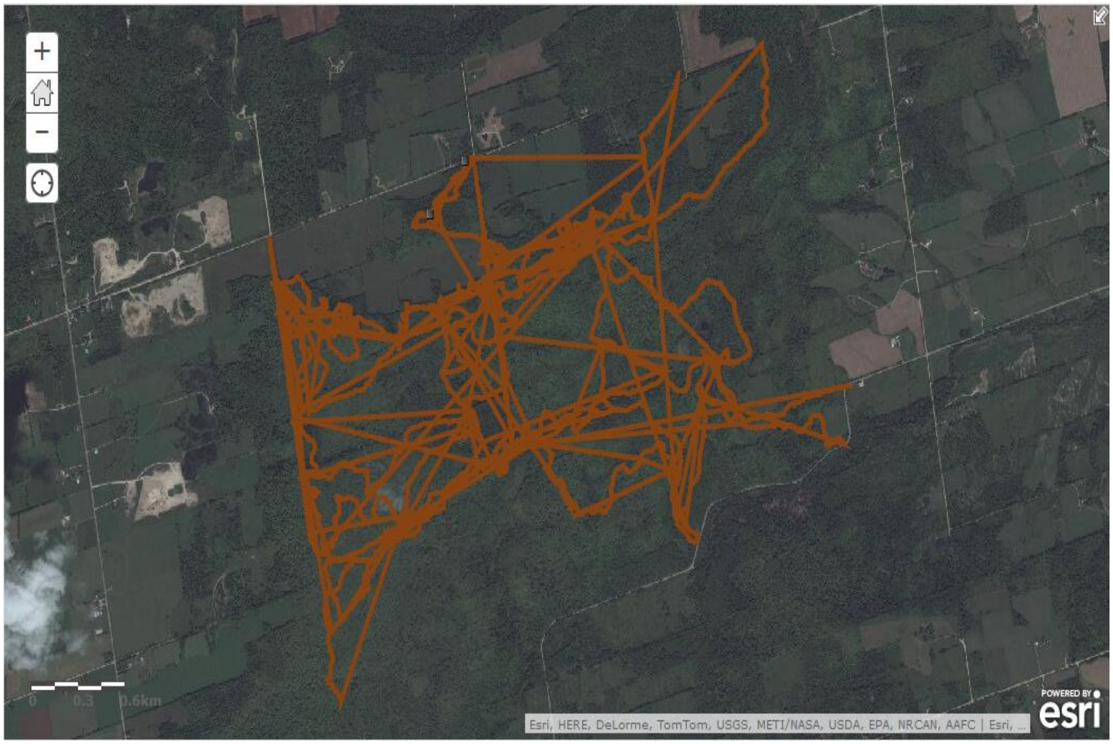
4.2.3 Methods

The different layers provided by importing the data into ArcGIS were visualized with underlying satellite data, as well as a basic buffer zone analysis which would allow for future comparisons with other VGD pertaining to this specific trail. As there are many bikers in Ontario, this information could be easily acquired by using a GPS tracker which can log GPX files.

Moreover, cyclists believe that they collectively have better knowledge necessary to build GIS databases, than any other group does, and that, as individuals, each have unique information which no one else can contribute, when it comes to GIS solutions that cater to biking enthusiasts (Priedhorsky 2007).

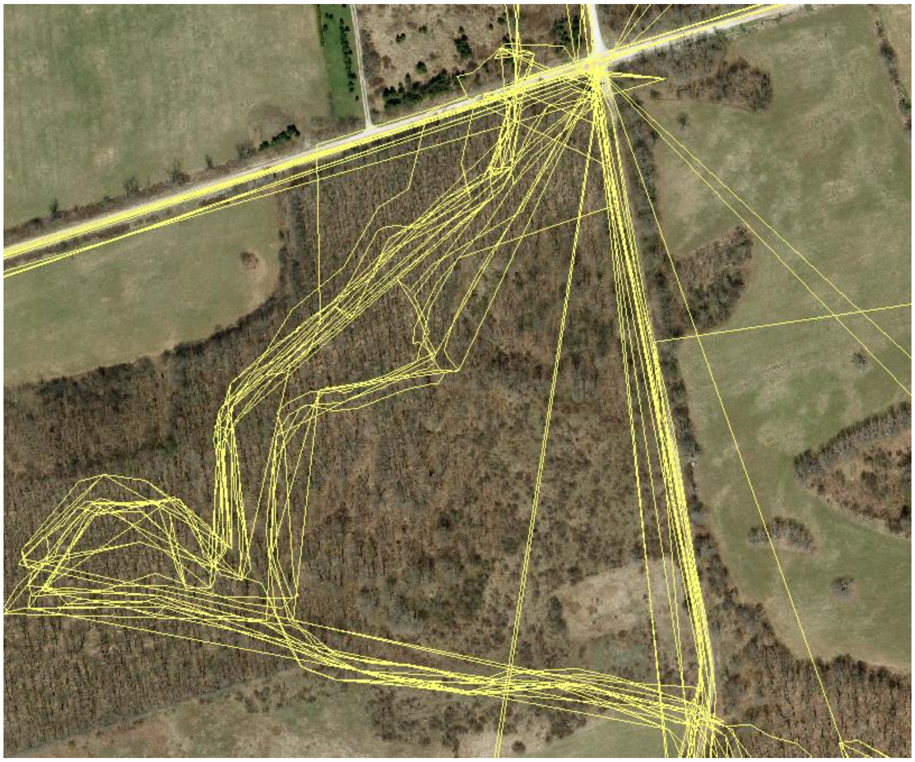
A number of figures were prepared which demonstrate how VGD can be accurate for making and following trails, even when the possibility of individual erroneous points is considered. Figure 14 shows the trail track map overlaid on an aerial photo base map, created solely for this analysis, which combines two forms of geographical data, one unofficial, and one official, respectively, in order to determine the accuracy of the user provided data. The track built from the GPS data provided by a user of trailfu.com is coloured Brown.

Figure 14 Trail Track Map Overlaid with Satellite Imagery



A small subset of this trail is loaded into Figure 15 which shows how many tracks can be overlaid for one single route. The road visible going north south on the right side shows just the viability that GPS devices can record a single track under optimal conditions.

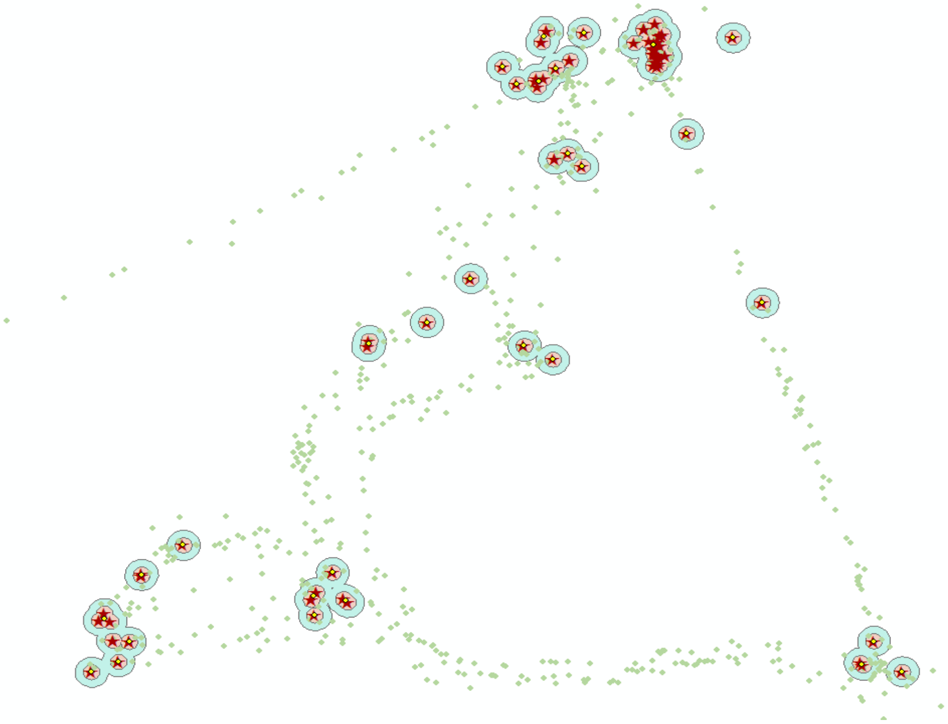
Figure 15 Subsection of Line Data From Trailfu.Com



In Figure 15, the track map is shown along with a SWOOP satellite basemap which also shows major roads. These individual tracks are refined into a track algorithmically, relying on a buffer system which can introduce error. Additionally, GPS units may malfunction in heavily wooded or downtown areas and report spurious locations along the track due to multipath errors caused by the GPS signal bouncing off objects as it come close to the receiver. Figure 16 shows how this algorithm works in order to create reference data to compare. After the line data has been converted to points, key points are found using the turning angle using Equation 1.

If the angle is above a threshold of thirty five degrees then this point is selected as a key point, and a subsequent buffer is place around these points which eliminate the possibility of many key points in a diminutive area. Using the centroid of the buffers seen in Figure 16 new lines are created to act as reference data for the trailfu.com data.

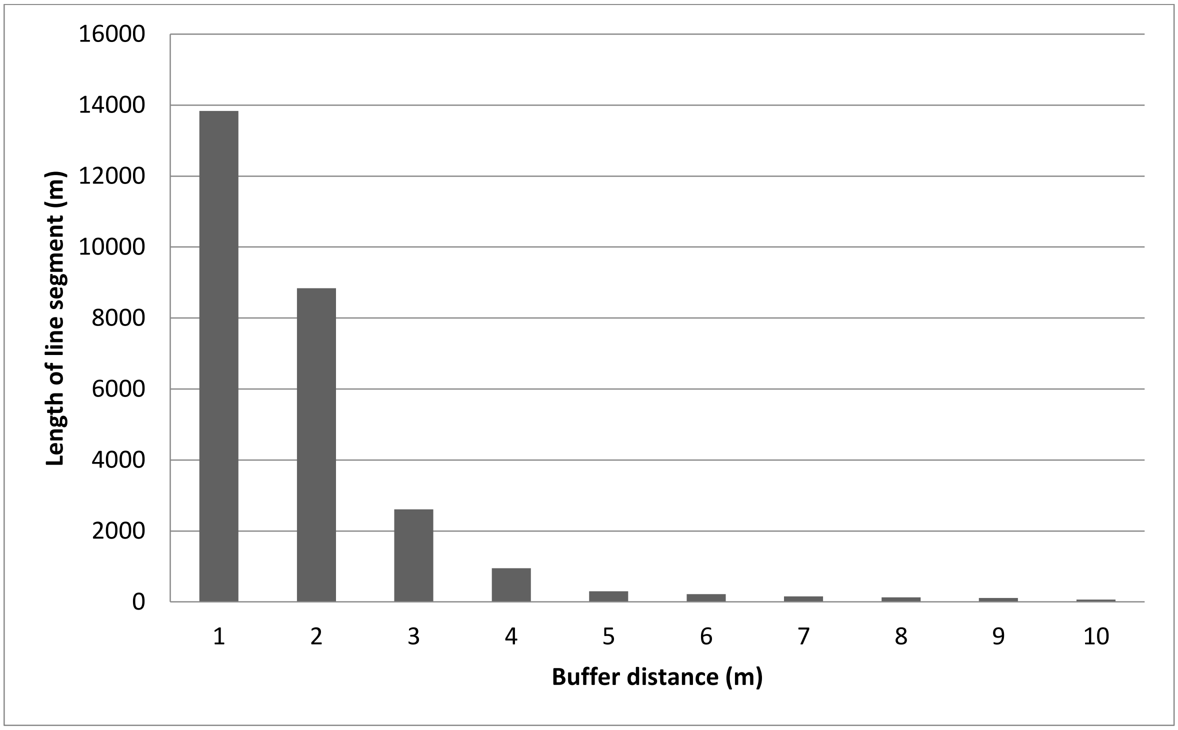
Figure 16 Key Points and Buffers of Trailfu.Com Data



From this point, the same process that was applied to the Waterloo OSM data was carried out on the OSM data with nearly identical results shown for the error away from the reference. Figure 17 shows how, even with many lines making the data appear to be skewed, there are still even more data lines which are in agreement with each other and show that with enough data, the majority will fall within the margin for error of this data type.

4.2.4 Results

Figure 17 Buffer analysis results for Trailfu.com



Future research should focus on ways of increasing the quality or ability to evaluate accuracy of such user provided data in the absence of official datasets to compare it with. Mountain bike trails are an opportunity for user-based mapping phenomena to be implemented in rural areas, which is one of the main alleged problem areas of VGI software like OSM. Trail maps are one way that social networking can occur in a community, whereby users alert each other to new trails, or alternate ways to navigate the city (Coleman et al. 2009). Paths made up of ample waypoints are accurate and smooth, because incorrect or aberrant points can be removed. Priedhorsky (2007 1) uses the term geowiki to describe a “system to enable cyclists to collaboratively build a database of geographic information relative to them.” Trailfu.com has a similar goal, which is to transform user interest into VGI, and vice versa. Trail data such as this is also available via OSM, and because small paths may be less attractive or easy to measure for large organizations or national mapping agencies, VGI may form a potent source of information. For example, when considering the entire German OSM street network, including pedestrian paths and small trails, the VGI data source exceeded the proprietary information by 27% in distance, showing that members of the public may provide a surplus of information in this area (Neis and Zielstra 2014). The use of informal sources of data to identify trails is not a new practice, with early hiking trails in America and Europe dating to the 18th century (Brown, 2006). However, tourist bureaus may be reticent to rely on VGI to produce official maps, as they are liable for inaccurate or incomplete information (Klien, Fitzner, and Maué 2007). Even if the author has walked the trails by himself and might have put much effort into digitizing the route, without professional credentials, the data may not be trusted for official use (Maué, 2007, p. 1). Maué (2007) contends that the reputation of the volunteer, which can be used as an overall measure of reliability, local knowledge, skill, and morality, can help to assess the accuracy and validity of their supplied data.

Bégin, Devillers and Roche (2013) explain how certain types of users are more likely to contribute trail data, rather than map data. This demonstrates how users may self-select for the type of information that they provide, with those who use the information, such as hikers, trail bikers, and nature enthusiasts, being the most likely people to provide the information. Amateur interest in natural features that an area has to offer thus informs the motivation of the users who provide VGI to a GI repository, like trailfu.com. By focusing on the local area, this study explores how users may have pet features such as trails, or pet locations that are preferred (Bégin, Devillers and Roche, 2013).

This case study utilizes aerial data to supplement VGI, which is a popular way of enhancing the usefulness of scant volunteered information in rural areas (Devillers, Bégin, and Vandecasteele, 2012). The combination of citizen sensors with the technological improvements in satellite unmanned air vehicle imagery will enable VGI to assist in the mapping of "all populated areas," whether urban or rural, within the next ten years (Devillers, Bégin, and Vandecasteele, 2012, p. 2). Government agencies, such as parks bureaus, could use VGI in order to improve services (ESRI, 2010). VGI forms one point of data that can be managed with a GIS, and can supplement traditional sources of GIS in important ways, particularly in areas that deserve public attention and involvement, such as parks (ESRI, 2010).

Some enterprising users of Google Maps have developed clever ways of using the API, which blur the lines between these case studies (Gibson and Erle, 2006). For example, the website *Bay Area Hiking Trails* allows users to "record [their] own specialized data: birds seen, bikers met," or geocaching activites (Gibson and Erle, 2006, p. 130). There are also proposed uses such as a service to identify invasive plant species or a slightly specialized social mapping application to document the rockiness, steepness, or exposure of certain trails (Gibson and Erle, 2006, p. 131).

The potential for mapping areas such as urban forests, as in this case study, is amplified by the use of VGI (Foster and Dunham, 2014). Google is currently testing a hybrid approach that provides volunteers such as tour guides and trail enthusiasts, with professional mapping equipment (McAvoy, 2013). By empowering trail guides with high-tech backpacks equipped with Google Street View style technology, this new trail-mapping effort will allow users to view 360 degrees of a trail's pathway (McAvoy, 2013).

New technologies, such as the upcoming 3D tablets from Amazon and Google, promise increasing capabilities for people to scan and catalog the world, including the use of fine geographic data. It appears that combining multiple sources of GI is the best way to compare the ISO standards' components of accuracy, such as completeness and reliability. However, ultimately, every measurement, whether generated by technology, experts, enthusiasts, or average consumers, includes an element of human bias, despite methods of mitigating error and striving for objectivity.

The creation of novel GI systems that cater to bicycle enthusiasts presents some challenges, for example, "bicycle navigation systems are particularly difficult to automate," as "there are strong personal preferences with regard to topography, traffic, distance, and other factors" (Raubal, Mark and Frank, 2013, p. 146). Perhaps VGI can provide software producers with the information needed to create products that serve bicycle riders' unique interests and concerns. This concept links all three case studies: the usefulness of VGI is predicated on its usability, and the reciprocal nature of its creation leads to GIS products that cater to increasingly specific requirements. Thus VGI is a positive development for a number of interests, with trail enthusiasts benefitting from the ability to share and compare their findings; provide updates on trails that have been closed or altered by natural or manmade changes; and enjoy their hobby more thoroughly by logging progress and competing with other bicyclists and hikers. Mapmakers should utilize the natural predilection for humans to understand their local world to the fullest extent by incorporating VGI with official data to amplify and verify both datasets of geographic information.

4.3 Case Study 3 Bird Atlas

4.3.1 Study Area

The study area for this project consists of the entire province of Ontario. The data is subdivided by UTM zones of which there are four in Ontario, UTM zone 15 through 18. The study area is further divided into 4912 ten kilometer squared grid cells. Each of these cells is denoted in the data by a centroid coordinate which causes for some non-square cells in areas around the edges of the UTM zone. By dividing the area into equal portions, a saturation map can be created using ArcGIS which will help to visualize the data collected more effectively. Referring to Table 3, the expected error can be computed through an estimation of error percentage (Tomlinson 2003, 81).

4.3.2 Software and Data Sources

ArcGIS is the primary software package used in order to analyze this data and to create the maps and provide grid cells into which the data table is joined. ArcGIS proved inefficient when dealing with complex tabular databases and Microsoft Excel, Access and R were each used in an attempt to analyze the data further. Davis provides an example of bird migration tracking using GIS data, which demonstrates how useful maps can be for interpreting large amounts of data (Davis 1996, 2006).

There were some limitations when using both of these software packages as Excel only allows for 255 records to be graphed along the y axis. Furthermore, the data also proved to be too large for access to query and deliver the products needed. Finally, R and Python scripting were used in order to deal with the massive dataset that was being used. R was used to create visualizations of the data in graph form, while Python was used to query the data to output the new variables that were needed.

The data itself was downloaded as a single .txt file that was approximately 500 megabytes. This data had to be converted to a .csv file for use in R and Python, and the final form of the .csv file had 541,743 records with each record having 172 variables. This lead to over 9 million data entries which was the cause for most of the complications associated with processing the data. The data was provided by The Atlas of Breeding Birds of Canada and had to be requested through their organization.

4.3.3 Methodology

The data was provided courtesy of The Atlas of Breeding Birds of Ontario, in order to compare records to the province at large, at least 15 point per square are needed (Parfitt 2013). The records contained roughly half a million entries, with each entry containing spatial information, the x, y, and centroid for the 10x10km grid used for each observation, information on the user, such as a unique surveyor identification number, and the amount of effort required in order to obtain each record. There is also the ornithological data, such as species scientific and common names. Within this data, there are some notable omissions, such as accurate date and locations. The data can only be accurate to the year and the location based on the grid in which the observation was recorded. This demonstrates a constant, error in the positional and temporal accuracy; however, since the error is constant it can be accounted for.

Since there is no reference data available to cross check, it is difficult to calculate the quality of data completeness. An estimate of completeness taken from the number of counts would be inaccurate as well as this number varies greatly depending on population density.

4.3.4 ISO Standards

The main focus of this data quality analysis will fall under the thematic accuracy quality element according to the ISO 19113/19114 standards, as the estimation of thematic accuracy can be derived using the number of surveyors in each grid cell and the number of unique bird species observed.

The R script was used to take the number of unique scientific names per grid cell, and new data tables were created based on the grid cell as opposed to observations. In addition, the number of unique surveyor numbers was also added to the new data table. Mapped individually, these variables are highly biased based on population density. When the unique bird count is proportioned against the count of unique surveyors, the population bias associated with Southern Ontario is taken into account. Figure 18 shows a scatterplot of the relation of number of observers to the number of unique bird counts within UTM zone 17n.

Figure 18 R-Generated Scatter plot Number of Unique Birds vs Number of Observers

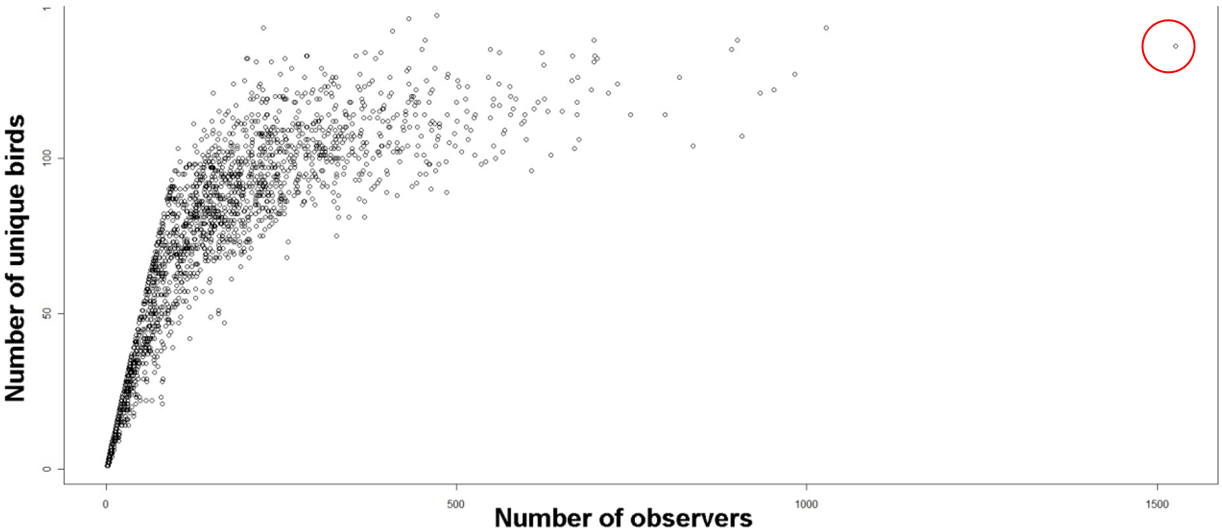
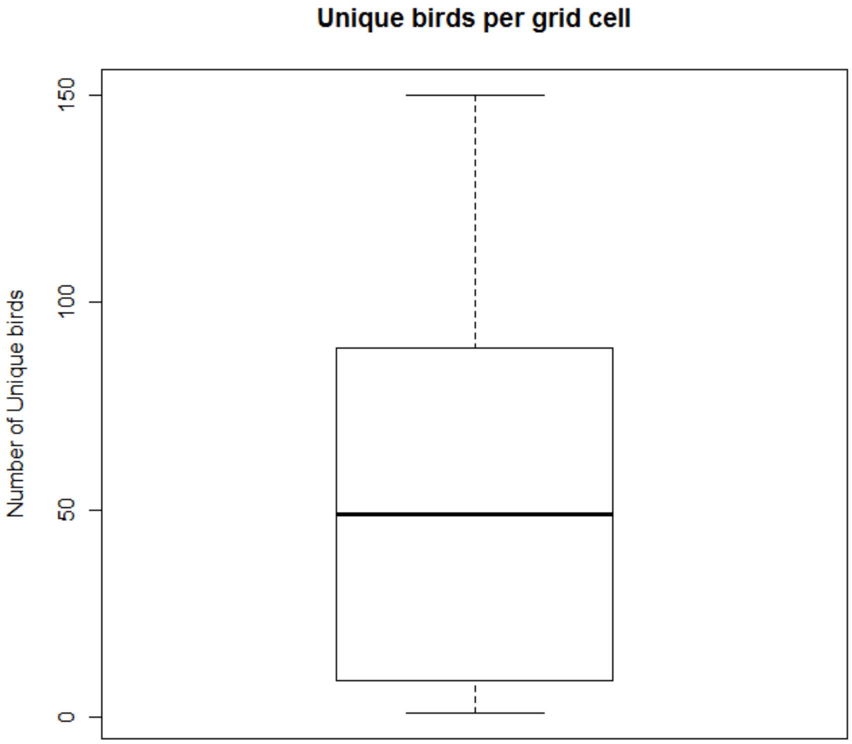


Figure 19 displays how as the number of observers increases, so too will the number of unique bird observations. There are some notable outliers within this data set. The greatest number of observers is located within a grid cell which encompasses the area of long point, Ontario. The reason for this number being so high is that the Ontario Breeding Bird Atlas study headquarters is located within this grid cell as well as several bird observatories, thus increasing both the number of birders and the number of observations. The metrics created in order to test the quality of the data allow for an assessment without bias and without having reference data. The overall trend in the data is that as the number of observers per grid cell increase, so do the number of unique observations. This pattern follows what would be expected of the data. What becomes unique to this dataset is the outliers located in the extreme high and low observation counts outside the first and third quartiles.

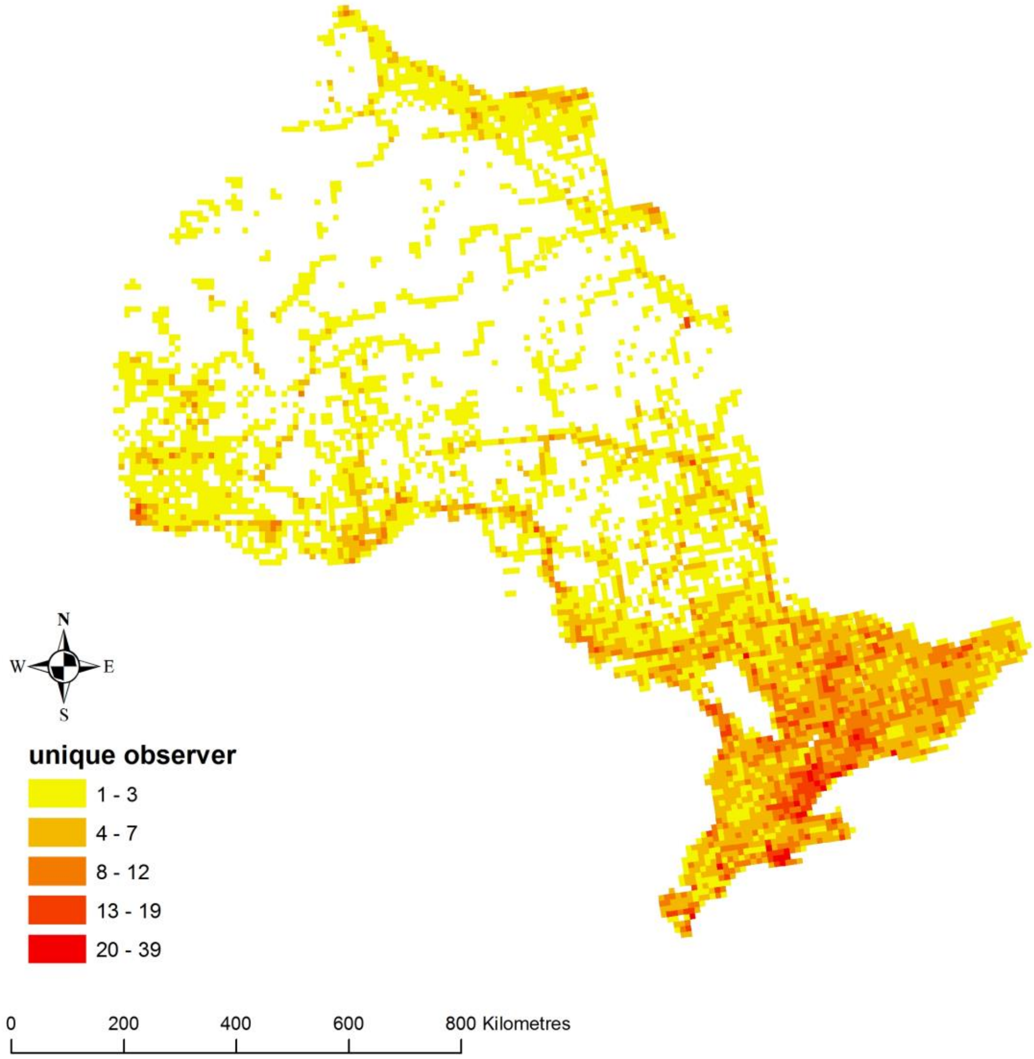
Figure 19 Unique Birds per Grid Cell Box Plot



This can be seen in the box plot, Figure 19, which shows the mean of the data which nears fifty. The more scattered distribution of the data can be observed in the scatter plot Figure 18. There is a propensity that the data demonstrates more random, and possibly less, accurate observations. This error applies to the thematic accuracy associated with the ISO 19113 and 19114 standards for geographic quality. When looking at the other elements of data quality, a systematic error can be seen in both the positional and temporal accuracy. Each entity within the data set shows an x and y location of the centroid of the grid cell with the possibility of the accuracy being somewhere within the ten kilometer square grid cell.

With recorded observations, that the citizen scientist may mistakenly identify the wrong grid cell where the observation was made. It is impossible for this to be confirmed or denied as there is no reference data or more accurate records of locations of observation. Due to the large amount of data collected, traditional statistical methods were necessary to sift through the dataset and gather relevant information. The data was difficult to interpret using conventional means as there was an excess of records. With 541,743 entries constraining 169 variables each, the roughly 96 million data points create issues when processing. Using SQL or traditional spreadsheet methods proved to be an inefficient method of digesting this data. Tomlinson (2003) explains how the method of importing data can matter just as much as the source, because basically any method of integrating large amounts of data will either cull, smooth out, or otherwise mitigate disparities which can create inaccuracies.

In order to process the data quickly and accurately, a script using the R programming languages was created; this script allowed for data to be brought in and processed without taking an overly long time. The script took the .csv file and calculated the number of unique observers and unique birds per grid cell. Some additional preprocessing was needed in order to further ease computations. Because the data was already divided into UTM zones, each zone was dealt with individually. These numbers can be visualized on a map of the study area as seen in Figure 20 Unique Observers per Grid Cell



and Figure 21

Figure 20 Unique Observers per Grid Cell

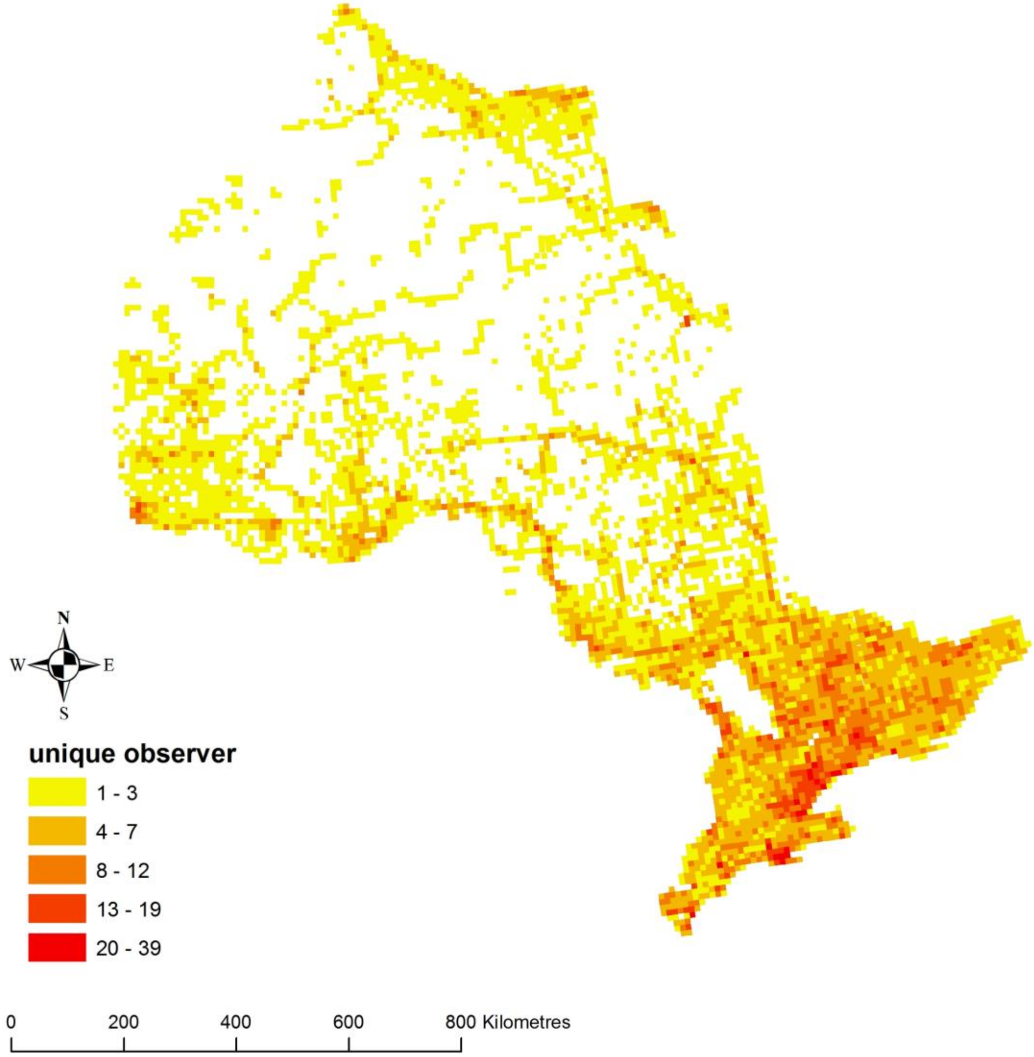
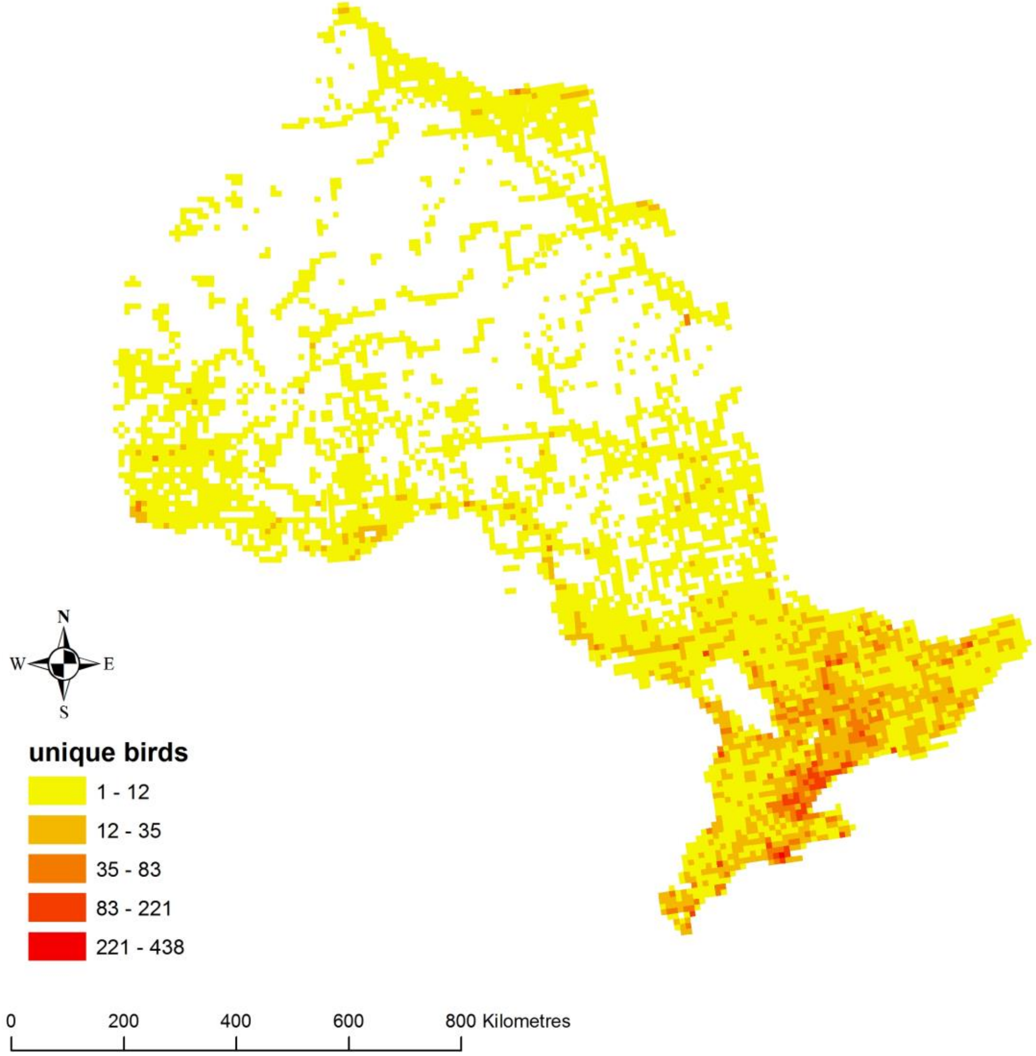
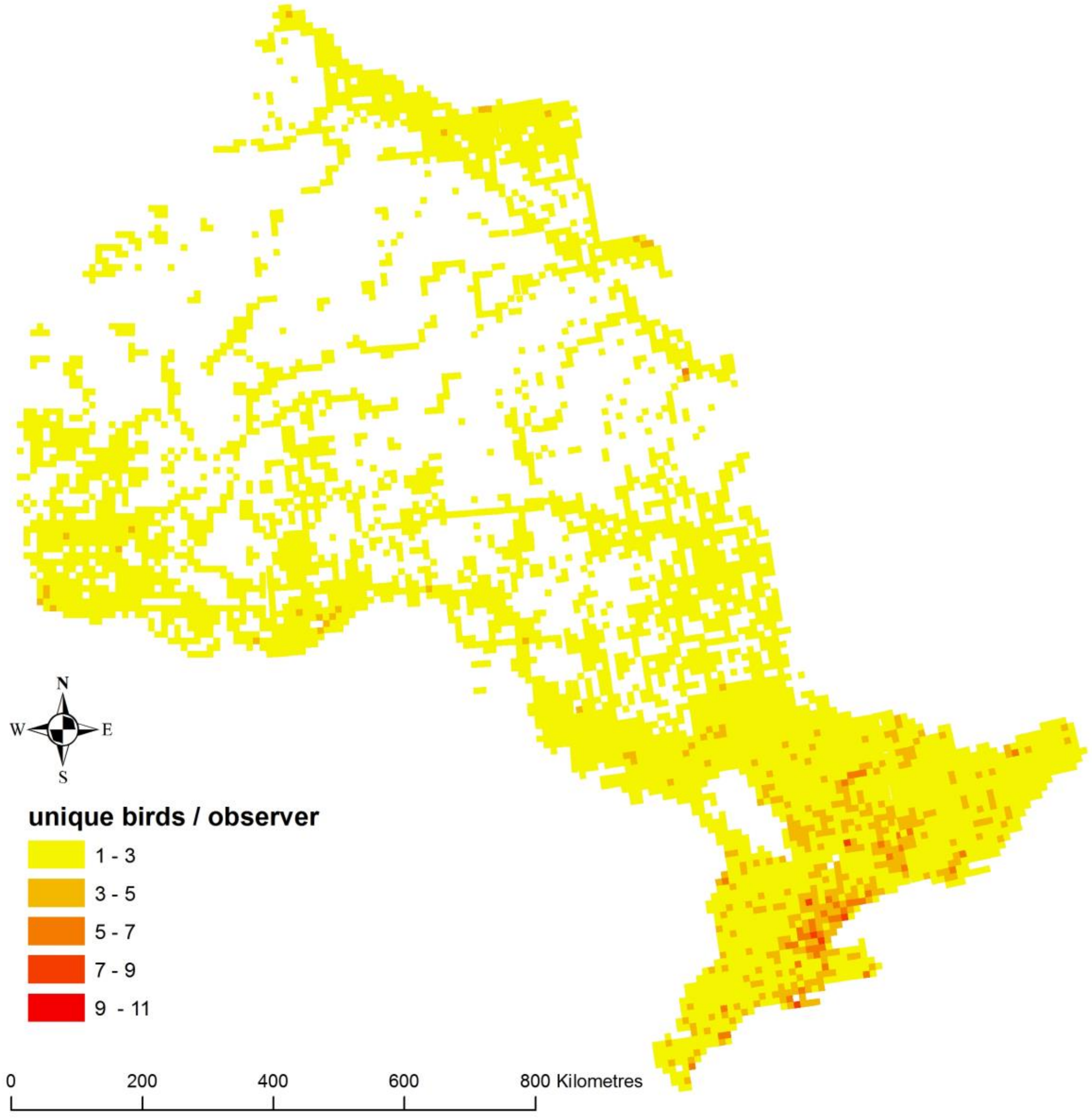


Figure 21Unique Birds per Grid Cell



These maps show the count of observers in the more populated areas near southern Ontario and a concentration of unique observations. When the number of unique observers and species were compared against each other, this change allowed the increased population to be taken into account. In Figure 21, there is an increase along the Trans-Canada highway, which runs along the south western shore of Ontario. This area has the highest number of unique observers of any area in Ontario. Some assumptions need to be drawn which would conclude that areas shown in yellow in Figure 21 are possibly more accurate than areas shown in red. In Figure 22, the results are more normalized, and a more accurate view of outliers can be seen. One notable outlier is visible along the south shore of Ontario. This grid cell contains Long Point Ontario, and this is the location of the headquarters of the bird study. The idea that more populous areas have a greater number of data samples, and therefore, increased accuracy, is supported by the literature regarding both OSM and bird counts (Wiersma 2010). One way that the accuracy of bird VGI could be improved is through systemic, post-hoc; methods, for example, data submitted to eBird are automatically assessed through a comparison of each new data entry with existing ones to filter out improbable sightings based on proximity to previous sightings (Wiersma 2010).

Figure 22 Ratioed Unique Birds against Unique Observers per Grid Cell



4.3.5 Results

Figure 22 shows areas of possible disparity of the data. The more yellow the grid cell, the more likely the provided data is erroneous. This may indicate that there are trace amounts of possibly low quality data, but overall the bird atlas data is mostly accurate. When looking at the histogram of unique birds, the bimodal nature of the graph seems to suggest that for most study sites, there are very few unique bird sightings. However, there is a point in the data where the number of unique bird sightings decreases, this decrease is visible right below one hundred sightings. This can be attributed to the number of observers needed in order to count this many unique birds. In other words, in more populous areas, more unique birds will be encountered due to the sheer amount of birdwatchers; more observers means more possible inaccuracies among the data set.

Figure 23 Unique Observers per Grid Cell Boxplot

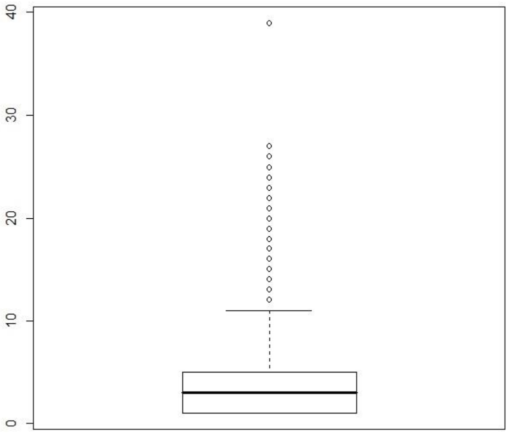


Figure 25 shows a negative exponential trend in the number of observers per grid cell. There are many study cells with only a few observers and only a few cells with many observers. This is a pattern that can be attributed to population density of the areas where these cells with many observers are located. Furthermore, the two box plots Figure 24 and Figure 25 associated with the previous graphs also demonstrates overall patterns among the unique bird sightings and submitters.

Figure 24 Frequency of Unique Birds

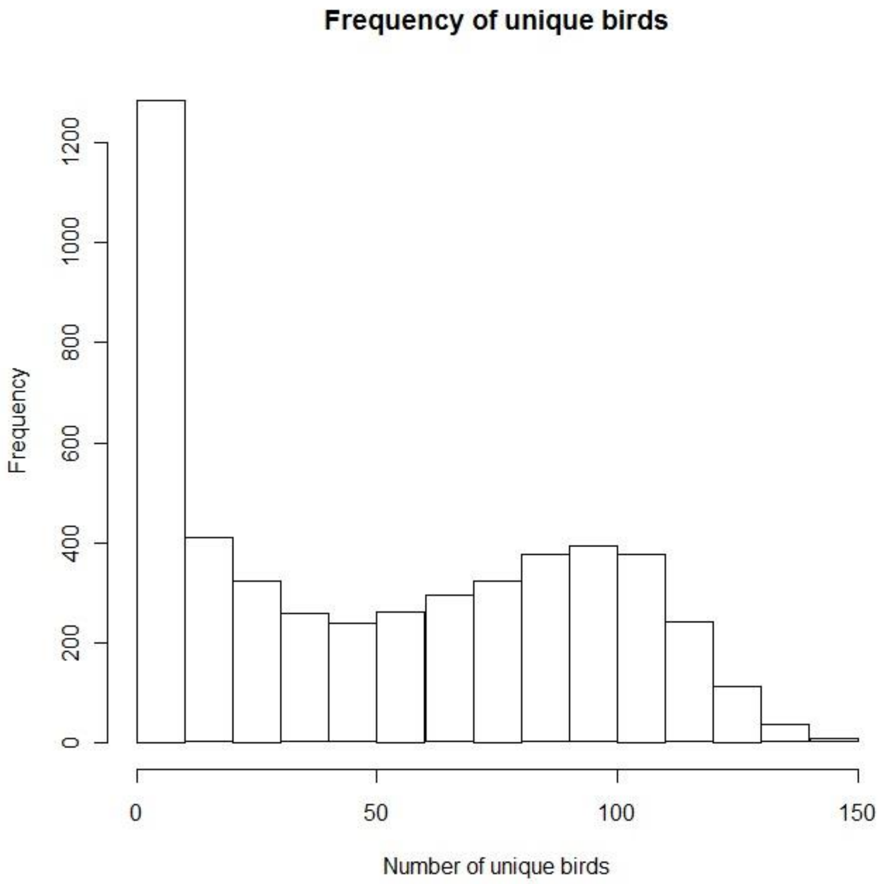


Figure 25 Collectors/Unique Bird Sightings Box Plot

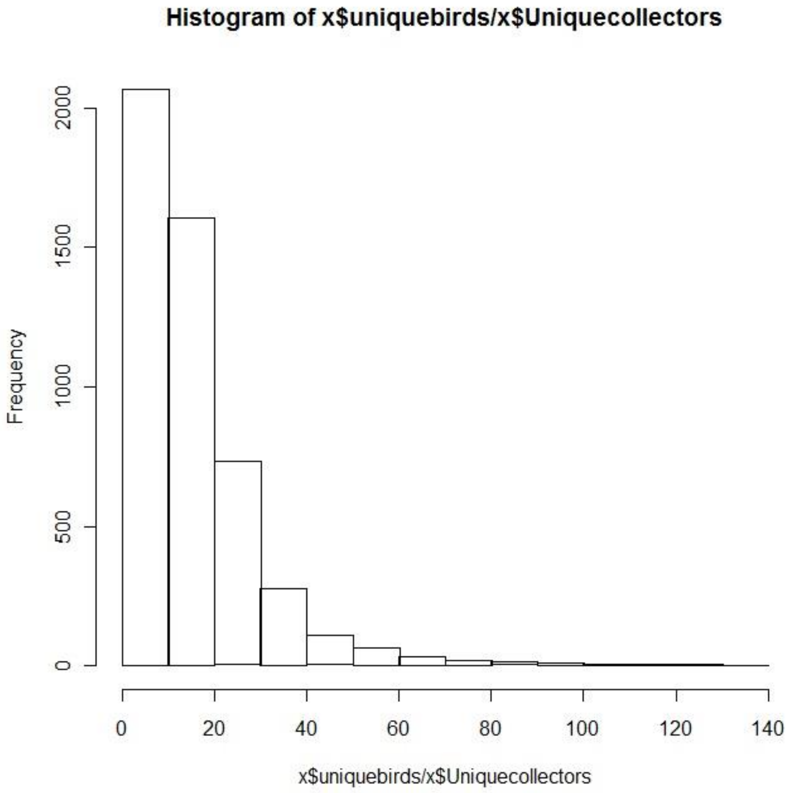


Figure 24 and Figure 25 establishes a median number of unique birds just above the fifty count, which correlates to the point on Figure 24 where the second peak of the ten starts. However, this second peak still falls within the 3rd quartile of the data and is close enough to the mean to consider it correct. Figure 25 exhibits a similar trend concerning the frequency of unique observers. This trend is what was expected as there are fewer cells with large number of observers versus the number of unique bird sightings, so when this is mapped out as in Figure 22, the areas which show a high number of unique birds per observer (shown in red) can in all probability be considered as erroneous, or of lower quality than the study cells marked with yellow. The results of the analysis of this data are able to determine areas where quality may be higher than in other areas. There is a need for vagueness when discussing the level of quality as whether or not a dataset is of high quality depends entirely on the context and intended use (Natural Resources Canada, 2012).

5. Discussion

The use of VGI to visualize the locations of birds is a fascinating area where technology, nature, and the public come together to better understand the world. The study of birds has long relied upon data submitted by the public in the form of sightings provided by proud amateur birdwatchers (Wiersma 2010). The use of citizen scientists has been common practice for more than one hundred years with technology enabling "citizen sensors," that can return photographic, spatial, temporal, contextual, and other valuable types of data (Wiersma 2010). In this case study, the trend of higher participation resulting in more unique species sighted is clearly on display. eBird is an international bird mapping effort undertaken by Cornell University, though some argue that it may not fit the traditional interpretation of a scientific researcher (Cornell Lab of Ornithology 2014; Sullivan et al. 2009; Wiersma 2010) . Ebird differs from the current study because data is checked against previous entries (Wiersma 2010).

This case study focuses on the data collected, rather than the sources who provided the counts. "Volunteer [bird survey] observers often perform worse during their first year on a survey route compared with later years," due to being inexperienced with the process of identifying birds (Farmer, Leonard,Mills Flemming and Anderson, 2014). The study of birds lends itself to study using human sources, as traditional scientific instruments, such as tracking cameras and track beds, may be ineffective ways to ascertain the level of bird populations, particularly in urban areas (Rogers, et al., n.d.).

In terms of semantics, when bird watchers contribute bird sighting information, they have to choose a species standard, but because "these standards differ in the way they assign species to birds, so that the same bird is classified differently in different parts of the world" Kuhn, 2007, p. 8). Differences like these make developing global birding databases based on VGI a difficult task. When it comes to citizen scientists making accurate observations of birds, one study noted that as the age of the participants increased, so did their rate of species detection (Farmer, Leonard,Mills Flemming and Anderson, 2014).

Farmer, Leonard,Mills Flemming and Anderson (2004) Research has demonstrated that birds with higher frequency vocalizations are less likely to be detected by older observers. This case study was unable to account for age as a factor, but considering local or study participants' demographics could prove useful in further research. The differences between young and old observers may be linked to hearing loss or other factors, but more importantly, points to the subtle ways that bias can creep into VGI, particularly when ISO or other accuracy standards may not in place (Farmer, Leonard,Mills Flemming and Anderson, 2014).

In this case study, there was no demographic information available to determine whether age or other factors could have affected the reliability of the data provided.

Foster and Dunham (2014) examine socioeconomic factors that lead to discrepancies in coverage of environmental VGI efforts. Similarly, other authors have identified socioeconomic issues as an area for further research when looking at VGI accuracy (Gazi and Roche, n.d.). Volunteers may come from a variety of backgrounds, and "relying only on GIS experts neglects the fact that involving interested users is an important step towards an open and democratic approach" to GIS (Gazi and Roche, n.d., p. 2).

Paudyal, McDougal and Apan (2012) explain how VGI can be a vital tool for natural resource management, including counting species and evaluating biodiversity. However, national mapping agencies, though recognizing the value of VGI, have yet to adopt its usage widely, with one survey pegging the use of VGI by mapping agencies at around 35% (Paudyal, McDougal and Apan, 2012).

Newman, Graham, Crall and Laituri (2011, p. 218) cite numerous examples of volunteer data being used for bird-related efforts in North America: "Project FeederWatch, PigeonWatch, NestWatch, NestCams, Great Backyard Bird Count, eBird, Celebrate Urban Birds, CamClickr, BirdSleuth, and Birds in Forested Landscapes." Further research in this area must be focused on ascertaining sociological information about the volunteers and metadata about the VGI itself so that a better assessment of accuracy, validity, and usefulness may be performed. Users have a number of reasons for contributing to VGI efforts that are environmentally focused, including "awareness and concern regarding environmental benefits, long standing love with the land and water, and social interactions/benefits" (Paudyal, McDougal and Apan, 2012, p. 279). Volunteers who count species and provide VGI generally have a vested interest in the plants and animals themselves, as well as concern for the local areas where they live and hike.

Technical instruments could be used to survey birds in a limited manner, such as monitoring nests, but human efforts are much more refined. In another comparison of technology versus VGI, when counting species, the “Infrared Triggered Camera (ITC) deer survey, a scientifically accepted survey method, was conducted in the same area” as a “VGI population estimate was 72% of the ITC population estimate” (Nicosia, 2013 xii). This shows that VGI is promising by itself, but may be amplified by using technological methods in conjunction with human ones. In the birding community, in VGI projects such as “2nd South African Bird Atlas Project (SABAP2), the data are gathered by pentad (areas 5' by 5') by individual, amateur birders and contributed directly” to the project in accordance with guidelines (Cooper, et al., 2012). Direct contribution of data by volunteers can be contrasted with data that is collected and synthesized by researchers (Cooper, et al., 2012).

Amateur birders have an intense passion for identifying and claiming bird sightings that can be both a boon and a boondoggle for the VGI they provide (Scott, Cavin, Cronan and Kerins, 2005). Some members of this community are highly competitive in terms of trying to see the largest number of unique birds in a year’s time (Scott, Cavin, Cronan and Kerins, 2005) . This form of super-leisure entails a prejudicial classism within the birding world, as “birders are quick to distance themselves from people whose participation is limited to feeding and watching birds around the home” (Scott, Cavin, Cronan and Kerina, 2005, p. 3).

Birding has become such a specialized hobby, it has transcended an amateur practice, becoming “a fixation on listing [that] is often accompanied by the acquisition, over time, of outstanding identification skills,” whereby “most elite birders can identify birds fairly easily in the field and rarely need to refer to a field guide to make a positive identification” (Scott, Cavin, Cronan and Kerina, 2005, p. 3).

Furthermore, the development of highly skilled amateur birders means that VGI dealing with bird counts may be more reliable than in other fields. In fact, “identification skills have become a standard by which many birders judge and accept others as rightful members of the birding social world…result[ing] in the emergence of performance standards characterized by exclusivity and elitism” (Scott, Cavin, Cronan and Kerina, 2005, p. 3). Indeed, there is a form of purism associated with counting birds, as in the case of James Vardaman’s Big Year in 1979, when he counted 699 birds by spending thousands of dollars on field guides, drawing the ire of dedicated amateur birders (Scott, Cavin, Cronan and Kerina, 2005, p. 3). Unlike scientists, who may rely on grants or limited field studies, “it is not uncommon for hardcore birders to drop everything they are doing and spend hundreds, if not thousands, of dollars for an opportunity to see [a rare] bird…those hardcore participants pursuing a Big Year spend hundreds of days birding, many of these away from home” (Scott, Cavin, Cronan and Kerina, 2005, p. 4). The passion and dedication of volunteers who provide geographic data for locating birds should not be underestimated.

Case study 2 intersects with case study 3 due to the connection between trails and bird-watching (Foster, 2014). Many birders rely on GPS units to visit sites where rare, migratory, non-local, or otherwise remarkable birds can be found, including looking to other birders for repositories of GPS data (Foster, 2014). The community is quite productive in terms of cataloguing birds in the wild, and it makes sense for academic, business, and government stakeholders to leverage VGI from the birding community while understanding the context of its creation.

Whether the metrics derived for these tests are conclusive in proving the quality of the data depends highly on the use of the VGI. Different uses of VGI have vastly different needs as far as quality. The term fitness-for-use is used to describe how well a particular dataset fits a certain use, and that is something that is imperative to consider when dealing with all VGI data. Some researchers imply that there is an excessive emphasis on positional or technical accuracy, which may distract from an awareness of fitness for use and the needs of the user. Without having some sort of reference data, it is impossible to definitively decree that one set of data is correct. If no reference data is available, the only way to utilize data sets which contain potentially inaccurate points is to identify them as outliers and examine why they fall outside of the normal range, possibly ignoring them depending on the specific context. The only significant conclusion, which can be drawn from the metrics created herein, is that some parts of each dataset may be of a higher quality than others. As well as care should be taken with any dataset when examining the particular areas which have a higher probability of being incorrect when the intended use of the data is for something that require and high probability of being correct throughout. In essence, the usage of data can partly determine the subjective level of accuracy. Tomlinson (2003, 39) writes, “you must establish how much error can be tolerated in the information product without losing its usefulness," and more accurate data can be expensive, time-consuming, or simply unavailable, which means that users must "think about error and consider what level of accuracy is important in terms of cost versus reliability."

In the literature and resources under analysis, there were instances of user submitted data which could not be compared to official GI, as in case studies 2 and 3 (De Longueville, Smith, and Luraschi 2009; Poser and Dransch 2010; Okolloh 2009; Kounadi 2009).

5.1 Applications for Research

This paper contributes to the current literature and VGI research by building upon the framework set out in the ISO 19113 and 19114 standards. By having a concrete foundation from which to base a definition of quality for VGI, VGI can be used more widely in analytical research. Establishing the links between the ISO standards and the methods and interpretations of VGD and VGI provides a framework from which other researchers may expand upon in order to further incorporate user-created data into research projects. An additional benefit to the use of VGD is that being able to utilize more user-created data will reduce the cost and man power associated with collecting accurate data through other means, such as using human experts or technological methods.

The crowd is a valuable source of geographical information, because nearly every person on Earth now has access to some sort of mobile or internet enabled device which can provide GI or relevant metadata. Encouraging the submitters and users of this aggregated data to understand and abide by the type of standards recommended by the ISO will lead to data which is more accurate and more usable for a wide range of important tasks. Standardized data can be used outside of its original intentions, as in case study 2, where satellite maps are contrasted with VGD in the form of GPS track points and tracks. The digitalization of data has made it much easier to import, export, and otherwise migrate data from source to source, meaning that the labourious process of entering and verifying data can be made easier, though the same caveats regarding varying levels of accuracy do still apply (Tomlinson 2003). The proper implementation of ISO standards by providers of GI allows for the regulation of aspects relative to the evaluation and description of the Geographical Information quality, thus avoiding ambiguous information and facilitating the adequate choice and use of the products. Furthermore, these standards seek full understanding between producers and users of this kind of information, favoring their commercialization, spread and efficient use. As with the development of any new technology, for example, the universal serial bus charging standard (IEC 62684) which allows different phones to use the same mini-usb charger, the IEC\ISO standards lead to interoperability between different sets of data and information.

While any large or significant findings cannot be predicted, further research into VGI quality standardization will assist the geomatics community as a whole in order to help provide new, worthwhile sources of data which may be used with confidence.

5.2 Data Quality of VGI

According to Grira, Bedard and Roche (2010, 62), “with spatial data accessible to members of the general public who have little formal training in quality issues, the GI science community is facing a new situation that raises questions about the communicated quality (Boin and Hunter 2007) and its different users’ perceptions.” Furthermore, “the increasing number of users leads to a wide range of requirements, to different assessment processes, and consequently, to a variety of quality perceptions” (Boin and Hunter 2007) which may differ from the ISO definition of quality. Overall, VGI provides a reasonable level of quality data according to the ISO standards, though certain areas such as type and nomenclature are often miscategorized by amateur cartography enthusiasts. Intentional and unintentional mistakes in data can be mitigated by comparing user submitted data to accurate official data sets through the use of algorithms or human intervention.

5.3 Previous Work

Previous research in the area of VGI generally supports the conclusion that user-generated data is an exciting and growing area for useful mapping data, which ISO standardization can influence toward greater accuracy. My results found in case study 3 were similar to Kounadi (2009) and Ather (2009) who compare OSM data to official maps and discover that central areas and those with major infrastructural elements contain a higher concentration of data points. In case study 3, the cell which contains Ontario, also the location of the bird mapping project headquarters, is a valuable example of how large amounts of user submitted data can be accurate overall, but may have individual accuracy shortcomings due to the number of untrained observers, this also relates to Kounadi's (2009) observations concerning road map data.

6. Conclusions

It is obvious that there is a great opportunity for incorrect data, material, and knowledge to infiltrate the world of cartography, especially with the ever increasing popularity of user-supplied information on Wikipedia type sites, that is to say sites that have the same free-sharing of information structure of Wikipedia. Until a consensus is reached regarding how to fully qualify the data provided and its correctness, there is no real way to asses VGI other than the metrics and methods presented in this study. A trend in studies conducted into the validity of VGI data has shown the need for further investigation into the number of users and demographic of users which are required for accurate data, both in short and long term studies, which could provide a continued assessment of VGI user content (Wiersma 2010). One implication of the variability of VGI applications is that there are many aspects to VGI data quality. VGI quality is heavily dependent on the intentions of the original volunteer. With VGI being used for many different applications at varying scales, appropriate measures of VGI quality are needed. In order for quality measures to be created, the elements of VGI quality must be clearly defined. (Wang and Strong 1996)

A start for defining the quality of VGI must include a more succinct definition of the term, expanding on Goodchild's definition to include; User generated spatial data with a purpose. This addition to the suffix of the definition allows to better differentiate between VGI and VGD which is a clear pathway for helping to define the quality of data and information. Referring back to Table 1 it is seen that not all elements and subelements of data quality can be evaluated at all times for all data but some intuitive ways to analyze data have been used in this thesis. Including expanding on the work of Hacklay, Mooney and Corcoran in order to utilize trail data with no reference in order to compare against itself. By no means is the method used herein an approached pertinent to all applications, but can be used in similar situations.

Hopefully, in the future, accuracy—and more importantly, ways to measure it effectively—will improve, or novel methods other than mere user submitted data, can be used to provide more accurate and quality information that does not pose a significant risk or to the consumers of geographical data. As the field of NeoGeography is still developing, the hopes for these changes are quite high.

The case studies demonstrate some of the benefits and drawbacks of user-generated volunteer geographical information. There is a glut of available data points, and by examining the patterns contained in this large set of information, a tendency toward consensus among individual contributors appears. However, the large amount of information provided can also turn the search for meaningful patterns into an exercise in finding needles among the haystack. Despite the emphasis on big data among current research paradigms, the software tools available can oftentimes require a laborious manual process of cleaning and importing the available information. Then, the data must be interpreted using a method which yields significant and interesting results that can mitigate any possible inaccuracies in the submitted data.

The idea that there is no concrete way of determining whether data is accurate, combined with the potential lack of credentials on the part of users who submit information, can be frightening. Future research in this area should ascertain better methods and metrics to categorize the data submitted to VGI databases, as well as ways to better assess the background of the individuals who contribute this information. Only then can VGI data truly be relied upon for problems which require critical and accurate information.

7. References

Anderson, T. 2008. *Theory and Practice Of Online Learning*. 1st ed. Edmonton: AU Press.

Ariza, F, J. n.d. Standards about Quality in Geographical Information (ISO 19113, ISO 19114, ISO 19138).

Ather, A. 2009. “A Quality Analysis of OpenStreetMap Data.” *Unpublished M. Eng. Dissertation, UCL*.

Barton, N, P. 1981. “Measurement of Transmittance and Veiling Glare Index.” *Assessment of Imaging Systems 2*. 239-47.

Bégin, D., R. Devillers, and S. Roche. 2013. “Assessing Volunteered Geographic Information (VGI) Quality Based on Contributors’mapping Behaviours.” *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 1 (1): 149–54.

Boin, Anna T., and Gary J. Hunter. 2007. “What Communicates Quality to the Spatial Data Consumer.” In *Proceedings of the 2007 International Symposium on Spatial Data Quality (ISSDQ 2007), Enschede, The Netherlands*. http://www.isprs.org/proceedings/xxxvi/2-c43/Session5/Boin_paper[1].pdf.

Brando, Carmen, and Bénédicte Bucher. 2010. “Quality in User Generated Spatial Content: A Matter of Specifications.” In *Proceedings of the 13th AGILE*

International Conference on Geographic Information Science, 11–14. Springer Verlag: Guimar aes, Portugal. http://www.agile-online.orgwww.agile-online.org/Conference_Paper/CDs/agile_2010/ShortPapers_PDF/105_DOC.pdf.

Burrough, P. A., R. van Rijn, and M. Rikken, 1996. "Spatial Data Quality and Error Analysis Issues: GIS Functions and Environmental Modeling." *GIS and Environmental Modeling: Progress and Research Issues*.

Busch, Andreas, and Felistus Willrich. 2002. “Quality Management of ATKIS Data.” In *Proceedings of OEEPE/ISPRS Joint Workshop on Spatial Data Quality Management, Istanbul*. http://www.ipi.uni-hannover.de/uploads/tx_tkpublikationen/paper_istanbul_quality_02.pdf.

Butcher, G., Fuller G, McAllister L, and Geissler P. 1990. “An Evaluation of the Christmas Bird Count for Monitoring Population Trends of Selected Species.” *Wildlife Society Bulletin*, 129–34.

Butler, H. "A guide to the Python Universe for ESRI users." In *Annual ESRI International Conference, San Diego, California, USA*. 2004.

Carmen, B and Bucher, B. "Quality in user generated spatial content: A matter of specifications." In Proceedings of the 13th AGILE International Conference on Geographic Information Science, pp. 11-14. Springer Verlag: Guimar aes, Portugal, 2010.

Casanovas, Pompeu, and Marta Poblet. 2012. “Crowdsourced Crisis Mapping: How It Works and Why It Matters.” In *The Conversation*. http://ddd.uab.cat/pub/artpub/2012/118890/conversation_a2012m5n1iENG.pdf.

Caspary, Wilhelm, and Gerhard Joos. 2002. “Statistical Quality Control of Geodata.” *Spatial Data Quality,* 106.

Ceballos, Joselyn A Robledo, and Giannina L Reyes Gatica. 'Implementation Of A Management And Quality Control System Under Iso Standards 9001: 2000, 19113, 19114, 19138 And 19115 In Cartographic Production'. Accessed May 28 2014.
http://icaci.org/files/documents/ICC_proceedings/ICC2009/html/refer/3_2.pdf

Chen, D, et al. "Assessment of open source GIS software for water resources management in developing countries." *Journal of Hydro-environment Research* 4.3 (2010): 253-264.

Coleman, D, Georgiadou Y, Labonte J, et al. 2009. 'Volunteered Geographic Information: The Nature and Motivation of Produsers'. *International Journal Of Spatial Data Infrastructures Research* 4 (1): 332--358.ujyoi7utfgcb

Coleman, D. 2013. “Potential Contributions and Challenges of VGI for Conventional Topographic Base-Mapping Programs.” In *Crowdsourcing Geographic Knowledge*, 245–63. Springer. http://link.springer.com/chapter/10.1007/978-94-007-4587-2_14.

Cooper, A, Coetzee S, and Kourie D. 2012. 'Assessing The Quality Of Repositories Of Volunteered Geographical Information'. *Ee Publishers*.

Cornell Lab of Ornithology. 2014. 'Ebird'. *Ebird.Org*. Accessed August 18 2014. http://ebird.org/content/ebird/.

Couch, P. 2011. 'GIS Buffer'. *Gistutor.Com*. Accessed May 28 2014. http://www.gistutor.com/concepts/9-beginner-concept-tutorials/46-gis-buffer.html.

Davis, B. 1996. *GIS: A Visual Approach*. 1st ed. Santa Fe: OnWord Press.

De Longueville, B, Smith, R, and Luraschi,G . 2009. "OMG, from here, I can see the flames!": A use case of mining Location Based Social Networks to acquire spatio-temporal data on forest fires." *Institute for Environment and Sustainability.* 73-80.

Devillers, R, B'egin D, and Vandecasteele A. 2012. 'Is The Rise Of Volunteered Geographic Information (VGI) A Sign Of The End Of National Mapping Agencies As We Know Them'.

DiNucci, D. 1999. "Fragmented Future" Accessed May 28 2014. http://www. darcyd. com/fragmented_future.pdf

Duckham, M. 2002. 'A User-Oriented Perspective Of Error-Sensitive GIS Development'.*Transactions In GIS* 6 (2): 179--193.

Etherington, T. 2011. "Python based GIS tools for landscape genetics: visualising genetic relatedness and measuring landscape connectivity." *Methods in Ecology and Evolution* 2, no. 1: 52-55.

Exel, M., E. Dias, and S. Fruijiter. 2010. The Impact of Crowdsourcing on Spatial Data Quality Indicators. *Proceedings of GiScience 2010*. 213-218.

Farmer, Robert G., Marty L. Leonard, Joanna E. Mills Flemming, and Sean C. Anderson. 2014. “Observer Aging and Long-Term Avian Survey Data Quality.” *Ecology and Evolution*. http://onlinelibrary.wiley.com/doi/10.1002/ece3.1101/full.

Flanagin, A, and J Metzge, M. 2008. 'The Credibility Of Volunteered Geographic Information'. *Geojournal* 72 (3-4): 137--148.

Foster, A, and Dunham, I. 2014. “Volunteered Geographic Information, Urban Forests, and Environmental Justice.” *Computers, Environment and Urban Systems*. http://www.sciencedirect.com/science/article/pii/S0198971514000945.

Gazi, Anisur Rahman, and Stéphane Roche. ND. “Towards a Socio-Economical Evaluation Framework of Volunteered Geographic Information (VGI).” Accessed January 14. http://rose.geog.mcgill.ca/ski/system/files/fm/2009/Gazi_abstract.pdf.

Gibson, R, and Erle, S. 2006. *Google Maps Hacks*. O’Reilly Media, Inc.

Goodchild, M. 2007. “Citizens as Sensors: The World of Volunteered Geography.” *GeoJournal 69*. 211-21.

Goodchild, M. 2009. 'Neogeography And The Nature Of Geographic Expertise'. *Journal Of Location Based Services 3 (2)*: 82.

Goodchild, Michael F., L. T. Steyaert, B. O. Parks, C. Johnston, D. Maidment, M. Crane, and S. Glendinning, Editors. GIS World Books. pp. 29-34.

Gordon, M. 2014. 'Bad GPS Data: What/Why/How'. *Strava Customer Support*. Accessed May 28 2014. https://strava.zendesk.com/entries/22645314-Bad-GPS-Data-What-Why-How#fix.

Grira, J, Bédard, Y, and Roche, S. 2010. "Spatial data uncertainty in the VGI world: Going from consumer to producer." *Geomatica* 64, no. 1: 61-72.

Haklay, M, Basiouka , S, Antoniou, V, and Ather, A. 2010. "How many volunteers does it take to map an area well? The validity of Linus’ law to volunteered geographic information." *The Cartographic Journal* 47, no. 4: 315-322.

Haklay, M. 2009. 'Beyond Good Enough? Spatial Data Quality And Openstreetmap Data'.

Haklay, Muki. 2013. “Citizen Science and Volunteered Geographic Information: Overview and Typology of Participation.” In *Crowdsourcing Geographic*

Knowledge, 105–22. Springer. http://link.springer.com/chapter/10.1007/978-94-007-4587-2_7.

Heipke, Christian. 2010. “Crowdsourcing Geospatial Data.” *ISPRS Journal of Photogrammetry and Remote Sensing* 65 (6): 550–57.

Helbich, M, Amelunxen, C, Neis, P, and Zipf, A. 2012 "Comparative spatial analysis of positional accuracy of OpenStreetMap and proprietary geodata." *Proceedings of GI_Forum 2012:* 24-33.

Hochmair, H, Zielstra, D, and Neis, P. 2013. 'Assessing The Completeness Of Bicycle Trail And Lane Features In Openstreetmap For The United States'. Accessed August 15 2014. http://flrec.ifas.ufl.edu/hochmair/pubs/hochmair_zielstra_neis_2013_TGIS_bicycle_prePrint.pdf

Hunter, Gaiy J., and M. F. Goodchild. 1993. “Managing Uncertainty in Spatial Databases: Putting Theory into Practice.” In *Papers From The Annual Conference-Urban And Regional Information Systems Association*, 15–15. Urisa Urban And Regional Information Systems. http://www.geog.ucsb.edu/~good/papers/199.pdf.

Iliffe, M. 2012. 'Towards Understanding The Factors Of Integrating Spatial Data Infrastructures And Crowdsourced Geographic Data'. *1 St AGILE Phd School*, 71.

International Standards Organization. 2014. 'ISO 19113:2002 - Geographic Information -- Quality Principles'. *Iso.Org*. Accessed May 31 2014. http://www.iso.org/iso/catalogue_detail.htm?csnumber=26018.

ISO/IEC IS 19113:2000. 2002. *Geographic Information — Quality Principles*. Geneva, Switzerland.

ISO/IEC IS 19114:2003. 2009. *Geographic Information — Quality Evaluation Procedures*. Geneva, Switzerland.

Jackson, Bethanna, Timothy Pagella, Fergus Sinclair, Barbara Orellana, Alex Henshaw, Brian Reynolds, Neil Mcintyre, Howard Wheater, and Amy Eycott. 2013. “Polyscape: A GIS Mapping Framework Providing Efficient and Spatially Explicit Landscape-Scale Valuation of Multiple Ecosystem Services.” *Landscape and Urban Planning* 112: 74–88.

Klien, Eva, Daniel I. Fitzner, and Patrick Maué. 2007. “Baseline for Registering and Annotating Geodata in a Semantic Web Service Framework.” In *Proceedings of the 10th Conference on Geographic Information Science, Aalborg, Denmark*. http://wwww.agile-online.org/Conference_Paper/CDs/agile_2007/PROC/PDF/119_PDF.pdf.

Koukoletsos, T, Haklay, H, and Ellul. C. 2011. 'An Automated Method To Assess Data Completeness And Positional Accuracy Of Openstreetmap', 20--22.

Kounadi, O. 2009. 'Assessing The Quality Of Openstreetmap Data'. *Msc Geographical Information Science, University College Of London Department Of Civil, Environmental And Geomatic Engineering*.

Kuhn, Werner. 2007. “Volunteered Geographic Information and GIScience.” *NCGIA, UC Santa Barbara*, 13–14.

Maué, Patrick. “Reputation as Tool to Ensure Validity of VGI.” In *Workshop on Volunteered Geographic Information*, 2007.http://www.ncgia.ucsb.edu/projects/vgi/docs/position/Maue_paper.pdf.

McAvoy, A. 2013. “Hawaii Hiking Trails to Be on Google Street View.” *The Big Story*. June 27. http://bigstory.ap.org/article/hawaii-hiking-trails-be-google-street-view.

Mooney, P, Corcoran, P, and Winstanley, A. 2010. 'Towards Quality Metrics For Openstreetmap', 514--517.

Natural Resources Canada,. 2012. 'Volunteered Geographic Information (VGI) Primer'. ftp://s5-bsc-faisan.cits.rncan.gc.ca/pub/geott/ess_pubs/291/291948/cgdi_ip_21e.pdf.

Neis, P, and Zielstra, D. 2014. 'Recent Developments And Future Trends In Volunteered Geographic Information Research: The Case Of Openstreetmap'. *Future Internet* 6 (1): 76--106.

Newman, Greg, Jim Graham, Alycia Crall, and Melinda Laituri. 2011. “The Art and Science of Multi-Scale Citizen Science Support.” *Ecological Informatics* 6 (3): 217–27.

Newson, S and Noble, D. 2003. *Producing statistically valid maps of species abundance from UK Breeding Bird Survey counts using Geostatistical Analyst in ArcGIS*. British Trust for Ornithology.

Nunez-Redo, M, Diaz, L, Gil J, Gonzalez D, and Huerta, J. 2011.“Discovery and Integration of Web 2.0 Content into Geospatial Information Infrastructures: A Use Case in Wildfire Monitoring.” *Availability, Reliability, and Security for Business, Enterprise, and Health Information Systems 6908.* 50-68.

Okolloh, O. 2009. “Ushahidi, Or’testimony’: Web 2.0 Tools for Crowdsourcing Crisis Information.” *Participatory Learning and Action* 59 (1): 65–70.

OpenStreetMap. 2014. 'Google Map Maker - Openstreetmap Wiki'. *Wiki.Openstreetmap.Org*. Accessed May 28 2014. http://wiki.openstreetmap.org/wiki/Google_Map_Maker.

Panoramio,. 2014. 'Help - Welcome To The Panoramio Community'. *Panoramio.Com*. Accessed May 28 2014. http://www.panoramio.com/help.

Parfitt, I. 2013. 'Citizen Science In Conservation Biology: Best Practices In The Geoweb Era'.*University Of British Columbia*.

Paudyal, D. R., K. McDougall, and A. Apan. 2012. “Exploring the Application of Volunteered Geographic Information to Catchment Management: a Survey Approach.” *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 1: 275–80.

Poblet, Marta. 2010. “Mobile Governance: Empowering Citizens to Promote the Rule of Law.” *Submitted to the CROSSROAD Call for Contributions on Future Internet on ICT for Governance and Policy Modelling. Retrieved August* 30: 2010.

Poser, K, and Dransch, D. 2010. "Volunteered geographic information for disaster management with application to rapid flood damage estimation." *Geomatica* 64, no. 1: 89-98.

Priedhorsky, R. 2007. 'Specialist Meeting On Volunteered Geographic Information Position Paper'. *Retrieved March* 3: 2008.

Raggatt, P. R. 1989. “Duplicates or Singletons?—an Analysis of the Need for Replication in Immunoassay and a Computer Program to Calculate the Distribution of Outliers, Error Rate and the Precision Profile from Assay Duplicates.” *Annals of Clinical Biochemistry: An International Journal of Biochemistry in Medicine* 26 (1): 26–37.

Randall, E. 2014. 'How A Raccoon Became An Aardvark'. *The New Yorker*. http://www.newyorker.com/online/blogs/elements/2014/05/how-a-raccoon-became-an-aardvark.html.

Raubal, Martin, D. Mark, and A. Frank. 2013. *Cognitive and Linguistic Aspects of Geographic Space*. Springer. http://link.springer.com/content/pdf/10.1007/978-3-642-34359-9.pdf.

Region of Waterloo,. 2014. 'Maps'. *Regionofwaterloo.Ca*. Accessed August 18 2014. http://www.regionofwaterloo.ca/en/discoveringTheRegion/maps.asp#interactive.

Reichenbacher, T. 2004. “Mobile Cartography: Adaptive Visualization of Geographic Information on Mobile Devices.” Muchen: Verlag Dr. Hut. 1-4.

Reichenbacher, T. 2007. *The Concept of Relevance in Mobile Maps*. Springer. http://link.springer.com/chapter/10.1007/978-3-540-36728-4_18.

Rinner, C, and Fast, V. "A Systems Perspective on Volunteered Geographic Information." (2013).

Roberts, J., Best, B , Dunn, D, Treml, E and Halpin P. 2010. "Marine Geospatial Ecology Tools: An integrated framework for ecological geoprocessing with ArcGIS, Python, R, MATLAB, and C++." *Environmental Modelling and Software* 25, no. 10:1197-1207.

Roche, S, and Feick R. 2012. 'Wiki-Place: Building Place-Based GIS From VGI'. *Position Paper From VGI Worshop Role Of Volunteered Geographic Information In Advancing Science: Quality And Credibility, In Conjunction With Giscience 2012.*

Sharma, A. 2011. "Open Access Journals in the Field of Geography with Special Reference to Doaj: A study." *Gyankosh-The Journal of Library and Information Management* 2, no. 2: 1-14.

Sheppard, A. 2012. “WQ: A Modular Framework for Collecting, Storing, and Utilizing Experiential VGI.” *Proceedings of the 1st ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information*. 62-9.

Stark, H. 2011. 'Quality Assessment Of Volunteered Geographic Information Using Open Web Map Services Within Openaddresses', 101--110.

Statistics Canada,. 2014. 'National Household Survey (NHS) Profile, 2011'. *Www12.Statcan.Gc.Ca*. Accessed August 16 2014. http://www12.statcan.gc.ca/nhs-enm/2011/dp-pd/prof/details/page.cfm?Lang=EandGeo1=CSDandCode1=3530016andData=CountandSearchText=waterlooandSearchType=BeginsandSearchPR=01andA1=AllandB1=AllandCustom=andTABID=1.

Sui, D , and DeLyser, D. 2012. 'Crossing The Qualitative-Quantitative Chasm I Hybrid Geographies, The Spatial Turn, And Volunteered Geographic Information (VGI)'. *Progress In Human Geography* 36 (1): 111-124.

Sui, D, Elwood S, and Goodchild M. 2013. *Crowdsourcing Geographic Knowledge*. 1st ed. Dordrecht: Springer.

Sullivan, B, Wood C, Iliff M, Bonney R, Fink D, and Kelling S. 2009. 'Ebird: A Citizen-Based Bird Observation Network In The Biological Sciences'. *Biological Conservation* 142 (10): 2282--2292.

Tenney, M. 2014. 'Quality Evaluations On Canadian Openstreetmap Data'. http://rose.geog.mcgill.ca/ski/system/files/fm/2014/tenney.pdf.

Tomlinson, Roger F. 2003. *Thinking About GIS*. 1st ed. Redlands, Calif.: ESRI Press.

TrailFu,. 2014. 'Mountain Bike Trails And Reviews'. *Trailfu.Com*. Accessed August 18 2014. http://trailfu.com.

Von Lindenberg, K. 2014. 'Comparative Analysis Of GPS Data'. *Undergraduate Journal Of Mathematical Modeling: One+ Two* 5 (2): 1.

Wang, R; Strong, D. 1996. “Beyond accuracy: What data quality means to data consumers.” *Journal of Management Information Systems.* 12 (4): 5-33.ghfy

Wiersma, Y. F. 2010. Birding 2.0: citizen science and effective monitoring in the Web 2.0 world. *Avian Conservation and Ecology* **5**(2): 13. [online] URL:
http://www.ace-eco.org/vol5/iss2/art13/
http://dx.doi.org/10.5751/ACE-00427-050213

Zilske, Michael, Andreas Neumann, and Kai Nagel. 2011. “OpenStreetMap for Traffic Simulation.” In *Proceedings of the 1st European State of the Map–OpenStreetMap Conference*, 126–34. http://svn.vsp.tu-berlin.de/repos/public-svn/publications/vspwp/2011/11-10/2011-06-20_openstreetmap_for_traffic_simulation_sotm-eu.pdf.

Zook, Matthew, Mark Graham, Taylor Shelton, and Sean Gorman. 2010. “Volunteered Geographic Information and Crowdsourcing Disaster Relief: A Case Study of the Haitian Earthquake.” *World Medical and Health Policy* 2 (2): 7–33.