ASSESSING ALTERNATIVE TECHNOLOGIES FOR USE OF VOLUNTEERED GEOGRAPHIC INFORMATION IN AUTHORITATIVE DATABASES

BOTSHELO SABONE

December 2009

TECHNICAL REPORT
NO. 269

ASSESSING ALTERNATIVE TECHNOLOGIES FOR USE OF VOLUNTEERRED GEOGRAPHIC INFORMATION IN AUTHORITATIVE DATABASES

Botshelo Sabone

Department of Geodesy and Geomatics Engineering
University of New Brunswick
P.O. Box 4400
Fredericton, N.B.
Canada
E3B 5A3

December, 2009

PREFACE

This technical report is a reproduction of a thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Engineering in the Department of Geodesy and Geomatics Engineering, December 2009. The research was supervised by Dr. David Coleman, and support was provided by World University Service of Canada (WUSC).

Sabone, Botshelo (2009). *Assessing Alternative Technologies for Use of Volunteered Geographic Information in Authoritative Databases.* M.Sc.E. thesis, Department of Geodesy and Geomatics Engineering Technical Report No. 269, University of New Brunswick, Fredericton, New Brunswick, Canada, 117 pp.

ABSTRACT

Volunteered Geographic Information (VGI) has been enabled by advances in positioning, Web mapping, cellular communications and wiki technologies. These technological advances have allowed ordinary citizens to become producers as well as users of geographic information. Predictions have been made that VGI could be used to fill gaps in existing spatial databases, for example, complementing Spatial Data Infrastructure (SDI) datasets. However there are critical issues surrounding its production and possible integration which need to be addressed before considering it for use in complementing SDI datasets.

This thesis presents research which investigated the extent to which VGI enabling technologies affect its accuracy and ensures accuracy compliant with Canadian Geospatial Data Infrastructure (CGDI) accuracy standards. The research examines the suitability of VGI as a resource for augmenting authoritative datasets, like CGDI datasets, by assessing its positional accuracy and other data quality factors. Factors influencing the accuracy hence quality of VGI, e.g. Location Based Service (LBS) positioning techniques, are analyzed and a framework for integrating VGI into suitable authoritative CGDI datasets is developed. The framework is designed to provide a platform for validation and integration of VGI into authoritative databases. The framework's limitations and strengths are also analyzed.

ACKNOWLEDGEMENTS

I gratefully acknowledge the following people for assisting and encouraging me to complete this research project:

- Dr David J. Coleman, my supervisor, for his support, guidance and encouragement. I am very thankful for the patience and time he took to give constructive feedback, comments and suggestions while I was writing this thesis.
- Dr. Susan E. Nichols and Dr. Peter Dare for their helpful advice towards this research.
- Mr. Jay Woodyer and Titus Tienaah for their assistance during data collection, for answering countless questions and giving good advice.
- Nyaladzani Nkwanana, my colleague, for assistance and support, as well as providing advice and discussions/debates on the research.
- Sylvia Whitaker and Lorry Hunt who've always kindly assisted me with all administrative issues.
- My dear family in Botswana; my parents, sisters, brother and nephew, for their continuous encouragement, love and support.
- My fiancé, Eric Webster, for his constant love, understanding and encouragement.
- The Botswana International University of Science and Technology (BIUST) and GEOmatics for Informed Decisions (GEOIDE) for providing financial support for my studies (BUIST) and the research (GEOIDE).

Table of Contents

List of Tables

List of Figures

List of Symbols, Nomenclature or Abbreviations

AOA- Angle of Arrival

A-GPS – Assisted GPS

EOTD- Enhanced Observed Time Difference

CanVec- Canada Vector Database

CTI- Canada's Centre for Topographic Information

CGDI – Canadian Geospatial Data Infrastructure

COO- Cell of origin

DSE- Department of Sustainability & Energy (State of Victoria, Australia)

F-VGI – Facilitated- Volunteered Geographic Information

GIScience- Geographic Information Science

GPS – Global Positioning System

GSM- Global System for Mobility

LBS- Location Based Services

Location Management Units (LMU)

OSM- OpenStreetMap

PDA- Personal Digital Assistant

PGIS – Participatory GIS

PPGIS- Public Participation GIS

POI- Points of Interest

PND- Personal Navigation Devices

NES- Notification & Edit System

NRN- National Road Network

NTDB- National Topographic Database

RTK GPS - Real Time Kinematic GPS

SDI- Spatial Data Infrastructure

WLAN- Wireless Local Area Network

TDOA- Time Distance of Arrival

UL-TOA- Uplink Time of Arrival

VGI- Volunteered Geographic Information

WiMax- Worldwide Interoperability for Microwave Access)

Chapter 1 Introduction

The rise of user generated geographic information, termed Volunteered Geographic Information (VGI), has been enabled by the convergence and emergence of technologies which allow ordinary citizens to create and disseminate maps and spatial data. One of the main issues surrounding VGI is centered on the purposes for which the data may be used. A wide array of uses which comprise of social networking applications, personal navigation and community based Participatory GIS (PGIS) have been identified as possible uses of VGI (Elwood, 2007a; Flanagin and Metzger, 2008). However, predictions have been made that data collected by ordinary individuals could be used by researchers, policy makers, citizen groups and private institutions for a number of purposes ranging from monitoring environmental change to filling gaps in existing spatial databases (Elwood, 2008a). The idea of using VGI to fill gaps in existing spatial databases or complementing Spatial Data Infrastructure (SDI) datasets is an increasingly discussed topic for research. This can be ascribed to the concern about basic supply of geographic information, and trends affecting the processes by which it is acquired and compiled (Goodchild, 2007a).

An assessment of the potential impact VGI could have for an SDI, in terms of using it to augment relevant datasets, is an important step towards addressing the concerns about lagging map updating programs or limited supply of new and updated geographic information for existing datasets (Goodchild, 2007a). There are underlying assumptions that VGI, obtained through collaboration and crowd sourcing, could be useful to enhance the information used by professionals and decision makers (Seeger, 2008; Gouviea et al, 2004). The availability of VGI is also thought to have improved geographic information in several ways. For example, VGI can consist of multi-media data such as photographs and video, information that cannot be sensed remotely. It also relies on local knowledge or understanding, as well as information about current local conditions, and it can add useful contextual information to spatial data (Goodchild, 2007b; Flanagin and Metzger, 2008; Elwood, 2008a). Therefore VGI could enhance the spatial data used in SDI's, especially when it comes to data that is lacking or in need of updating. However, data integration is a key area of concern especially when it comes to dealing with data with different levels and/or combinations of positional and attribute accuracy, and completeness. These factors have great influence on the quality and characteristics of VGI and are important to take into account if it is to be used in helping populate and maintain an SDI.

Chapter 1.1 Research Background

Investigating the nature of VGI includes examining technologies such as location based services (LBS) infrastructure which have become an important instrument for facilitating VGI. Location based services provide users with spatial and contextual information about their location from a range of mobile devices. Other related technologies such as mobile cartography and TeleCartography can also be noted as VGI facilitors. TeleCartography describes the distribution of cartographic materials via wireless data transfer interfaces and mobile device (Gartner, 2007). Mobile cartography is described as the theories and technologies of dynamic cartographic visualization of spatial data and its interactive use on portable devices anytime anywhere under the consideration of the actual context and user characteristics (Gartner, 2007).

All together LBS, TeleCartography and mobile cartography can be considered under a single umbrella termed “ubiquitous cartography” (Gartner, 2007). Ubiquitous cartography refers to the use and creation of maps by users anytime and anywhere and is strongly influenced by the advances in information technology, such as the development of wireless systems, high density storage and broadband communication (Gartner, 2007). The mentioned technologies, in addition to others such as, interactive web services, positional technologies like GPS, geotagging, geoblogging etc. all contribute to the ability of users to use, create and disseminate spatial data. These developments have been used to define VGI and have also been referred to with a number of terms including ubiquitous cartography, wiki – mapping, web mapping, and neogeography (Elwood, 2008a; Turner, 2006).

The technologies that influence or enable VGI affect its nature and characteristics and are therefore important when assessing the effect or impact VGI will have on digital spatial data. Elwood (2008b) has cited three ways in which different researchers assume VGI might alter digital spatial data: its production and sharing, its content and characteristics, and the purposes for these data resources might be used. When considering the use of VGI in a spatial data infrastructure, all these factors need to be investigated taking into account the technologies responsible. Other factors including those of demands, specifications and standards for LBS or other enabling technologies should be taken into consideration. They should be compared to the demands, specifications and standards for SDI in order to assess the compatibility of the data.

Standards and specifications for data must be complied with in order to allow integration as well as sharing and dissemination of geographic information in an SDI environment. Thus analyzing and quantifying the accuracy and uncertainty of the spatial and attribute data -- as well as the accuracy and uncertainty regarding the processes of acquiring the data -- is essential for comparing the quality and characteristics of VGI with SDI datasets and identifying the opportunities and limitations for integration. Furthermore an evaluation and synthesis of characteristics, limitations, constraints and opportunities of LBS positioning technologies; VGI generation processes; and SDI functional and data requirements, needs to be undertaken in order to assist in identifying opportunities and limitations for integration.

Quality of VGI is a critical issue which will have an impact on its perceived value and use. Geographic data are often modelled as features with attributes, and are abstracted representations of the real world (Zhang and Goodchild, 2002). The quality of geographic data is measured as the difference between the data and the world they represent and becomes poorer as the data and corresponding reality diverge (Maue and Schade, 2008). "Quality of VGI" may include such aspects as: positional and attribute accuracy; currency (i.e. how up to date the data is); credibility of sources; and completeness of data. All these aspects are affected directly and indirectly by the technologies and processes used to acquire data. Particularly, the accuracy aspect of quality is directly affected by the positional technologies and methods in which data is collected and mapped.

Quality control mechanisms are tools used to ensure a minimum level of quality in data. Some rely on automated methods or thorough peer reviews by specialists (Maue and Schade, 2008). In the world of VGI, quality control is difficult to achieve due to the heterogeneous and dynamic nature of the data as well as the volume. There are some mechanisms for quality control which rely on peer reviews, such as wikis or rating systems. These methods may not be suitable for authoritative databases which rely on well known and mature methods for modelling and representing quality. However, they might be appropriate as an interim step until such time that the databases proprietor can verify the accuracy through more traditional means. Positional and attribute accuracy are of key importance when considering integration of data, as there has to be a high degree of consensus of these two quality aspects between the data that is to be merged. Therefore, there is a need to have quality control mechanisms which take these factors into account.

Several areas of Geographic Information Science (GIScience) research are concerned with the representation, manipulation and analyzing the qualitative, heterogeneous or shifting forms of VGI in a digital environment (Elwood, 2008a). Questions such as how multiple datasets can be integrated if the same semantics are used differently in each are being attempted. Solutions that are comprised of approaches that focus on enriching data with information that will help the user assess heterogeneity as they make decisions about integration, analysis and application are also being sought after (Elwood, 2008a). Kuhn - (2007) considered research induced by VGI from four perspectives that include technology, semantics, cognition and society. From a technological perspective the main issue he highlighted, similar to those posed by Elwood – (2008a), was how information could be integrated through locational reference at a scale that was never considered in traditional GIS and SDI architecture. The questions raised by this issue included how to characterize the quality of VGI; annotate VGI with useful metadata; discover pertinent VGI; and integrate VGI across multiple sources and with traditional information.

It is clear from the trend in research that VGI is expected to have an impact on GIScience. The type of impact and how significant it will be for geographic information markets, GIScience and society is the focus driving many research questions. For GIScience researchers the main concern is in understanding the implications of VGI's user interactivity and user generated content for GI Systems and GI Science (Kuhn, 2007). From a societal perspective, the emerging research agendas are on how VGI might foster new forms of surveillance and further erosion of privacy; enable new forms of activism; participatory democracy and civil life; or exacerbate existing inequalities and creating new forms of exclusion (Elwood, 2008b).

Most research focused on VGI data look at its production, content, quality, credibility, characteristics and how it can be used and integrated (Elwood, 2008b; Gartner, 2007; Goodchild, 2007a; Maue and Schade, 2008). There are gaps in research areas, however, that deal with integration of data in terms of quality control. Identifying methods of modeling and representing VGI quality in terms of accuracy, as a basis for integration of datasets, is a key area missing from research agendas at the moment. There has been a lot of research done on integration of data from a technological perspective especially in regards to VGI use in Participatory GIS. This type of research involving VGI use for PGIS is focused mainly on the technologies that are used to enable data integration through the web, such as Web 2.0 technologies and open source standards, which allow users to generate and share content via the internet (Gouviea et al, 2004; Seeger, 2008). It is clear though, that an interdisciplinary approach that investigates all the technologies which enable VGI is required in order to assess any aspect regarding its quality.

Chapter 1.2 Research Agenda and Objectives

The purpose of this research is to investigate the extent to which VGI enabling technologies, including LBS positional and data acquisition techniques, ensure compliance with Canadian Geospatial Data Infrastructure (CGDI) accuracy standards.

Chapter 1.2.1 Objectives

1. To identify and characterize the 3 leading LBS positional technologies and processes used for VGI generation 95% of the time, indicating their strengths and limitations.
2. To identify the most common VGI data types, available 95% of the time on 5 well known websites with more than 1 million entries.

3. To use the identified data types to develop criteria which assess VGI quality in terms of positional and attribute accuracy and uncertainty, which ensures compliance with CGDI data accuracy standards 90% of the time.

4. To test a small sample of existing and newly created VGI at Canadian local, provincial and federal levels against that criteria and use them to compare with a small sample of existing CGDI data and accuracy standards.

4. To establish a framework which seeks to implement the use of VGI in a SDI, using mechanisms that ensure data accuracy compatibility.

Chapter 1.3 Significance and Contributions

The overall goal of the research objectives is to characterize the quality of VGI from a technical perspective that considers positional and attribute accuracy. This research will assess a combination of technological processes and data characteristics based on accuracy and uncertainty to determine quality. It also provides test bed to fully assess characterizations of data quality that are made. This research aims to contribute to the work on seeking solutions that can help determine the suitability of VGI for integration across multiple sources and with traditional SDI data.

This research addresses the gaps in existing research dealing with integration in terms of quality control. Existing research focuses on integration through the Web; however, this research is focused on VGI quality in terms of accuracy, as a basis for integration.

The contributions of this research include the evaluation of VGI quality aspects, (e.g. positional accuracy), which allow the quality of VGI and its suitability for integration with traditional SDI data to be determined. The other contribution includes the framework of integration for validating and integrating VGI. It assesses methods for validating and integrating spatial data, as well as current methods of validating and integrating VGI with other data. The framework aims to provide a holistic and best fit solution for VGI integration.

Chapter 1.4 Thesis Outline

This thesis is organized into five chapters. This chapter introduces the background, objectives, and discusses the significance of the research. Chapter 2 reviews literature which places the research in context. An in-depth look into VGI types and uses; VGI facilitating technologies, including LBS positioning techniques and Web 2.0; and the role/use of VGI in a SDI is provided. CGDI datasets standards and specifications are also reviewed for the purposes of developing criteria.

Chapter 3 discusses the methodology of carrying out the research objectives. The development of criteria from CGDI datasets standards and specifications and methods used to evaluate VGI are outlined. The framework for integration is presented, as well as the process undertaken in designing it.

Chapter 4 presents the results of the evaluation of VGI and highlights the limitations and strengths of the integration framework. The performance of VGI against the criteria is also analyzed in this chapter.

Finally, chapter 5 discusses the research outcomes and issues encountered. Recommendations for future research are provided as well.

Chapter 2 VGI; Facilitating Technologies; and the Potential Role/Use of VGI in SDI

Chapter 2.1 Introduction

In this chapter the nature and characteristics of volunteered geographic information (VGI) will be discussed in terms of its possible impact on spatial data infrastructures (SDI). The chapter draws on background information from different research domains which include VGI, LBS, SDI, and spatial data quality and integration. The chapter sets out to define VGI and describe existing types and uses of VGI, as well as the value it is believed to possess. A review of technologies that facilitate VGI is also included, with focus on LBS positioning techniques and Web 2.0. The review highlights the strengths and weaknesses of LBS positioning techniques and Web 2.0, as well as their impacts on VGI production.

The chapter goes on to describe VGI in the SDI context by first outlining the concept and structure of a SDI. It uses the Canadian Geospatial Data Infrastructure (CGDI) as a specific example to illustrate the structure and components of a SDI. Furthermore, the example is used to demonstrate how data accuracy standards and specifications play an important role in determining the suitability of use of VGI within a SDI. The chapter then discusses the implications of SDI's using VGI and how it can be integrated with other datasets. Finally, the chapter concludes with a summary of the main topics and issues associated with the use of VGI within a SDI.

Chapter 2.2 VGI

VGI can be considered a subset of the more general Web phenomenon of user generated content (IAB, 2008; OECD, 2007; Goodchild, 2007a and others.). It represents a different approach from the traditional way of collecting and disseminating geographic information. The collection, creation and dissemination of geographic information are traditionally reserved for official mapping agencies (Goodchild, 2007a). VGI diverges from official agencies with trained and qualified professionals to ordinary untrained citizens collecting, creating and disseminating geographic data.

The mapping agencies rely on top-down organizational models which preserve a distinction between geographic data providers and consumers (Kuhn, 2007). However, VGI represents bottom-up organizational models which blur the distinction between data providers and consumers with the "consumers" becoming "providers" as well. This role reversal has been enabled by technologies that allow users to collaborate in creating and sharing geographic data. Bruns (2006) proposes that the term "Produsage" be used to describe collaborative user-led content creation. (See also (Budhathoki et al, 2008).) Produsage as a concept stands in direct contrast to traditional modes of industrial production and is built on iterative, evolutionary development models (*ibid*). It involves communities of participants who usually make very small, incremental changes to the established knowledge base where knowledge always remains in the process of development and information is always unfinished, extensible and evolving (*ibid*).

VGI has the characteristics of produsage. Other examples where produage occurs and its characteristics found, include open source software development, citizen journalism and Wikipedia (Bruns, 2008). In all the listed examples, users engage in content creation that's characterized by collaborative construction and maintenance of both content and the social relationships among participants (*ibid*)

The interest in VGI for use in SDI's raises major concerns about its suitability for use by professionals since in most cases it wasn't created for professional use. More importantly, the issues regarding quality of data and its validation and integration impact significantly on its perceived value. However, the fact that there is considerable research into uses of VGI and how to integrate it with other data suggests that it could cater to some professionals' data needs.

Chapter 2.3 Types of VGI

There are different kinds of spatial data being disseminated by ordinary citizens. From linear features like roads on websites like *OpenStreetMap* (OSM) to place names and points of interest to the citizens. *Wikimapia*, *The People's Atlas* (Platial.com), *Wayfaring*, *Tagzania* and *OSM* are few of the many existing examples of VGI initiatives. The previously mentioned sites can be categorized into different groups: wikis (*OSM* and *Wikimapia*), where anyone can contribute and modify content, or social networking sites (*Wayfaring, Platial.com* and *Tagzania*).

The goal of wikis is to create collaborative websites and where all users have a stake in the quality of data. Therefore, having accurate data is a concern for all users. The concept dubbed Linus' Law is believed to apply to wikis. It states:

"*Given enough eyeballs, all bugs are shallow*," (Shirky, 2008).

It implies that, given enough people to solve a problem, the quicker and easier it is to find a solution. For example, an explanation on the OSM website states that:

"If one person puts in inaccurate data, maliciously or accidentally, the other 99.9% of people can check it, fix it, or get rid of it," (OpenStreetMap, 2009).

They believe that vast majority of good-intentioned participants can automatically correct for the few bad contributions. The *Wikipedia* project has been exemplified by Goodchild, (2007a); Bruns (2008); Shirky (2008) and OpenStreetMap (2009), among others, to show that a large amount of good quality data can be collected *but* it can also be difficult to weed out the inevitable errors.

The social networking sites are more focused on connecting people who share interests and/or activities, or who are interested in exploring the interests and activities of others (Bruns, 2008). They also offer collaborative map making of features or items and activities that are of interest to the users. Example can be found on *Wayfaring.com* where users can create maps of lives and *Platial.com* where members of the social network can find out who and what is nearby. Table 2.1 gives a more detailed description of the sites' functions and data types they have.

Table 2.1 Sites with different types of VGI

Website	Type	Function	Data Types
OpenStreetMap	Wiki	• Free editable map of the world • Users collaborate to build Road/street network	Streets/Roads P.O.I
Wikimapia	Wiki	• Online editable map using satellite imagery • Users collaborate to describe places and provide place names	Bounding boxes Polygons
Tagzania	Social Networking	• It is about an individuals' places • Allows users to build & keep individual personal maps using personal tags • Users collaborate to make maps using common tags	P.O.I Lines Polygons
The People's Atlas (Platial.com)	Social Networking	• Gives a guide to who and what is nearby • Map making, map searching and map exploration site	P.O.Is annotated with pictures and text
Wayfaring	Social Networking	• Creating and sharing personalized maps by users in social network • Allows collaboration with others to build new maps	P.O.Is annotated with pictures and text

Chapter 2.4 Uses of VGI

The definition of VGI implies that, for contributed data to be considered "VGI", volunteers must instigate the documentation of the spatial features of their own accord (Seeger, 2008). This spontaneous creation of VGI is predominant in the social networking arena and some wikis as people create and share spatial data of their interests and/or activities within an online community. While this form of VGI may be of use and interest to individuals or communities, a directed or facilitated approach to VGI creation may be more appropriate to suit the needs of professionals. Seeger (2008) terms this variant of VGI, facilitated-VGI or f-VGI. He differentiates an f-VGI approach from other VGI practices because of the way in which the collection of volunteered information is shepherded by a facilitator. This information might be contributed in response to a predefined set of criteria, such as an explicitly defined question, or limited to an established geographic extent (Seeger, 2008).

F-VGI builds upon the established principles of Public Participation GIS (PPGIS) and can be used by local governments, state agencies, community based organizations, special interest groups, private companies, etc (Sieber, 2004; Seeger, 2008; Elwood, 2008a; Bruns, 2008). However whether VGI is spontaneous or facilitated it still involves the engagement of private citizens collaborating to contribute spatial data, thus the issues of spatial data quality still remain. So even if the f-VGI approach is adopted by professionals, they still have to deal with the inherent issues associated with spontaneous VGI such as, accuracy of data, credibility of contributors and data validation. Table 2.2 describes a few programs and examples of uses of VGI by professional individuals and organizations.

Table 2.2 Examples of uses of VGI by Professional Individuals and Organizations

VGI Initiative	Description	Sector
Senses @Watch	A research project intended to create and evaluate strategies promoting the use of VGI collected through citizens' senses. (Gouveia et al, 2004; Gouveia and Fonseca, 2008)	Academia
The GLOBE Program	Worldwide network of students, teachers, and scientists working together to study and understand the global environment. Students, teachers and scientists collaborate on inquiry-based investigations of the environment and the Earth system. http://www.globe.gov/r/	Interagency program (funded by NASA, NSF and US State Department)
MilWikiKB	Military command and control wiki knowledge base designed to use contributed data to gain good situational awareness about physical locations of objects of interest and their conditions. (Brannstrom and Martenson, 2006)	State Agency - Military
TomTom Mapshare™	The TomTom MapShare service allows customers to update, add to and delete from their device's map data at any time and also share changes with the TomTom community. http://www.clubtomtom.com/general/get-to-know-tomtom-mapshare%E2%84%A2/	Private Sector
Mapping Vernal Pools	Volunteers participated in state wide mapping of vernal pools connecting with the State Department of Environmental Protection and Rutgers University through a web mapping interface.(Lathrop et al, 2005) http://www.dbcrssa.rutgers.edu/ims/vernal/viewer.htm	Local Government (New Jersey)
Christmas Bird Count	Volunteers throughout the Americas take part in a traditional annual bird counting mission. Audubon and other organizations use data collected in this wildlife census to assess the health of bird populations, and to help guide conservation action. http://www.audubon.org/Bird/cbc/	Special Interest Group

Chapter 2.5 Technologies Facilitating VGI

Volunteered geographic information has been brought about by several technologies which previously existed for separate specific applications. For example, LBS technologies of spatial positioning and mobile communication can exist as stand alone technologies serving separate and specific applications (Gartner et al, 2007, Stojanovic et al, 2001. However, because they have merged to enable LBS they now share the capability of generating VGI. Evolving web technologies resulting in Web 2.0 also play a significant role in VGI production (Goodchild, 2007b; Elwood, 2008b; Rinner et al, 2008). Investigating these technologies provides a useful insight on the platforms on which VGI is built as well as how it is produced. This will also provide an understanding of limitations and strengths of VGI by examining the capabilities, strengths and limitations of these technologies.

Chapter 2.5.1 Web 2.0

Web 2.0 technologies provide an environment that better enables seamless information sharing and communications, social networking, and mass collaboration (Goodchild, 2007b; Elwood, 2008b; Rinner et al, 2008). User generated content, including VGI, is a core feature of Web 2.0 applications (Rinner et al, 2008). Web 2.0 applications allow bi-directional collaboration in which users are able to interact and provide information to central sites, and to see that information collated and shared with other users (Goodchild, 2007b). Interactive geo-visualization interfaces such as Google

Maps or Microsoft's Virtual Earth make it possible for nearly anyone with an Internet connection to disseminate their own maps and geographic information (Elwood, 2008b).

The collaboration to produce and share geographic information are capabilities directly induced by the Web 2.0 environment and technologies. It influences ways in which VGI can be created, modified and used. This can be useful for organizations such as mapping agencies, wishing to utilize VGI because they can use or develop Web 2.0 applications to suit their own requirements. Furthermore, the collaborative nature in which VGI is produced by such applications can assist with data validation; therefore measures of quality control can be established. This is a major strength of Web 2.0 technologies. However, on the downside, mapping agencies and other official organization cannot simply rely on users collaborative efforts for data production and especially not for data validation. In addition, the top-down approaches that mapping agencies use to manage their spatial data may conflict with bottom-up Web 2.0 approaches to manage VGI. Striking a balance between the different approaches may be essential to utilizing VGI as resource.

Chapter 2.5.2 Location Based Services

Whilst Web 2.0 applications provide users the capabilities of creating maps and mashups, advances in wireless communication infrastructure and spatial positioning technologies have enabled users to capture and collect geographic data in real world environments. A growing number of mobile phones, PDAs, digital cameras and other handheld devices use GPS positioning to provide users with information based on their location, or to add locational information to other media (Smith et al, 2002; Gartner et al, 2007; Elwood, 2008a). Location based services (LBS) , Tele-Cartography and Mobile-Cartography are examples of technologies which provide users with information based on their location and enable them to create and use maps anytime anywhere (Gartner et al, 2007). All together these technologies can be considered under the single umbrella termed “ubiquitous cartography” (Gartner et al, 2007).

LBS are essentially enabled by (1) mobile communication infrastructure, (2) wireless internet, (3) spatial positioning techniques and (4) Web- GIS. LBS deliver geographic information and geo-processing power to mobile and static users via the Web and mobile network in accordance with location of users (Stojanovic et al, 2001). The spatial positioning techniques component of LBS influence the accuracy of VGI generated via these means and shall be discussed next. This is not to say that the other components are not important; because it is the convergence of all four components which make LBS possible. However, they have less impact on the quality of VGI created from LBS. More information on the mobile communication infrastructure, wireless internet and Web-GIS components of LBS cab be found in Appendix I.

Chapter 2.5.2.1 LBS Spatial Positioning Techniques

Positional techniques used by LBS are network based, terminal/handset based, and hybrid in nature. Network based techniques use the cellular network to determine the location of the mobile device. Handset based techniques use the radio navigation system provided by GPS satellites (Unni and Harmon, 2002). Hybrid techniques use both the network and GPS system. Additionally there are short range positioning techniques, such as WLAN or (Wi-Fi) and Bluetooth. These technologies are limited in their coverage and more suitable for indoor and small area coverage (Unni and Harmon, 2002). The positional techniques employed by LBS for facilitating VGI are a key focus of investigation in this research. In particular, better understanding of the respective accuracies offered by these techniques for pinpointing locations of users is a principal objective of this research. Knowing the positional accuracies offered by LBS positional techniques will aid in assessing the accuracy of VGI created by LBS.

- **Terminal/Handset based Positioning – GPS**

The terminal/handset based positioning techniques employed by location based services relies on GPS which can used anywhere in the world. The systems' satellites transmit navigation messages that contain their orbital elements, clocks and statues, which a GPS receiver uses to determine its position (Barnes, 2003). Determining the receivers' longitude and latitude requires at least three satellites, and adding a fourth can determine its altitude (Barnes, 2003). This positioning technique requires users to purchase mobile devices that have in-built GPS receivers in order to use the system. The accuracy of GPS can be less than 10 m outdoors and up to 50 m in urban canyons (D'Roza and Bilchev, 2003; Barnes, 2003; Dao et al, 2002; Reed et al, 1998; Unni and Harmon, 2002)

- **Network Positioning**

Network based positioning techniques exploit the cellular network infrastructure to obtain geographic location. The most common network-based solutions for mobile positioning are: Cell of Origin (COO), Time of Arrival (TOA) and Angle of Arrival (AOA) (Reed et al, 1998; Unni and Harmon, 2002; Smith et al, 2001). COO is the most straightforward cellular location solution and works by detecting the base stations with which the mobile device is registered (D'Roza and Bilchev, 2003). It provides low accuracy, typically up to 500 m in urban areas and up to 15 km in rural areas (D'Roza and Bilchev, 2003). TOA uses the differences in the time of arrival of the signal from a user's mobile device to at least three base stations and calculates the location of that device (Dao et al, 2002). The AOA technique seeks to determine the location of a mobile device based on the angle at which signals transmitted from the device arrive at the base stations. A minimum of two base stations is required to calculate location. This technique offers accuracy not less than 150 m (Unni and Harmon, 2002). There are variations of TOA and AOA positioning techniques such as Time Distance of Arrival (TDOA) and Uplink Time of Arrival (UL-TOA) which offer better accuracies than their respective original techniques; these typically vary from 50 m to 150 m (Swedberg, 1999).

- **Hybrid Positioning**

Hybrid techniques use a combination of network-based and handset-based technologies for location determination. An example is Assisted GPS (A-GPS) which uses a GPS positioning technique whereby there is assistance data provided from a special GPS server/base station by the mobile network (Barnes, 2003;Unni and Harmon, 2002; Dao et al, 2002). Another example is Enhanced Observed Time Difference (E-OTD) technique which determines the location of a mobile device using location receivers that are geographically dispersed across a wide area (Dao et al, 2002).

- **A-GPS**

A-GPS enables GPS positioning even in urban and indoor areas where the signal is too weak to be acquired using standard signal tracking procedures within the receiver (Barnes, 2003; Unni and Harmon, 2002; Dao et al, 2002). For example, the approximate location information of the GPS-enabled handset (derived from the COO technique) can aid the tracking of the satellite signals, and the ephemeris data (transmitted to the mobile device from a GPS base station receiver) can permit fast position computation even in a so-called cold start (Dao et al, 2002) . Sometimes the actual measurement and position computation is done not in the handset but at a location server integrated within the mobile telephone network (Dao et al, 2002). The process takes just a few seconds, whereas conventional GPS receivers can take many minutes (if at all) (Barnes, 2003; Dao et al, 2002). Conventional GPS receivers may take longer to obtain a position fix due to factors such as minimal satellite coverage or low satellite signal strength. Hybrid-based techniques give the best accuracy (less than 10 m) of the three types of handset-based technologies (Barnes, 2003; Swedberg, 1999; Dao et al, 2002).

- **E-OTD**

The E-OTD technique uses geographically dispersed location receivers, known as Location Management Units (LMU), which each have an accurate timing source (Dao et al, 2002). E-OTD (software-enabled) mobile devices and LMU's triangulate at least three base stations to measure and compare the arrival time of a signal from a user and location is computed by the handset (Unni and Harmon, 2002; Dao et al, 2002). The E-OTD technique offers an accuracy level from 50 up to 125 meters (Unni and Harmon, 2002; Dao et al, 2002; Swedberg, 1999).

- **Short Range Positioning**

Wireless Local Area Networks (WLAN or Wi-fi) and Bluetooth technology also have the potential as platform enablers for LBS, providing services such as personalized mobile advertising and promotion (Unni and Harmon, 2002). WLAN provides wireless access to the internet whilst Bluetooth enable devices in short range to communicate. These technologies are limited in their coverage and as such cannot be viewed as competing with network based or GPS technologies; however, they are good for coverage in small areas (*ibid*). Internet Protocol based networks such as, WiMax (Worldwide Interoperability for Microwave Access), surpasses Wi-Fi in terms of coverage and data speed (*ibid*). These are still in the early stages of deployment but they promise to deliver ubiquitous internet coverage. Being able to offer ubiquitous coverage will enable them to compete on the same level with GPS and networked based technologies. This is because they overcome the shortfalls of Wi-fi, where users access the internet from Wi-fi “hot spots” that cease internet connectivity when the user moves away from them.

Spatial positioning techniques have a direct impact on the quality of data, particularly the positional accuracy of VGI generated via LBS. Therefore, it is important to know and understand the context in which spatial positioning techniques are employed. By examining existing uses and applications of LBS we can obtain information on which uses or applications use techniques offering high accuracies and why, as well as which uses and applications require low accuracies. This helps to categorize existing services and gives an indication of the types of services that could provide high or low data accuracies if they were used to generate VGI. Further discussion of LBS uses and applications follows in Section 2.5.3, and Table 2.3 summarizes LBS positioning techniques accuracies and limitations.

Table 2.3 Summary of LBS Positioning techniques accuracies and limitations

Positioning Technique	Accuracy	Limitation
Terminal Based		
GPS	High accuracy outdoors (<10m), low accuracy in “urban canyons” (up to 50m).	Line of sight issues, Handset modification required (in-built GPS receiver chip)
Network based		
Cell of origin (COO)	500 -1500m	Low Resolution and accuracy
Angle of Arrival (AOA)	< 150m	Medium resolution, Expensive network modifications required
Hybrid technology		
Enhanced Observed Time Difference (E-OTD)	50 - 125m	Suited for GSM only. Network and handset modification required. Cell coverage necessary.
A - GPS	10 - 20m	Significant changes to network and handset modifications required

Chapter 2.5.3 LBS Uses and Applications

Several authors, including Unni and Harmon (2001), Bernados et al (2007) and Barnes (2002) have cited different areas of application of LBS. To date, safety (emergency services), navigation, tracking, transactions or location based charging, and information or content delivery have constituted the majority of LBS applications.

Many of these applications have opportunities to generate and use VGI. In some instances, like navigation, VGI has already made successful strides. OSM is an example where LBS navigation applications have allowed users to collect road network information from their Personal Navigation Devices (PND) and smart phones, in an effort to collaboratively map the world with road network information. Other areas of applications that may contribute to VGI include location enhanced communication and social networking; where extended ways of interacting among individuals are made possible by adding location aware capabilities to common services like instant messaging and push-to-talk (Bernados et al, 2007). Examples of location services and their typical accuracy and positioning requirements can be found in Table 2.4.

Table 2.4 Typical Positioning Requirements for location services (Barnes, 2003)

Area of Use	**Application**	**Purpose**	**Typical Accuracy & Positioning Requirement**
Safety	Emergency Services	Obtain help from emergency services	High (GPS/A-GPS)
	Roadside assistance	Obtain breakdown assistance	Medium (EOTD)
Navigation and Tracking	Vehicle navigation	Reach destination	High (GPS/A-GPS)
	Fleet management	Manage fleet resources	High/Medium(A-GPS/EOTD)
	Asset tracking	Locate and direct assets	Low (COO)
	People Tracking	Locate and direct people	Medium/Low (EOTD/COO)
Transactions	Location Sensitive billing	Competitive pricing	Medium/Low (EOTD/COO)
	Zone based traffic calming	Automatic pricing of road usage	Medium (EOTD)
	Cross-selling	Sales of products and services	High/Medium(A-GPS/EOTD)
Information	Locational advertising	Targeted advertisement	Medium/Low (EOTD/COO)
	Public information	Provide public information	Medium/Low (EOTD/COO)
	Geographic messaging	Localized information and alerts, and Social Networking	Medium/Low (EOTD/COO)
	Yellow pages	Find proximity of something specific	Medium/Low (EOTD/COO)

Chapter 2.6 Potential Role/Use of VGI in SDI

The interest, benefits and challenges of using VGI in augmenting authoritative datasets have been discussed in the preceding sections. However, the exact context in which it will be used was not framed. In the following sections, the context of VGI usage will be framed around the Canadian Geospatial Data Infrastructure (CGDI). SDI components and structure will be described, and then a background on CGDI and its data accuracy standards and specifications will be provided. This serves to illustrate the requirements CGDI will have for incorporation of VGI with its datasets. Furthermore, it provides the basis for criteria which can be used to evaluate VGI under consideration for incorporation. Implications of using VGI in CGDI are also discussed; with a focus on the impact of using VGI could have on CGDI's data, existing collection and updating methods, data integration methods and other components or structure. Methods for data integration are briefly examined, and how VGI could be integrated and applied in existing methods is also discussed. This provides useful information for the framework that could see VGI integrated with authoritative datasets.

A Spatial Data Infrastructure consists of the policies, institutional framework, technical standards, fundamental datasets, and clearinghouse networks to make information available to users (ANZLIC, 1996). The Canadian Geospatial Data Infrastructure (CGDI) initiative is an inter-governmental (federal, provincial and territorial), private sector and academia effort to provide easy, consistent and harmonized access to geographic information and services (Labonte et al, 1998). Standards are a key component of any SDI as they facilitate the sharing of geospatial data. CGDI Geospatial Standards ensure compatibility of datasets from disparate sources and promote technical interoperability which allows users to access data and services (GeoConnections, 2009). In order for VGI to be utilized as a resource for the CGDI (or any other SDI) it must comply with the standards.

Chapter 2.7 CGDI Data Accuracy Standards and Specifications Background

The CGDI has the following five inter-related technical components:

1. ***GeoExpress- Data Access*:** Access and discovery of metadata, images, files, and database query and extraction;

2. ***Geospatial Framework***: Data alignment layer for simplified integration, and data framework for application development ;

3. ***Geospatial Standards*:** Foster harmonization and expanded usage of international standards for geospatial data collection, description, quality, access and dissemination;

4. ***Partnerships***: Foster cooperation to collect, build, share, and maintain the geospatial data);

5. ***Supportive Policy Environment***: Foster access, lower cost, common licensing, and other activities to facilitate wider use of the geospatial data (Labonte et al, 1998).

The geospatial framework and standards components are important for this research, as they will be used to determine the suitability of VGI. Data quality standards will be used to assess the accuracy of VGI and how well it complies with the standards while the framework data will be compared with VGI datasets.

- **Geospatial Standards**

The CGDI uses existing standards and specifications such as those from the United States Federal Geographic Data Committee (FGDC), the International Organization for Standardization (ISO) 19100 series, and the Open GIS® Consortium Inc. (OGC) standards and specifications (CGDI Technical Guide, 2004). In this way, the CGDI is able to integrate and function with other federal, provincial, territorial, municipal and industrial geospatial infrastructures and initiatives throughout Canada and abroad (CGDI Technical Guide, 2004).

- **Geospatial Framework – framework data**

The data framework component is fundamental to the CGDI. Its goal is to make it easy to integrate and use data from different agencies and disciplines across Canada (Labonte et al, 1998). This framework, using standards previously described, ensures that the enormous volume of existing data can be accessed and used effectively, and also provides the digital foundation that data producers and users will use to collect and maintain data (*ibid*).

The CGDI promotes sharing and compatibility of geospatial data by defining a common set of framework data (CGDI Technical Guide, 2004). Framework data is the set of continuous and fully integrated geospatial data that provides context and reference information for the country (*ibid*). It is expected to be widely used and generally applicable, either by underpinning or enabling most geospatial applications. Framework data from GeoBase will be used as a basis for comparing "CGDI datasets" with VGI to carry out one of the objectives of this research. This data will be used because it is easily accessible and it is road data which is easy to collect as VGI and has already been collected by other volunteers, thus making it easy to compare the data. The data product specifications from the framework data, particularly the data quality specifications, will inform the criteria for which the comparison is based on. Framework data takes three principal forms:

1. ***Alignment layers*** include geometric controls required to adequately position geospatial information.
2. ***Land feature layers*** contain well-defined and readily observable natural or man-made physical features that are not subject to interpretation or speculation. These layers include many of the same features that are visible on topographic maps, such as roads, rivers and elevation.
3. ***Conceptual layers*** are the frameworks that society develops and uses to describe and administer the country. These layers complement a vast amount of application specific data. They are often interpreted from observations of physical, economic or social factors, and include features such as municipal boundaries, federal electoral districts and ecological areas (Labonte et al, 1998).

Chapter 2.7.1 Framework Data standards and specifications

Framework data standards and specification used to inform criteria for assessing VGI come from the National Road Network (NRN), National Topographic Database (NTDB) and CanVec data products. NRN data was also used to compare the accuracy of VGI with the accuracy of a CGDI dataset.

- **NRN**

The National Road Network (NRN) product contains road data for Canada. It's a result from the GeoBase initiative. It is distributed in the form of thirteen provincial or territorial datasets and consists of two linear entities (Road Segment and Ferry Connection Segment) and three punctual entities (Junction, Blocked Passage, Toll Point) with which is associated a series of descriptive attributes such as, among others: First House Number, Last House Number, Street Name Body, Place Name, Functional Road Class, Pavement Status, Number Of Lanes, Structure Type, Route Number, Route Name, Exit Number (NRN Data Product Specifications, 2007).

- **NTDB**

The contents of the NTDB largely correspond to that of the topographic maps in the National Topographic System (NTS). The NTDB contains information relating to 112 entities in 14 themes (Standards and Specifications of NTDB, 1996). It comprises of digital vector datasets that cover the entire Canadian landmass. Geomatics Canada has digitized and structured thousands of topographic maps, and includes features such as watercourses, urban areas, railways, roads, vegetation, and relief (GeoGratis, 2009). The organizational unit for the NTDB is the National Topographic System (NTS), based on the North American Datum of 1983 (NAD83) (Standards and Specifications of NTDB, 1996). Each file (data set) consists of one NTS unit at either the 1:50,000 or 1:250,000 scale (GeoGratis, 2009). The data is now available by themes within a file (GeoGratis, 2009).

- **CanVec**

CanVec is a digital cartographic reference product produced by Natural Resources Canada. It is a multi-source product coming mainly from the National Topographic Data Base (NTDB), the GeoBase initiative (www.geobase.ca) and the data update using Landsat 7 imagery coverage (CanVec Data Product Specifications, 2007). CanVec product contains more than 90 topographical entities thematically organized into 11 distribution themes: Administrative Boundaries, Buildings and Structures, Energy, Hydrography, Industrial and Commercial Areas, Places of Interest, Relief and Landforms, Toponymy, Transportation, Vegetation and Water Saturated Soils (CanVec Data Product Specifications, 2007). Each topographical entity is described by a name, a definition, a list of topological relationships, a list of attributes and a spatial component than could be a point, a line or an area (polygon) (CanVec Data Product Specifications, 2007). The data quality specifications of the different products are outlined in Table 2.5.

Table 2.5 Outline of Data Quality Specifications for GeoBase NRN, NTDB and CanVec products (CanVec Data Product Specifications, 2007; Standards and Specifications of NTDB, 1996; NRN Data Product Specifications, 2007)

	Data Products		
Data Quality	**GeoBase NRN**	**CanVec**	**NTDB**
Positional Accuracy	Planimetric accuracy aimed for the product is ±10 m or better in clear, unobscured areas.	Planimetric accuracy: ±10m for roads or ±25m or better for other features in clear, unobscured areas.	Planimetric Accuracy: ±30 m or better in clear, unobscured areas.
Thematic Accuracy	Unknown	Thematic accuracy and Completeness assessment is performed during data production process. The validation method applied depends on the data source used. If error detected is less than 5%, then all datasets in the batch are normally considered acceptable.	
Completeness	NRN contains: geometric and attributive description (current, accurate, consistent). Evaluation of errors of commission and omission conducted by data providers.		
Logical Consistency	The spatial relations of the entities of NRN datasets are systematically validated by means of in-house software.	Rules of the CanVec conceptual schema used to validate consistency between the conceptual schema and the CanVec product.	
Temporal Accuracy	Not applicable	Not applicable	Not applicable

Chapter 2.8 Implications of VGI use in SDI

The most serious issue associated with utilization of VGI is that of the data quality. Without appropriate mechanisms to validate data, VGI may well become a liability rather than an asset. However, in reasonably well-defined data environments -- such as exists with the address-matched street and road networks-- data validation mechanisms could be established which may provide quality control (McDougall, 2009). Methods for testing data quality differ and depend on the data type. They can also provide insight on how validation mechanisms can be established for VGI.

Using VGI to augment CGDI datasets implies that their data quality requirements must be met. Therefore, CGDI data quality standards and data product specifications may provide a basis to establish quality control. However, the specifications for CGDI datasets can be complicated, e.g. completeness and thematic accuracy specifications involve using different kinds of validation methods depending on data source. Positional accuracy specifications are the clearest - they stipulate minimum tolerances which data products must meet, and can guide criteria used to select VGI data for augmenting CGDI datasets.

Other implications of VGI for use in CGDI are concerned with how it will affect existing datasets (increasing positional accuracy inconsistencies in datasets) and the effect it has on usual validation and integration processes; how much time and resources it takes to effectively validate and integrate VGI with other datasets. The other impact VGI could have for CGDI as a whole is in engaging more citizens to become familiar with its data and services, as well as using them; through collaboratively providing data to government portals. It would give the CGDI a more user-centric feel than that which already exists.

Issues concerning how VGI will be validated and integrated, considering its dynamic and heterogeneous nature also need to be addressed. Existing methods for validating and integrating traditional map data may not apply to VGI because of the differences in the data structures and because they may require more human intervention (i.e. they cannot be performed automatically). Efficient processes for validation and integration of VGI are important; VGI should not become a burden for an organization.

Data integration is commonly used for dataset maintenance; typically, new datasets are integrated with existing datasets to update them with new information or to improve their quality (Chen et al, 2003). For example, updating an old street network dataset (from vector dataset) could be achieved through integrating it with new street information (from high resolution imagery). However, accurately integrating data from different sources is a challenging task. This is because spatial data obtained from various sources can have different projections, different accuracies and different formats (e.g. raster or vector format). Many different methods have been developed to integrate different types of data and others were developed to integrate specific types of data (e.g. integrating vector data with orthoimagery (Chen et al, 2006)). However, from the methods investigated, conflation and derivatives of conflation techniques seem to be the most common method used to integrate various types of datasets.

Conflation describes the integration of spatial data from different sources (Saalfeld, 1988). Conflation may be applied to transfer attributes from old versions of feature geometry to new more accurate versions; to the detection of changes by comparing images of an area for a number of different dates; or to automatically register one dataset to another based on the recognition of common features (Jensen et al, 1998).

Conflation works by first identifying corresponding features, or control points, in the data sources to be integrated (Saalfeld, 1988). The second step is to bring the dataset in alignment using rubber-sheeting transformations based upon these matching control points (Cobb et al, 1998). Rubber-sheeting transformations preserve topology and align datasets with each other - which is key to successful matching and merging of map features (Saalfeld, 1988).

Conflation methods could be applied to VGI, for purposes of integrating it with other datasets. However, the characteristics of a VGI dataset are very different than those of a dataset from a mapping agency. For example, a VGI dataset's positional accuracy may be unknown and the dataset may only possess geometric information with little or no attributes. This may increase the difficulty of using existing conflation methods and decrease the effectiveness of the integration. However, capabilities offered by conflation, such as, transferring attributes from old versions of feature geometry to new more accurate versions, could be utilized effectively in cases where there is sufficient geometric information and minimal attributes in a VGI dataset. The main issue with VGI is how to determine whether it can be used in integration with existing datasets, i.e. is it useful, and subsequently how to integrate it with existing datasets without compromising the original datasets quality. Because VGI also has characteristics of being dynamic and evolving over time, the integration method used should take this into account.

The background information which places this research in context was provided in this chapter. The characteristics, types and uses of VGI were discussed, and were followed by a review of VGI facilitating technologies. Web 2.0 and location based service technologies were investigated, with more emphasis on LBS spatial positioning techniques. This is because they are directly responsible for the accuracy of VGI created via LBS. The chapter concluded with discussions on the implications of using VGI in a SDI, focusing on the CGDI as an example. CGDI data accuracy standards and specifications were described to illustrate the requirements CGDI will have for incorporation of VGI with its datasets. And lastly data integration methods were reviewed, to provide insight on how VGI could be integrated with authoritative datasets.

The next chapter describes the methodology adopted in order to meet research objectives and solve the research problem. Criteria development, methods for testing the accuracy of VGI, and the processes undertaken to carry out the research are described. Furthermore, a review of methods which could be used to integrate VGI with authoritative datasets is given. And the integration framework, derived from these methods, is presented.

Chapter 3 Investigating VGI geometric properties

Chapter 3.1 Introduction

The goal of this chapter is to describe the methods that were used to meet the objectives and ultimately answer the research question “To what extent do VGI enabling technologies, including LBS positional and data acquisition techniques, ensure compliance with CGDI accuracy standards?” The main objectives are: (1) Identifying and characterizing three leading LBS positional technologies and most common VGI data types; (2) Developing criteria which assess VGI quality in terms of positional and attribute accuracy; (3) Testing VGI against criteria and comparing with CGDI data and accuracy standards/specifications and; (4) Designing a framework that seeks to integrate VGI with CGDI datasets that satisfies the criteria and CGDI data accuracy requirements.

First, criteria development achieved through objectives (1) and (2) is discussed. It entails describing criteria selection accomplished by comparing specific CGDI datasets specifications, LBS positioning techniques’ accuracies and VGI characteristics. Then, the process of testing VGI against the criteria and comparing with CGDI data will be outlined. This includes the collection of VGI samples as well as methods and processes used for the testing and comparison. Finally, a framework for integrating VGI with CGDI datasets will be described. It draws upon methods of data integration, updating and validating CGDI datasets, and updating and validating other authoritative datasets with VGI. The different methods are compared and a “best-fit’’ solution for incorporating

VGI into CGDI datasets is derived. The chapter will also highlight the main constraints of carrying out the research objectives and solving the research problem.

Chapter 3.2 Developing Criteria

Criteria were developed to assess the compliance of VGI with CGDI datasets accuracy standards and specifications. The process of selecting criteria was conducted by comparing NRN, NTDB and CanVec data accuracy standards and specifications. (See Chap. 2.3.2.1 for an introduction to these datasets.) In particular, NRN GeoBase datasets specifications were pivotal as the basis for comparison with VGI data in this research. Other factors that influenced the development of criteria included LBS positional accuracies and VGI characteristics. These factors took into account the realistic nature of VGI positional and attribute accuracies and not just what *should* be expected.

Chapter 3.2.1 LBS Positioning Techniques

LBS spatial positioning techniques have a direct impact on the positional accuracy of VGI generated via LBS. For criteria development purposes, knowing the ranges in accuracy offered by different positioning techniques is important. This will facilitate balancing what can actually be achieved and what is expected of the accuracy of VGI created from LBS. From Table 2.3 in Chapter 2, it is observed that the highest accuracy is offered by GPS and A-GPS techniques, whilst the lowest accuracy is achieved by the network based positioning techniques. From Table 2.4 in Chapter 2, positioning requirements for location services are given. Services requiring high accuracy like navigation typically use GPS/A-GPS positioning. Those requiring low accuracy -- such as geographic messaging for social networking -- use network based techniques, e.g. EOTD and COO. VGI data assessed by the criteria include street network data, points of interest and OSM downloads. The street network data and points of interest were collected using GPS, and OSM data is also typically created from GPS. Therefore, the estimated accuracy of the GPS techniques was factored into the criteria. With the estimated GPS accuracy, the performance of the collected VGI data against the NRN, NTDB and CanVec data accuracy specifications could also be estimated.

Chapter 3.2.2 VGI Characteristics

The criteria also took into account important characteristics of VGI, such as: little or no attributes; low accuracy of the geometry; possibly incorrect or erroneous attributes; and the lack of completeness in the data contributed. Attributes automatically stored by LBS devices, e.g. a time stamp, were also considered. This facilitated the comparison of the VGI data and the CGDI datasets by knowing what can and cannot be compared. For example, the VGI street network data collected will not possess the same amount or type of attributes as the NRN datasets. Therefore, it would better to focus on comparing only attributes available in both datasets *or* only those that can be easily validated, e.g. road type.

Chapter 3.2.2.1 Selecting standards and specifications for assessing VGI data quality

NRN, NTDB and CanVec standards and specifications were selected because the datasets cover the entire Canadian landmass. They also contain main topographic map features e.g. roads (from NRN), and other features such as watercourses, urban areas, vegetation, etc. (from NTDB and CanVec). Specifications for data quality were assessed and selected based on whether they apply to all the data products or if they could be applied to VGI.

From the CanVec and NRN Data Product Specifications documents, data quality specifications reviewed included; positional accuracy, thematic accuracy, completeness, logical consistency, temporal accuracy. Positional accuracy and completeness for NRN as well as positional accuracy for CanVec were selected to be included as criteria. From NTDB, data quality specifications selected were: geometric accuracy, and entity and entity occurrence. Positional/geometric accuracy was selected because it is being investigated and occurs in all the data product specification documents. The completeness specification for NRN and the entity and entity occurrence specification for NTBD were selected because they describe the structure of the data. The description includes such important characteristics of data structure as geometric representation, attributive description and metadata. Thematic accuracy specifications were used to derive attribute accuracy criteria. Temporal accuracy was not included because it was not applicable for all the data products; therefore there would be no basis for a comparison. Logical consistency specifications were also not included because they only described how it should be validated.

Chapter 3.2.3 VGI Data Quality Criteria

Criteria derived from NTDB, NRN and CanVec standards and specifications are listed below. They were developed by modifying the specifications mentioned above, in order to assess VGI and compare with the CGDI datasets.

1. **Data Structure**: The data should have entities which have a name to distinguish it from other entities or features, such as a road or bus stop, and the entities should comprise of geometric representation, descriptive representation, an identifier and metadata. (NRN and NTDB specifications). This criterion will be used to judge the completeness of VGI. It has been simplified from a complete scheme of a generic entity (NTDB Entity Occurrence Specification) so as to accommodate VGI. Figure 3.1 gives a diagram of the data structure required.

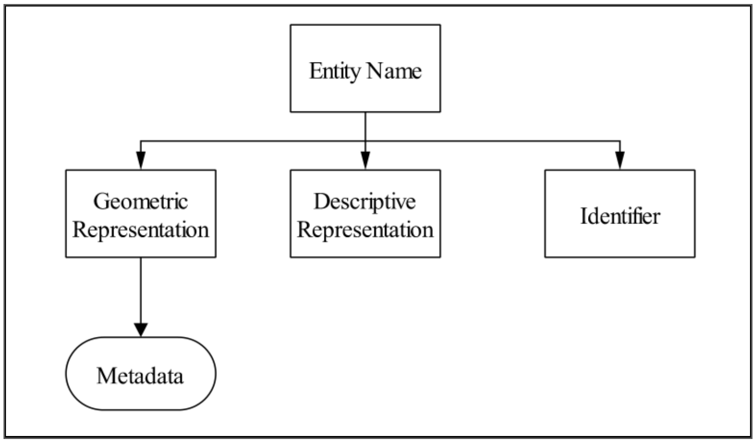


Figure 3.1: Data Structure Criterion (Standards and Specifications of NTDB, 1996)

2. **Metadata**: The metadata for a dataset or feature should at the minimum include: Source of data (e.g. GPS) and Accuracy of Data during acquisition. VGI doesn't often contain metadata since volunteers may not be interested in providing it; which is one of its big disadvantages. However, that does not mean that metadata should be excluded from the overall criteria. If VGI is to be used by official organization metadata will be very important. This criterion gives a simplistic view of what should be expected in terms of metadata. It will also be used to check existing VGI samples for metadata and to compare with the NRN data.

3. **Positional Accuracy**:

   a. As different positioning techniques and processes are used to create VGI, it is essential that they meet a planimetric accuracy criterion to facilitate their integration and utilization. According to the respective data specification documents, VGI must have a planimetric accuracy of no less than ±10m for roads (in clear unobscured areas), and ± 30m or less for other features (in clear unobscured areas) in order to be integrated and used with NTDB, CanVec and NRN datasets.

4. **Attribute Accuracy**: Attributes should include a name or identifier, e.g. road name for a road; and descriptive information, such as type or function. In the case of a road it would describe the type of road i.e. residential street or highway. Requesting too many attributes from volunteers may discourage and deter them from contributing information. However, not having any or too few attributes may make the data difficult to use. This criterion addresses the nature of VGI attributes which is that they are often limited. It also does not include specifications for attributes by NRN, NTDB and CanVec because they are complex and too many. The criterion will be used to assess the accuracy of VGI attributes by comparing them with corresponding CGDI datasets attributes.

5. **Uncertainty**: The maximum acceptable proportion of errors of classification for attributes should be no more than 5% and the maximum acceptable proportion of positional errors should be no more than 10%. The criterion examines the errors that exist within the VGI datasets as a whole. The clause for attribute uncertainty is based on NRN and CanVec specifications; whilst the clause for positional uncertainty is based on traditional mapping specifications which state that 90% or more of all well defined features in a dataset should be within a specified distance (in this case 10m) of their true position. Data with more than 10% of errors will be considered unsuitable.

Chapter 3.3 Data Collection

The next step after developing criteria was collecting VGI data to be evaluated and the datasets to compare it against. VGI samples were collected from the field and from the Web. VGI from the Web came from OSM, whilst VGI from the field came from measurements of streets and points of interest using hand-held GPS devices and a LBS device. The field data was considered as VGI because the instruments and methods used to capture it were similar to those that ordinary individuals might use. Once the data was collected it would be processed then analyzed by testing against the criteria which includes comparing with CGDI datasets.

Chapter 3.3.1 Requirements

The types of data collected as VGI (residential streets and points of interest) was influenced by use of LBS devices as well as availability of similar CGDI data to compare with. For making positional accuracy comparisons with VGI samples, NRN data downloaded from the GeoBase portal were used. Aerial imagery and data of higher accuracy were used to compare positional accuracy of POI (Point of Interest) data. The aerial imagery was obtained from the City of Fredericton and the higher accuracy data were collected using a high grade GPS. Whilst the aerial imagery and high accuracy data are not strictly CGDI datasets, their sources can be considered authoritative. The aerial imagery came from a local agency and the higher accuracy data was captured using surveying methods by graduate students and a university staff member. Table 3.1 gives a summary of the datasets collected and used for analysis.

Table 3.1 Summary of Datasets collected

Dataset and Area	Source	Date Collected/Obtained
Points of Interest (Downtown Fredericton)	• Garmin GPSMAP 76CSx • Garmin eTrex • TOPKON RTK GPS	• (February/09) • (June/09) • (August/09)
Residential Streets (Skyline Acres)	• GPS track from Garmin GPSMAP 76CSx • GPS tracks from iPhone • GPS tracks from Garmin eTrex • GeoBase National Road Network • OSM streets	• (February/09) • (February/09) • (April/09) • (February/09) • (June/09)
Orthophoto (Downtown Fredericton)	• City of Fredericton – Rob Lunn Image Characteristics: • Image Resolution: 15cm • Prescribed Horizontal Accuracy: Unknown • Dates of Photography: Unknown	• (April/09)

The requirements for data collection in the field were: devices used to collect desired data; a map of the area showing data to be collected, and a standard procedure for data collection. This would ensure consistency in the way data was collected for all the devices used. Figure 3.2 shows a map of the areas where data was collected.

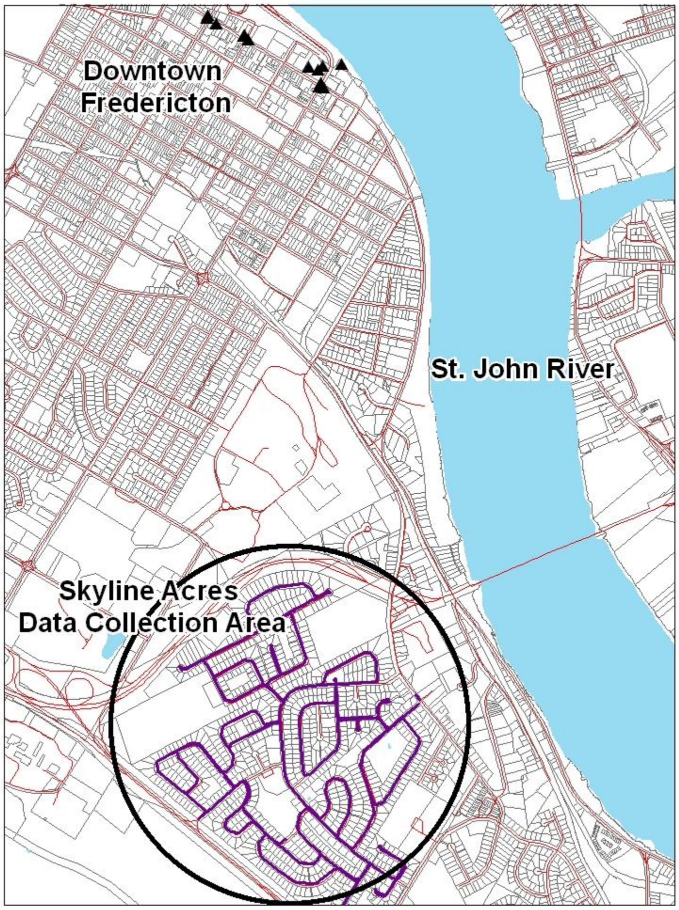


Figure 3.2: Data Collection Area

Chapter 3.3.2 Data Collection Process

Pre-Field

- Familiarity with the devices before going into the field was achieved by revising their manuals and conducting practice data capturing. An experiment design was created detailing how data should be collected and included a map of the area of interest and the features to be collected. Device settings were checked to ensure the date and time were correct and the default coordinate system used was WGS84. The devices memory was also cleared of existing data to increase storage capacity.

Field

- The first field collection took place in February 2009, where streets and POIs were collected. The streets were initially collected as one single track by both: (1) an Apple iPhone using *GPS MotionX* software and, (2) a Garmin 76CSx GPS device. However, this proved to be troublesome as multiple lines were recorded as one street. Editing the streets to yield single lines to represent the streets proved to be very time consuming and difficult when trying to sort out the individual arcs and topological relationships involved. Maintaining the original geometry of the streets during editing was also difficult to achieve, which meant the positional accuracy would be altered. Furthermore, analysis results would be compromised as the edited field data would not give a true depiction of the positional accuracy captured by the devices. Therefore, it was decided to recapture the streets as individual separate GPS tracks using both devices. This way the tracks could be merged later – for easier storage and faster processing- but still maintain their original geometry and their positional accuracies. The streets were collected again in March and April.
- Points of interest were collected using the two different handheld Garmin GPS devices. POI locations were collected in Averaged Position mode which allows for slightly more accurate position as measurements are taken for a period of time

then averaged. The time period for recoding a point location was 60 seconds. The POIs that were collected are listed below:

- ✓ 4 traffic lights (Regent St and Queen St intersection)
- ✓ Monument (William Maxwell Aitken: Lord Beaverbrook)
- ✓ Public toilets
- ✓ Museum
- ✓ Phone booth
- ✓ Fountain (in front of City Hall)
- ✓ Bike Rentals shop
- ✓ Police station flagpole
- ✓ 3 traffic lights (Westmorland Street and Queen St intersection)
- ✓ Lamp post (in Fredericton Officers Square and by the Police Station)

- As features were collected their subsequent attributes were noted. Attributes such as name and map symbol were either manually inputted or default values were given. Linear feature attributes that were generated automatically included: Location – start and finish- (Lat/Long), Elevation, Date and Time, Length, Bearing and Average Speed. For points, automatically-generated attributes included: Location (Lat/Long), Elevation, Date and Time.

Chapter 3.4 Data Processing

The steps undertaken after data were collected and prepared for analysis are described below:

1. Data collected from the field were loaded into GPS mapping software *Expert GPS* - by Topografix- to enable conversion from the GPS exchange file format (.gpx) to ESRI shapefile format (.*shp*). This was done to enable editing in ArcGIS as well as to ensure all datasets were in the same format. The datasets' datum was converted from WGS84 to the North American 1983 CSRS98 (D_North_America_1983_CSRS98) geographic coordinate system. Since taking measurements for comparing positional accuracies would be easier in metric units rather than geographic coordinates, the NAD_1983_CSRS98_ New _Brunswick _Stereographic projected coordinate system was used.
2. Data downloaded from GeoBase was exported as a .*shp* file directly from the GeoBase portal and subsequent editing included projecting to the NAD_1983_CSRS98_ New _Brunswick _Stereographic projected coordinate system and clipping the dataset to the desired extent (Fredericton area).
3. OSM data was extracted in the desired extent (Fredericton area) in OSM XML format (.*osm*). Converting .*osm* to .*shp* required an avenue script – osm2shape.avx (GIS-LAB, 2009) – which was added in ArcView 3.3 and enabled connection of an OSM tools extension. Through the tools the OSM data was imported and converted to .shp. The coordinate system of the OSM dataset was converted to

NAD_1983_CSRS98_ New _Brunswick _Stereographic projected coordinate system following that.

4. Once the datasets preparation for comparing their locations was completed, the actual process of measuring the difference between two datasets locations would begin. Figure 3.3 outlines the workflow for processing data to make it ready for comparisons. It includes the software required and steps undertaken.

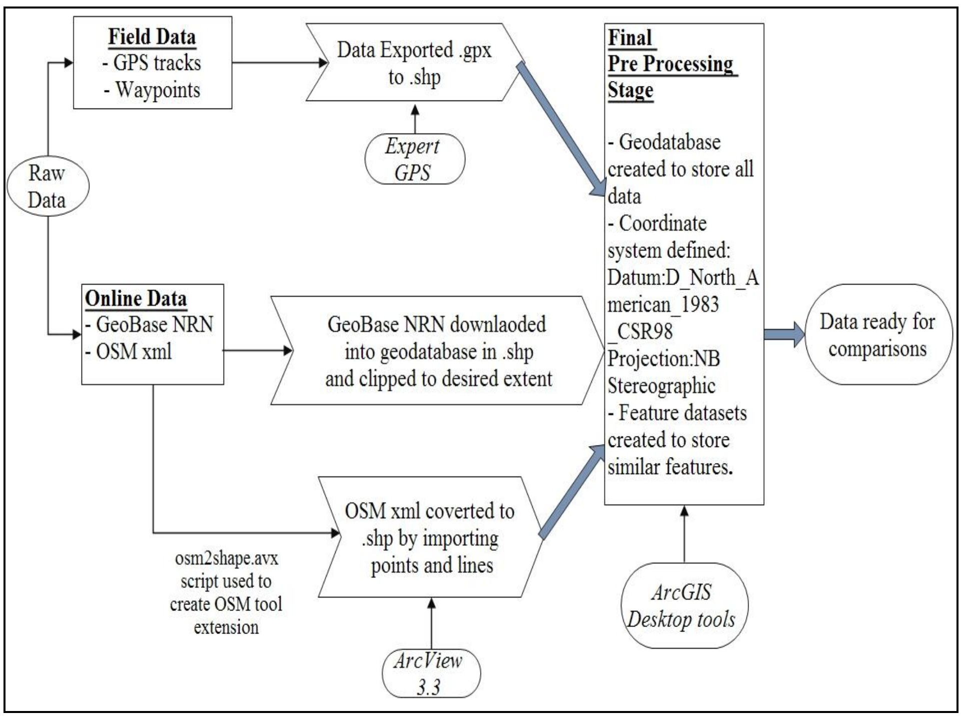


Figure 3.3: Data pre-processing workflow

Chapter 3.5 Evaluating VGI

Evaluating VGI entailed testing it against the criteria and comparing it against CGDI datasets. These functions are required to fulfill Objective (3) and intended to assess the extent that VGI complies with CGDI data accuracy specifications and standards. Testing VGI against the criteria would facilitate measuring the positional and attribute accuracy as well as comparing it with CGDI datasets. The criteria guide how the comparison with CGDI datasets should be made. VGI street network data collected from OSM and the field was compared with the NRN dataset, whilst VGI points of interest were compared with aerial imagery and higher accuracy data of the same features. Methods for assessing attributes involved visual inspections - as there weren't that many. However, methods for assessing a large number of attributes were investigated and included as proposed solutions if there were many attributes to assess.

Goodchild and Hunter (1997) and Hunter's (1999) methodology for evaluating positional accuracy was used for assessing the linear features (VGI streets). Positional accuracy of a feature's digital representation was evaluated by measuring the difference between locations from the dataset being assessed and locations from a dataset determined to have higher accuracy. The VGI streets were the data being assessed and the NRN dataset was used as the dataset with higher accuracy. The comparison was carried out by:

- Sampling points along the tested sources (VGI street data) and in each case measuring the distance to the closest point on the reference source (NRN data) to obtain distributions and statistics such as the mean or percentiles (Hunter and Goodchild, 1997);

- Using buffers to determine the percentage of line from one dataset that is within a certain distance of the same feature in another dataset of higher accuracy (Hunter and Goodchild, 1997; Hunter, 1999; Haklay, 2008). In this case, the percentage of line of VGI street data within 10 m of the same line in NRN data would be determined. The buffer distance of 10 m is from the positional accuracy criterion which stipulates: VGI must have a planimetric accuracy of ± 10 m for roads (in clear unobscured areas).

Chapter 3.5.1 Buffer Comparison Method (Lines)

The buffer comparison method was selected because it was easier and faster to use for analyzing large or multiple datasets. There were four VGI street datasets; (1) streets from OSM, (2) Garmin GPSMAP 76CSx streets, (3) iPhone streets, and (4) Garmin eTrex streets. Therefore, using the buffer method would be easier and quicker than using the manual method. It also depicts the amount of errors in the VGI datasets as a whole; compared to the NRN dataset.

The method entailed creating a layer with a 10 meter buffer around the NRN streets and then intersecting the buffer layer with the VGI streets data. The extent of the NRN dataset was ensured to be the same as the extent of the VGI streets datasets. Additionally it was ensured that the same streets appeared in both datasets. This was done so that the difference in length between the NRN dataset and VGI streets datasets could be attributed to positional errors and not to factors like: (1) either one dataset containing more/less features; or (2) the extent being larger in one dataset than the other. The four VGI street layers were then intersected with a NRN buffer layer using a CLIP overlay operation in ArcMap. This would result in a layer showing only the portions of streets that were contained within the buffer. A diagram showing the overall procedure is given in Figure 3.4. The total length of intersected road track data in this layer was then divided by the total length of the original VGI street before clipping (the VGI and NRN datasets were ensured to have the same extent), and then multiplied by a hundred to give a percentage value. The equation used is shown below:

(Result of Intersection Layer total length ÷ Original Layer total length) * 100

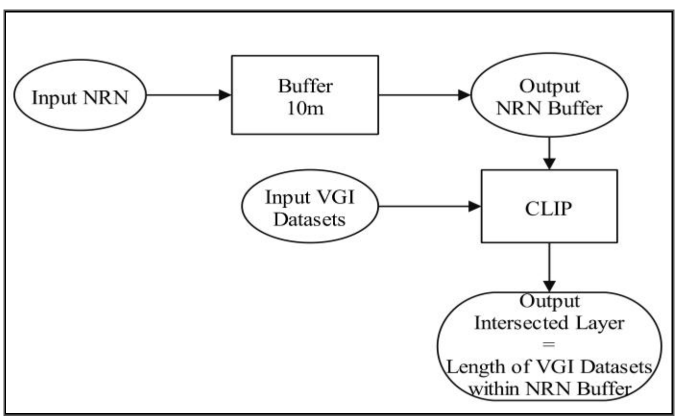


Figure 3.4: Buffer Comparison Method workflow

Chapter 3.5.2 Manual Comparison Method (Lines)

Manually comparing the VGI datasets and the NRN dataset was conducted to obtain the average distance between corresponding street features within the datasets. This process was very long, therefore only three VGI datasets were measured; (1) streets from OSM, (2) iPhone streets, and (3) Garmin eTrex streets. The Garmin GPSMAP 76CSx streets were excluded because the original dataset had double lines representing the streets. It would be difficult and even more time consuming to measure distances between street features represented as double lines in one dataset and as single line in another. The edited version of the Garmin GPSMAP 76CSx streets dataset was not measured either because it would be difficult to discern human errors (introduced by editing) from positional (recorded by the device) and other errors. The process of sampling points and measuring the distances is described below:

1. Points were sampled along the streets in the VGI datasets at 50m intervals by using the Divide tool from ArcMap Editor Tools. It automatically divides linear features and places reference points at the intervals specified by users. This was useful as the three VGI datasets had an average of twenty streets each. However, only OSM and iPhone streets were divided this way. This is because the eTrex GPS streets contained an attribute that prompted an error message when the

Divide Tool was run. For this dataset, reference points were manually placed at 50m intervals on lines by measuring the distance using the Measure Distance tool. The main issue with this process was that human errors could be easily introduced through not accurately measuring the distance between reference points. To mitigate this problem, it was ensured that during the digitizing the zoom level was high, the Measure Distance Tool's cursor followed the same path as the line being measured, and after adding reference points the distances between them were checked to be equal to 50m.

2. The distance between the reference points on lines in the VGI street datasets to the closest point on the same lines in the NRN dataset was measured by:

    - Creating new and empty line feature shapefiles to be digitized (for each street being measured) in the geodatabase. This would make sure that when features are digitized, i.e. the distance measured; the length would be automatically recorded. This is because lengths of linear features are automatically recorded in the geodatabase when the feature is created.
    - For each street, a perpendicular line was digitized from the reference points to points on the same street in the NRN dataset. The ArcMap, Editor Toolbar, Create New Feature Tool was used to digitize the perpendicular distance between streets in the two datasets. Measuring the perpendicular distance was done to make certain that measurements were taken accurately and to maintain consistency of the measurements. Figure 3.5 shows how the perpendicular distance between points was measured.

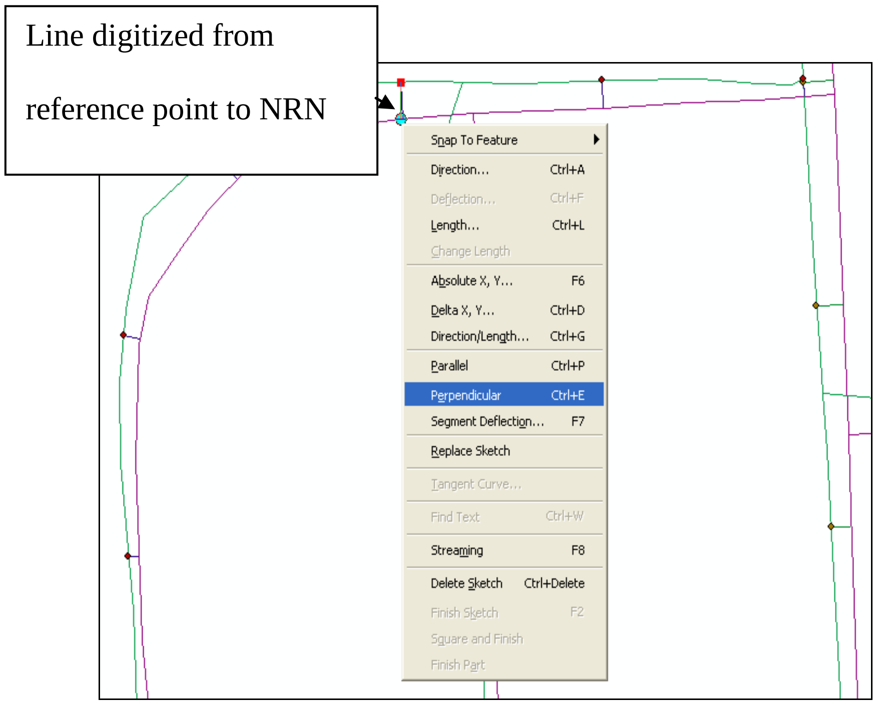


Figure 3.5: Measuring Distance manually

Chapter 3.5.3 Manual Comparison Method (Points)

The POI data was evaluated using a method similar to the manual comparison for the VGI streets. It was decided to collect the same POI data with a Real Time Kinematic (RTK) GPS to obtain a dataset of higher accuracy. These points would then be compared with the VGI POI data. The aerial imagery was not used because its accuracy was unknown. Without this knowledge, the imagery could not be used as a reference source to determine how accurate the VGI points were. The imagery could have been georeferenced to determine its accuracy, but a lack of ground control points within the imagery coverage area prohibited this task. Collecting ground control points for geo-referencing the image would not be feasible due to time limitations and because collecting them with RTK GPS system, available at the time, required known ground control to be observed as well.

As there were multiple sets of POI datasets collected using the hand-held GPS devices, the average position for a point was calculated before distance could be measured. The distance between points from the VGI POI and RTK GPS POI datasets was measured using the same procedure as Step (2) for the manual line comparison method. The average distance was also calculated. Errors were calculated by obtaining the difference in northings and eastings.

Chapter 3.5.4 VGI attributes

Assessing the accuracy of attributes for VGI data was included in the objectives of this research to see whether or not they could be considered useful. Criteria for assessing VGI attributes included checking for a name or identifier, such as road name, and descriptive information, such as the type or function. Visual inspection of the VGI datasets' attribute tables was conducted to test them against the criteria. Comparing them with the NRN datasets could not be achieved since the NRN data had codes and their values for attributes. Default attributes, automatically added by the hand-held GPS devices, were kept for data collected in the field and could be verified by the data collector; however, there were no similar datasets to compare them with to ensure their values were accurate.

For measuring attribute accuracy for large datasets, comparisons with datasets from well known sources are usually done. An example is using satellite imagery to check for the correct classification of a feature. For validating other attributes such as the features name and "building type", a more labour-intensive process of checking the correctness of the names against a reference source would need undertaking. However, if VGI was to become another source of NRN update information, a more semi-automatic process should be used to test the accuracy of VGI attributes else the entire process would be inundated by the amount of data to test.

Chapter 3.6 Framework for Data integration

The final objective of this research is to design a framework for integrating VGI with CGDI datasets that satisfies the criteria and CGDI data accuracy requirements. Designing this framework will take into consideration different approaches for updating, validating and integrating spatial data by different organizations. First, methods for updating and validating data by an organization within the CGDI were overviewed. Then, methods of updating, validating and integrating VGI with authoritative datasets were illustrated through two existing examples. These consist of the Tele Atlas Map Insight™ initiative and the State of Victoria, Australia Department of Sustainability & Environment (DSE) Notification & Edit Service (NES) system. The methods were compared and an integration framework was derived from them.

Chapter 3.6.1 CTI Update Method

Natural Resources Canada's Centre for Topographic Information (CTI) is the country's national topographic mapping agency. It is responsible for the acquisition, management and dissemination of topographic information for the Canadian landmass (Natural Resources Canada, 2009). CTI initiated a national program to update topographic data at the 1:50 000 scale using Landsat – 7 orthoimages (Martin and Loubier, 2002). The task was very daunting considering the extent of the Canadian landmass; therefore, a suitable and cost effective method for updating was selected (*ibid*).

1. Updating

A change detection method was used because it made it possible to validate the existence, the geometry, and the description of features visible within the image, whether they were modified or not (Martin and Loubier, 2002). However the change detection process of topographic data was too complicated to be totally automated. The nature of the entities to be mapped were difficult to discriminate by automated methods, thus visual inspections had to be done to update the data adequately (*ibid*). Updating through change detection required studying the evolution of a given entity (*ibid*). CTI used a change detection decision tree to study the evolution of every entity. The main questions set by the decision tree were concerned with visibility, entity identification, importance of the change, minimal dimensions, generalization and identification of the change detected. Figure 3.6 summarizes the change detection decision tree used in the updating process.

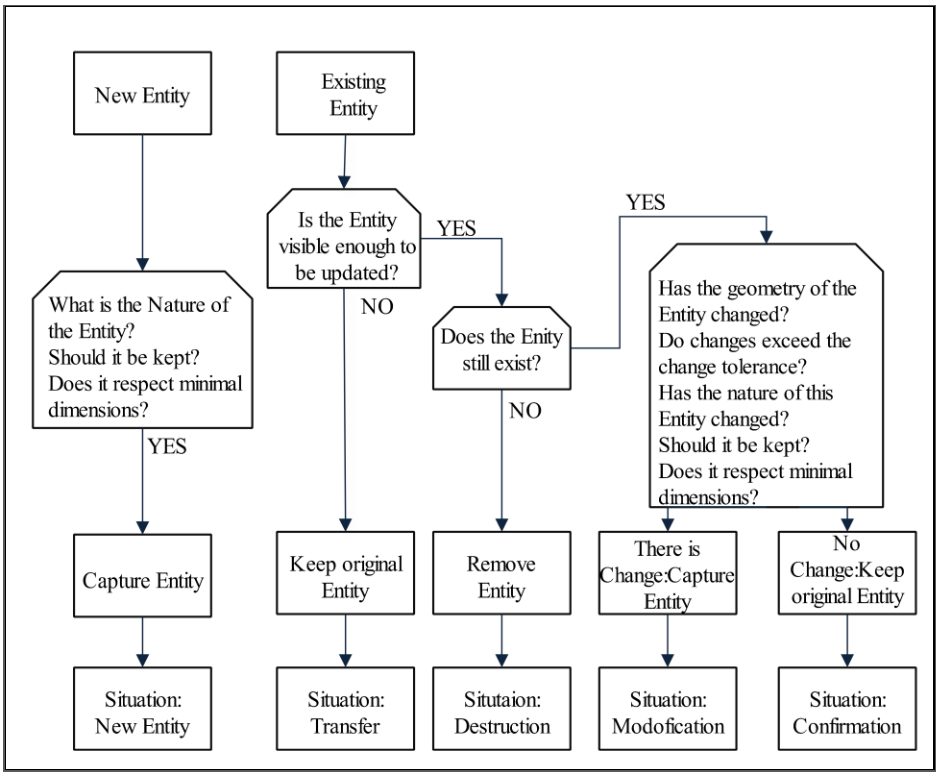


Figure 3.6: CTI Change Detection Decision Tree (Martine and Loubier, 2002)

2. Validating

Updating work was given to private companies; however, CTI was responsible for validating changes and inspecting results from the updating. The same decision tree was used for quality control. This meant that the inspection process was simpler and faster, but, it would have been subject to same errors as during the update process. Constraints of the update and validation process using change detection were: complexity of the decision tree, definition of entities, image heterogeneity (i.e. images taken in different seasons were used to detect changes), size of the territory and production rate (Martin and Loubier, 2002). The objective of CTI after this initiative was to replace visual interpretation to the degree possible by automated methods, depending on the constraints mentioned above, which would reduce or eliminate the use of human interpretation and enhance the quality and homogeneity of the results (*ibid*).

Chapter 3.6.2 Tele Atlas Map Insight™ Update, Validation and Integration Method

The Map Insight program by Tele Atlas® is a consumer feedback and change reporting system (Tele Atlas Map Insight™, 2009). Map Insight is a Web-based application that streamlines the collection of feedback by providing end users with a tool to transmit map changes and updates directly to Tele Atlas (Tele Atlas Map Insight™, 2009). Map Insight allows partners like TomTom to offload collection and reporting of user feedback directly to Tele Atlas (*ibid*). In conjunction with consumer feedback and change reporting, Tele Atlas® uses commercial and proprietary aerial imagery to update and improve their maps. Furthermore, enhancement and validation of data is conducted through extensive data editing processes and actual in-field data collection (*ibid*). A workflow diagram showing the validation process is shown in Figure 3.7.

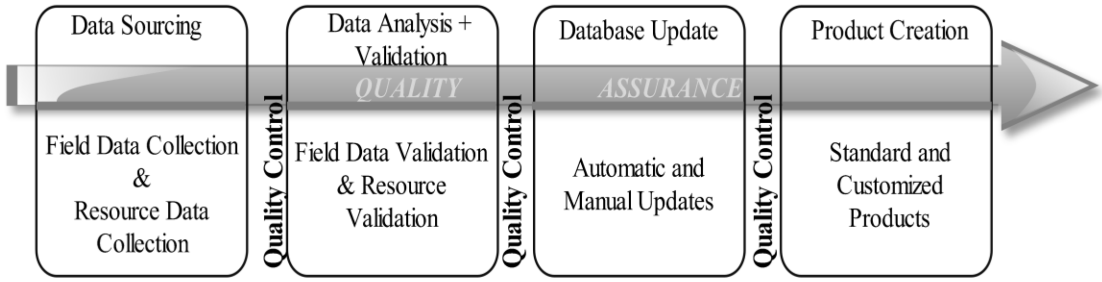


Figure 3.7: Tele Atlas Validation Workflow process (Henry and Temmink, 2008)

1. Updating and Validating

The four main processes are 1) Data sourcing; (2) Data analysis and validation; (3) Database update and; (4) Product Creation. During data sourcing, information is acquired from data collection vans, governments and partners, web crawler tools, cars, satellite imaging and consumer feedback (Tele Atlas Map Insight™, 2009). Once data is obtained data mining and integration tools are used, to rapidly review, validate, and implement database changes (*ibid*). For example, Tele Atlas® Multinet® data (street network vector database which incorporates user feedback) is updated and validated using data management and certification software by 1Spatial (*ibid*). Multinet® data is updated and published every three months with new attributes, features, and content improvements (*ibid).*

1Spatial's data management and certification software includes Radius Studio™. This product is a spatial processing, analysis and compliance engine. It provides a data certification platform that ensures the quality and consistency of spatial data (1Spatial Radius Studio Concept Guide, 2008). It is an implementation of a rules-based processing environment (*ibid*). Rules-based processing provides a means of measuring and enhancing data quality (Woodsford, 2007). It follows the FACT-PATTERN-ACTION dynamic. That is: Given some facts, if they meet any of the patterns/rules, perform the defined action(s) (*ibid*). FACTs are known as the data source. PATTERNs are the rules that the data source obeys or should obey. ACTIONs happen as a result of PATTERNs being applied to FACTs (*ibid*). Figure 3.8 illustrates the Fact-Pattern-Action dynamic using a VGI example.

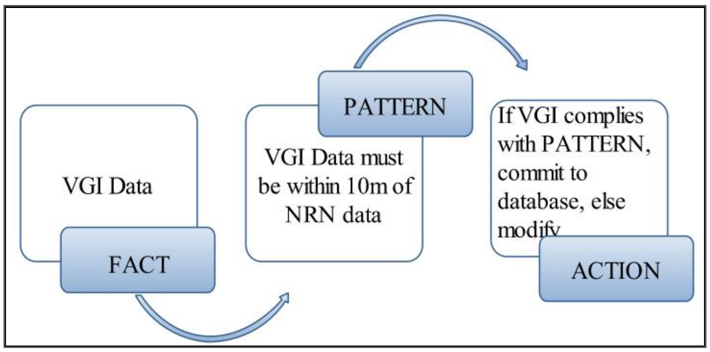


Figure 3.8: Fact-Pattern-Action Dynamic

Using a rules-based processing environment such as that provided by Radius Studio™ could provide a quantitative measure of data quality or conformance (Woodsford, 2007). Specific tasks performed by Radius Studio ™ include: Open data, Discover rules, Check rules, Apply action, Apply action map, and Commit data (to database) (1Spatial Radius Studio Concepts Guide, 2008)

Chapter 3.6.3 NES Update, Validation and Integration Method

NES is a service provided by DSE to improve processes for notifying and maintaining changes to the Australian state of Victoria's authoritative spatial datasets (Coleman et al, 2009). NES is available to state and local government organizations that already participate in data sharing and data maintenance programs with DSE (NES, 2009). Users can report an error in a dataset or change that is required to any dataset accessible by the NES system and reporting is achieved through raising a notification (NES User Manual, 2008), Users of the NES portal are required to be registered before they can log onto the system. Once a user has logged onto the portal, the service is able to detect the user's role (notifier, custodian, maintainer and system administrator). The service also detects any groups and organizations to which users belong and the associated functions which are permitted (*ibid*). Certain users, such as the general public are only ever be able to generate a notification from the Easy Editor, which allows change requests to be created and viewed. Other users, who have a Custodial Role, are be able to effect changes on datasets for which they are custodian, directly within the system using the Advanced Editor (*ibid*). In the Advanced Editor data attributes and geometry can be edited; new change requests can be created and existing change requests viewed; and the custodian for the dataset can then verify and approve any changes made. Edited data can be extracted by a data maintainer to be committed to the authoritative databases (ibid).

1. Validation

A change request/notification must be edited and approved in the Advanced Editor, before it can be sent to the data maintainer (NES User Manual, 2008). To be approved in the Advanced Editor, the change request must pass the validation rules that apply to the data set applicable for the Type of Change (*ibid*). Validation will check that topology and attribute information contained within the change request conforms to the required data standards for the given dataset (*ibid*). Validation ensures data integrity in the authoritative datasets is retained and ensures data maintainers receive change requests of the required quality (*ibid*). Rules-based processing validation based on 1Spatial's Radius Studio ™, enables data validation prior to commitment to NES Corporate Spatial data Library or Maintainer (NES User Manual, 2008).

Chapter 3.6.4 Framework for integration

The main reason for selecting the methods previously described is because they exemplify current practices of updating data and how VGI is utilized by official organization in industry and government. CTI updated and validated their topographic data using satellite imagery; Tele Atlas uses aerial imagery, field data and VGI (in the form of consumer feedback and change reports) for updating and enhancing their data; and DSE uses VGI (in the form of change reports) in updating the NES database. Different methods were used to validate the VGI and data used. The examples illustrate that the method of validation selected depends on the types of data being used for updating. The common feature of all the validation methods is that they employed rules-based processing during the validation process. The change detection decision tree used rules for determining features' suitability and Radius Studio™ uses the Fact-Pattern-Action dynamic. The framework for integrating VGI with CGDI datasets is based on the methods for updating and validating data described in the previous section.

The main features of the framework are:

1. A wiki style web application that allows users to add and edit VGI.

2. A validation engine which serves as the backend of the web application mentioned in (1). It will be a temporary database for validating VGI before it can be integrated with the CGDI datasets.
3. An integration phase performed after the validation using conflation techniques. Data will be integrated in the temporary database first, and then it will be committed to the database containing original datasets as a finished product. Figure 3.9 gives an outline of the integration framework.

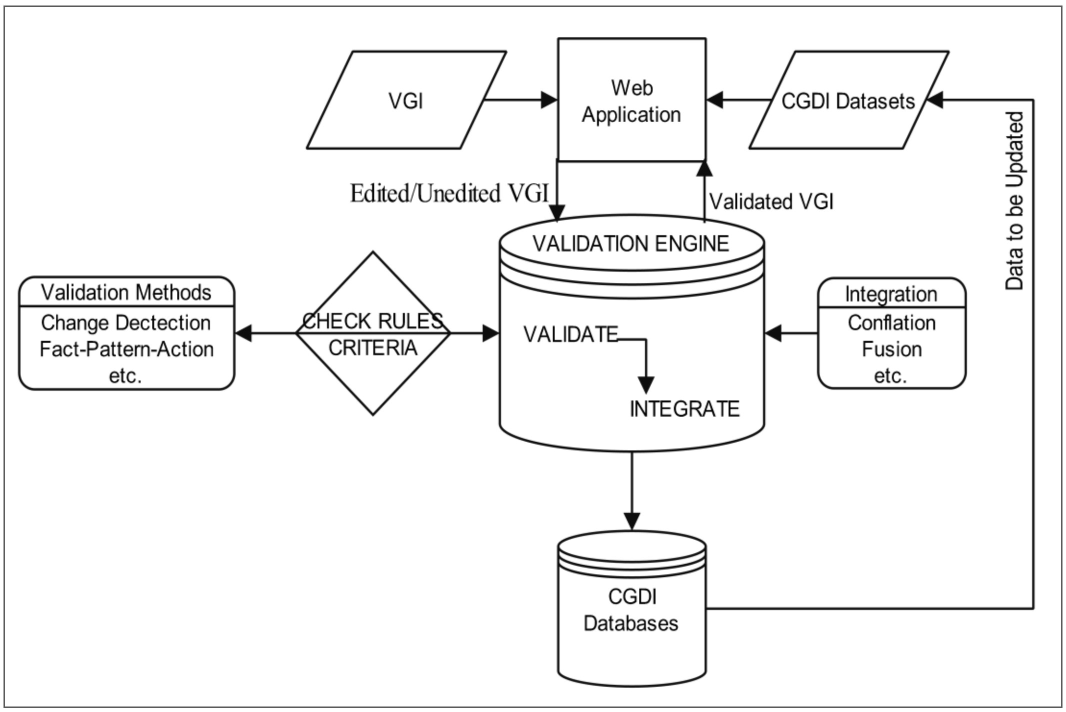


Figure 3.9: Integration Framework

Chapter 3.7 Lessons Learned

The main issues regarding accomplishing objectives were related to developing criteria, collecting data, measuring the positional accuracy of VGI and designing the framework for integration. Finding criteria that would assess the suitability of VGI for integration with CGDI datasets was challenging because the two sets of data are very different. It was important to use standards and specifications for CGDI datasets as a basis for criteria, so that VGI’s suitability could be measured and compared. However, finding standards and specifications that accommodated VGI and maintained the CGDI data specifications' exact wording and meaning was difficult. Therefore, modifications were often made to partly maintain the specifications and accommodate VGI.

Data collection constraints included not being able to use the aerial imagery for measuring the accuracy of the POI data. The imagery could not be used because the prescribed horizontal accuracy was unknown. This led to using RTK GPS to obtain the same POI and using that data as the source of higher accuracy data during comparison. One of the challenges faced was in inadvertently using old control monuments to set up the base station: where the accuracy may have deteriorated due to changes and disturbances of their surroundings. Obstruction of communication between the base station and rover, or obstruction of the GPS signals by tall buildings downtown was also a constraint. Avoiding built up areas and setting up new control by using high accuracy static GPS methods would have mitigated these issues. However, due to time limitations, sampling data in a new area and making static GPS observations could not be achieved.

For data processing and evaluating the data, the main drawback was human errors introduced during editing and manually measuring equal distances to place reference points in VGI streets datasets. Unfortunately this could not be avoided; however, rechecking the edits and measurements was done to minimize the human errors.

The challenge of designing the integration framework was that it was derived from similar examples. The Map Insight and NES are similar in that users' feedback and change reports are what can be considered as VGI for those systems. They also both use the same software, Radius Studio TM, as a validation mechanism. Other examples of how VGI is utilized in government and industry may have yielded different methods for updating, validating and integrating VGI. However, utilization of VGI by governments and industry is not yet popular, so only a few examples exist and they mostly use the same methods.

The methodology adopted in order to meet research objectives and solve the research problem was discussed in this chapter. Criteria development, VGI data capture and processing, and methods used for testing VGI against the criteria and comparing with CGDI datasets were discussed. Finally, the process of designing the framework for integrating VGI with CGDI datasets was described. It entailed assessing data integration methods as well existing methods for updating and validating authoritative datasets using VGI. The next chapter presents and analyzes the results from evaluating VGI. The performance of VGI throughout the evaluation processes is examined, and a critique of the integration framework is also given.

Chapter 4 Results of Investigating VGI Geometric properties

Chapter 4.1 Introduction

The evaluation of VGI datasets and the development of the framework for integration, conducted in the previous chapter, are crucial objectives which facilitate in ultimately solving the research problem. The evaluation consisted of testing the VGI datasets (Street and POI data) against previously developed criteria and comparing them with similar CGDI datasets. Positional accuracy, attribute accuracy, data structure, metadata and uncertainty of VGI data were measured via the criteria. This entailed making comparisons with CGDI datasets and CGDI data accuracy specifications. Where comparison with a CGDI dataset could not be performed, i.e., the POI data case, data of higher accuracy was used instead, but CGDI data accuracy specifications pertaining to the criteria were still used. A framework for integrating VGI that satisfies criteria hence satisfying CGDI data accuracy requirements was designed by evaluating different methods for updating authoritative datasets with VGI. It was judged on the basis of how well it satisfied the last objective of the research: to develop a framework for integrating VGI that satisfied the criteria with CGDI datasets.

This chapter examines the performance of VGI throughout the evaluation processes. Results of the evaluation will be presented and subsequently analyzed. Methods used for testing criteria and comparing with CGDI data will be analyzed along with the results. This is because they may have had an impact on the results. Errors induced by the experiment as well as unanticipated errors shall also be discussed. A critical analysis of the integration framework shall be conducted, with emphasis on measures for improvement. The goal of the research described in this chapter is to determine, based on these criteria, the suitability of VGI for use in CGDI databases by examining results from the evaluation.

Chapter 4.2 Testing Positional Accuracy criterion: Buffer Comparison Summary

Results for investigating VGI geometric properties were obtained from testing the positional accuracy criterion (from criteria developed in chapter 3). The criterion stipulates that VGI must have a planimetric accuracy of no less than ±10m for roads in clear unobscured areas, and ± 30m or less for other features (again, in clear unobscured areas). VGI street centerlines (tested data) were compared with the GeoBase NRN (reference data) and three sets of results were obtained: (1) Buffer comparison (streets) results; (2) Manual comparison (streets) results; and (3) Manual comparison (POI) results. The table below shows results from carrying out the Buffer Comparison method described in Section 3.5.1. This method was used to test the positional accuracy criterion.

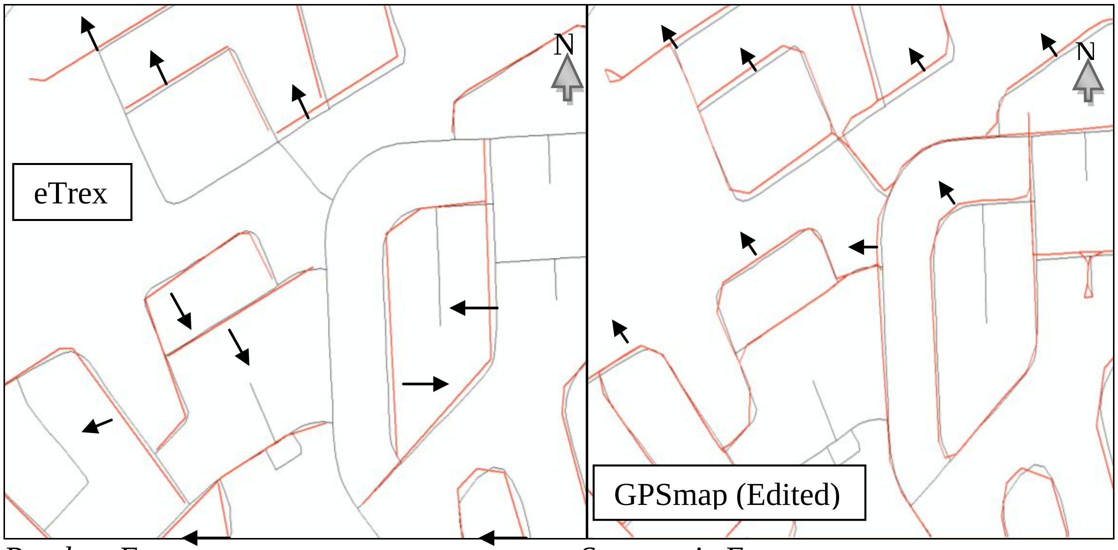
Table 4.1 Percentage of VGI streets within 10m of NRN Streets

VGI Street Centerline Data Source	**Percentage Within 10m**
iPhone	82.86%
OpenStreetMap	94.04%
Garmin eTrex	90.47%
Garmin GPSMAP 76CSx (Edited)	89.81%
Garmin GPSMAP 76CSx (Unedited)	90.37%

The OpenStreetmap (OSM) downloads had the highest percentage within the buffer zone, followed by the eTrex, GPSMAP and iPhone streets respectively. The high percentage of OSM data within the buffer zone means that most of the data was within the 10m threshold and was more accurate than the rest of the VGI street datasets. This could be a result of data sources from which the data is derived. OSM sources could include digitizing Yahoo! Aerial imagery (Yahoo! have agreed to let OSM use their imagery for tracing purposes) and other government or commercial satellite imagery. They could also include using higher accuracy GPS devices for recording road-centerline data. Also, the wiki-style process of editing and reviewing by peers in OSM network could contribute to higher accuracy in the data as errors are usually detected and removed. However, there is no metadata or tags indicating which sources the data came from or that it was edited by others. Therefore, there is no way to tell what actually contributed to the high accuracy of these particular data.

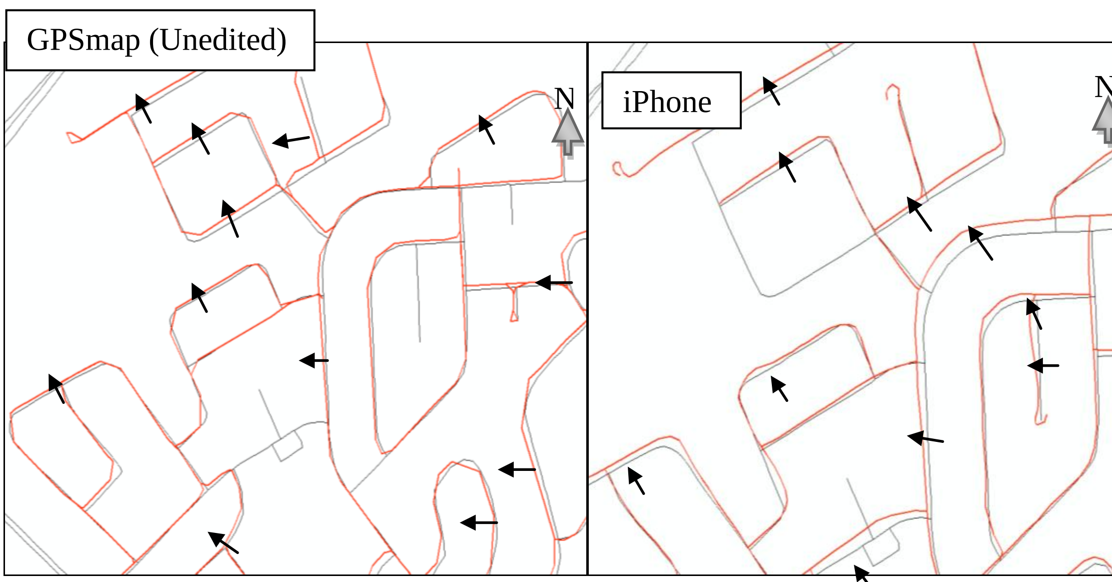
Street centerline data collected using the Apple iPhone had the lowest percentage of streets within the buffer zone. This means that more iPhone streets fell outside the 10m threshold and were less accurate than the rest of the VGI street datasets. This could be attributed to low accuracy of the GPS receiver in the mobile phone or the positioning technique employed by the device. The iPhone is not a dedicated GPS device and third party applications are used (in this case GPS MotionX) to record GPS tracks. Assisted-GPS (A-GPS) is the positioning technique used by the iPhone and it finds the closest satellites to identify locations. However, if there is no clear line of sight from GPS satellites, locations are determined via Wi-Fi; and if Wi-Fi is out of range, locations are determined using cellular towers (Zandbergen, 2009). The accuracy of positioning methods deteriorates when GPS cannot be used (*ibid*). This might explain the low accuracy of the iPhone data because during data capture the device was inside a vehicle; where obstruction to satellites may have occurred due to the roof, and resulted in poor quality positioning or lower positioning methods being used.

The values of percentage of streets within the buffer for the eTrex, GPSMAP (edited) and GPSMAP (unedited) streets were 90.47%, 89.01% and 90.37% respectively. These values are close and suggest that the GPS devices yielded roughly the same results in terms of positional accuracy. However, upon inspection of the datasets, it appeared that errors in the eTrex data were random while errors in the GPSMAP (edited) and GPSMAP (unedited) appeared to be systematic. The ideal situation would have been that the VGI streets were superimposed on the NRN streets, however, due to previously mentioned errors, a trend of displaced VGI streets is observed. Figure 4.1 and 4.2 show the direction of displacement of the VGI street datasets compared to NRN dataset. The VGI streets are represented as red solid lines, NRN streets are black solid lines and the arrows indicate the direction of displacement.



Random Errors *Systematic Errors*

Figure 4.1: Direction of displacement



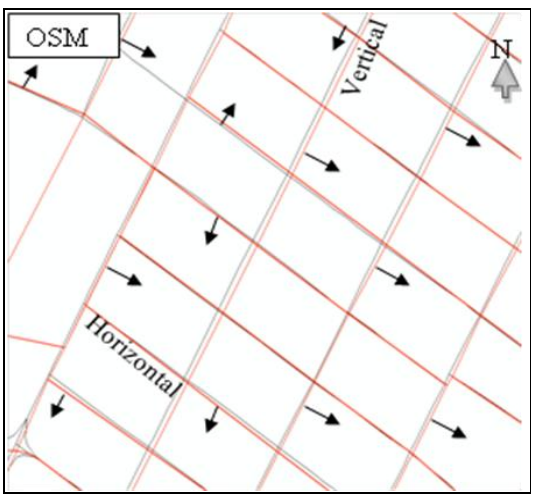
Systematic Errors *Systematic Errors*

Figure 4.2: Direction of displacement

Random errors are caused by inherently unpredictable fluctuations in the readings of a measurement apparatus or in the experimenter's interpretation of the instrumental reading (Zhang and Goodchild, 2002). They also may be in part due to environmental interference or unexplained phenomenon. The displacement errors in the eTrex dataset appear random because it cannot be predicted which direction the displacement will occur. Systematic errors are predictable and could be caused by imperfect calibration of the measuring device, imperfect methods of observation and environmental factors interfering with the measurement process (*ibid*). If the cause of systematic errors can be identified and modelled, they can be removed (*ibid*).

Systematic errors in the GPSMAP (edited), GPSMAP (unedited) and iPhone datasets exhibit a northwesterly trend which can be explained by the manner in which the observations were taken. First, the magnitude of the errors might be influenced by the reduction in accuracy of GPS positioning caused by a limited clear view of the sky during data capture, which was a result from obstruction by the roof of the vehicle. Second, the direction of the displacement of the streets could be a result from the direction traveled by the vehicle. During data capture, the vehicle traveled in different directions and was on the right-hand side of the road. This explains the trend of the VGI streets being displaced towards one side of the NRN streets. It would be difficult to model these errors since the magnitude of the displacement varied along each street throughout the dataset.

The displacement of the OSM streets indicate that the errors could be systematic since the majority of the streets (shown in Figure 4.3) are displaced towards the south east direction and south west direction. However, since the source of data is not known, nor is the method in which streets were recorded; the cause of the errors cannot be explained. Figure 4.3 shows the direction of displacement for the OSM streets. The systematic error might also be due to the fact that the OSM dataset's datum was converted from WGS84 to the North American Datum 1983 CSRS98 (NAD83_CSRS98) geographic coordinate system. The difference between the reference ellipsoids of these coordinate systems is approximately between 1m and 1.5m (NRCAN: Canadian Spatial Reference System, 2009). Using geographic transformations, e.g. 7 parameter shift, produces a datum shift from WGS84 to NAD83, however, transformations may be error prone as well (*ibid*). The observed systematic error in the OSM, iPhone and GPSMAP datasets could be a result of error from the transformation, used by *ArcMap* GIS software - to covert from WGS84 to NAD83.



Systematic/Random Errors

Figure 4.3: Direction of displacement

Random errors would also exist for all the datasets because the devices and methods of observing were not perfect. During the evaluation, data was edited to ensure that the VGI street datasets had the same extent as the NRN streets. This process introduced additional errors and was the main contributing factor to the shortcomings of the buffer comparison method. During buffer comparison, the total length VGI streets within the 10m buffer is compared to the total length of streets in the NRN dataset. It is assumed that NRN data is more accurate, therefore the VGI data should have a greater value for the total length since it more prone to positional errors. However, to make the comparison fair and accurate, both datasets must have the same extent and it must be ensured that data being compared covered the same streets and had the same number of streets in both datasets.

While it is useful to know the percentage of VGI streets within the required 10 m buffer, it would also be useful to know the value of the distance between those streets the corresponding NRN streets. The limitations of the buffer comparison methods include not being able to obtain a value for the distance between the VGI and NRN streets. Therefore, the manual comparison method was used to overcome the limitation of the buffer method. Using both methods to evaluate positional accuracy would improve the analysis and provide better understanding of the results.

Chapter 4.2.1 Testing Positional Accuracy criterion: Manual Comparison Summary for Streets

Mean separation was measured between individual streets in both NRN (reference data) and VGI streets (tested data). iPhone, OSM downloaded and Garmin eTrex GPS streets were the three VGI datasets compared via this method. Results are shown in Figures 4.4, 4.6 and 4.7.

a. **iPhone**

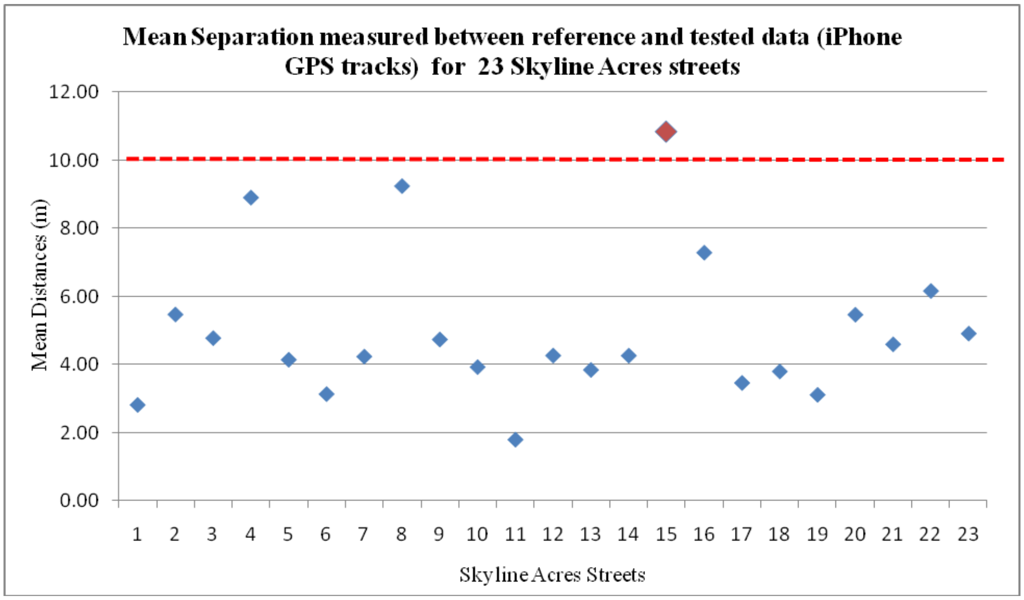


Figure 4.4: Graph showing Mean Separation measured between VGI (iPhone) and NRN streets

As depicted in Figure 4.4, there was only one mean separation value which exceeded the 10m limit stipulated by the criterion. This means that, out of the twenty-three streets measured, only one street did not comply with the criterion. The street which did not comply with the criterion, shown as the red point, was Street 15 - Liverpool Street – and the mean separation measured between the VGI (iPhone) data and the position of corresponding points in the NRN file was 10.83m. Practically speaking, eighty-three centimetres can be considered neglible considering the ultimate goal of integrating VGI datasets like this with CGDI datasets like the NRN. This is beacuse the spatial resolution denominator of the NRN dataset is estimated to be approximatelty 10000 (NRN Data Product Specification, 2007). Therefore, an error of 83cm on the ground, which translates to 0.083mm on the map, at that scale, would not significantly compromise the planimetric accuracy of the dataset. However, given that the buffer method returned a value of 82.86% for iPhone streets which are actually within 10m of NRN streets, it means that there were distances measured, for some iPhone streets, which exceeded 10m. Averaging the distances measured between VGI street and NRN datasets does not give a conclusive representation of whether or not the criterion was fullfilled. Figure 4.5 illustrates the difference between the buffer and manual comparison methods. The diagram on the left, in Figure 4.5, shows how much of of the street is within 10m (represented by the blue buffer), whilst the diagram on the right shows distances measured between the datasets.

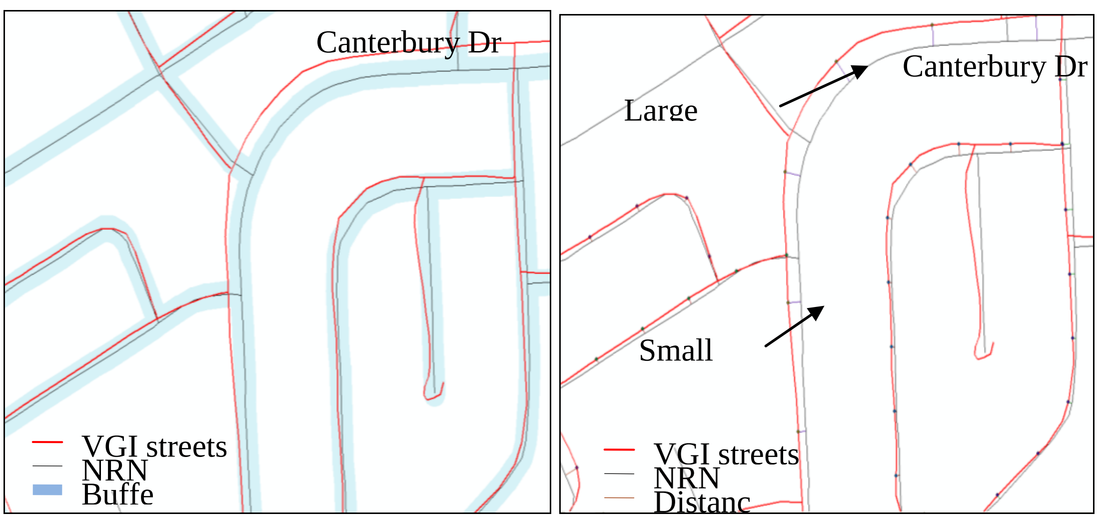


Figure 4.5: Comparison of buffer and manual method

The separation between a VGI street and its NRN counterpart can vary across the entire street. For example the diagram on the right, in Figure 4.5, shows that the meseaured separation between the VGI (red) and NRN (black) streets are larger in some parts of the street than in others. The largest separation distance measured between VGI-

Canterbury Dr and NRN-Canterbury Dr was 19.918m and the smallest separation distance measured was 0.903m. Taking measurements at shorter intervals along the streets might yield more accurate results. The results for the OSM and eTrex datasets which were measured at the same interval as the iPhone dataset, 50m, are shown in Figure 4.6 and figure 4.7.

b. **OSM**

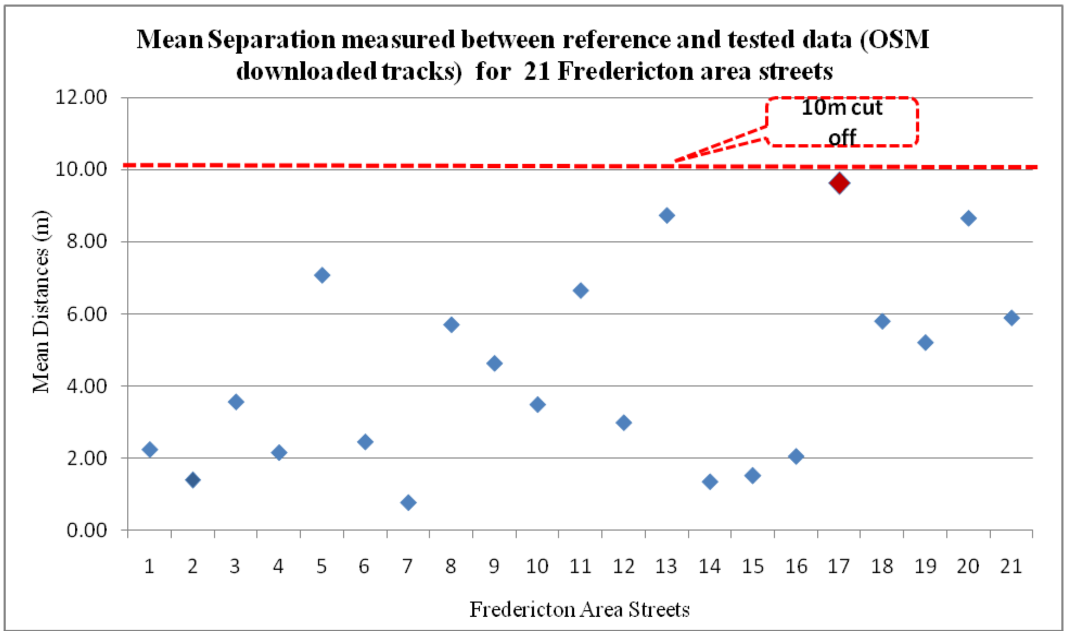


Figure 4.6: Graph showing Mean Separation measured between VGI (OSM streets) and NRN streets

c. **eTrex**

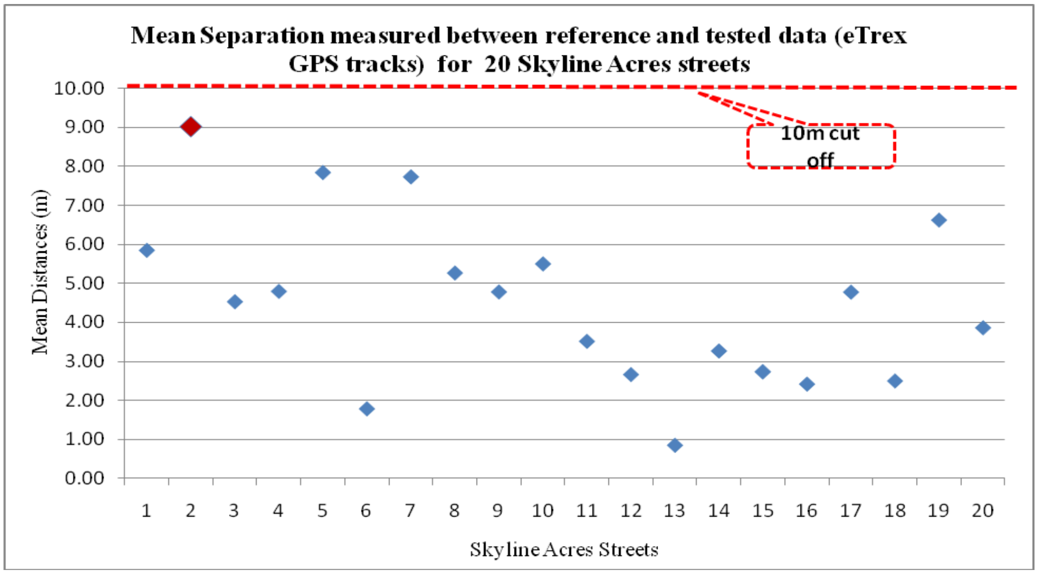


Figure 4.7: Graph showing Mean Separation measured between VGI (eTrex streets) and NRN streets

The OSM and eTrex street datasets complied with the criterion according results shown in the graphs of Figure 4.6 and Figure 4.7. The mean separation measured for each street in both datasets was below 10m. The measured discrepancy between VGI and NRN streets also varied along the same street; however, it appears that most measured discrepancies were less than 10 metres. For example, Table 4.2 shows the measured discrepancy (separation), mean and other statistics for three eTrex streets.

Table 4.2 Absolute Values of Measured Discrepancy between corresponding points for a sample of eTrex streets: including Mean, Standard Deviation and RMS.

Streets	Woodbridge St		Southampton Dr		Bliss Carmen Dr	
	Reference Points	Measured Discrepancy (m)	Reference Points	Measured Discrepancy (m)	Reference Points	Measured Discrepancy (m)
	1	3.795	1	5.061	1	7.175
	2	2.308	2	4.972	2	3.979
	3	2.519	3	3.891	3	3.277
	4	1.100	4	3.382	4	2.394
	5	0.583	5	2.537	5	2.379
	6	2.133	6	2.013	6	3.801
	7	2.274	7	2.298	7	4.962
	8	5.186	8	0.236	8	6.111
	9	4.732	9	0.576	9	6.824
	10	4.211	10	2.935	10	5.951
	11	4.956	11	6.744	11	5.754
	12	5.343	12	4.165	12	5.077
	13	4.494	13	5.937	13	4.672
	14	8.929	14	5.598	14	4.795
	15	13.750	15	5.829		
	16	9.717	16	5.329		
	17	5.162	17	4.179		
Statistics						
Mean		4.776		3.864		4.797
Standard Deviation		3.252		1.834		1.450
RMS		5.778		4.277		5.011

The mean discrepancy (separation) between VGI and NRN streets in Table 4.2 are plotted on the graph in Figure 4.7 along with mean discrepancy for other streets. Table 4.2 also gives the Root Mean Square (RMS) and Standard deviation values. Standard deviation measures the variability or dispersion of a statistical population or dataset (Zhang and Goodchild, 2002), in this case, spread of measured distances between corresponding points in the respective datasets. The standard deviation for Woodbridge Street is higher when compared with Southampton Drive and Bliss Carmen Drive. The larger spread of measured distances for Woodbridge Street could be caused by inconsistency of vegetation cover in the area during data collection. For example, in some places along Woodbridge Street (length approximately 850m) there was heavy vegetation cover and in other places there was clear view of the sky. This may have caused the accuracy of the device to vary along the street during data capture. The lower standard deviation and Root Mean Square (RMS) values for Southampton Drive and Bliss Carmen Drive could be connected with relatively sparse vegetation and openness in those areas; therefore, the accuracy of the device was likely to remain more constant during data capture along these streets.

For the iPhone street dataset, 18 out of 23 (18/23) streets were determined to have a mean separation less than 5m. The average of mean discrepancies for the iPhone dataset is 5m. Eleven of twenty-one (11/21) OSM streets have a mean separation below 4m and the average of mean discrepancies for the dataset is 4.38m. The eTrex dataset has 13 out of 20 streets with mean separation below 5m; the average of mean discrepancies is 4.52m. These results further illustrate the compliance of the VGI streets with the positional accuracy criterion as the average values are less than 10m. Calculating the average of averages is usually risky, as it can yield skewed results if the averages don't have equal weight. However, the average of weighted averages was NOT calculated in this instance because the mean separation values had equal weight. Distances were measured at equal intervals on all streets, which means the sampling rate was the same for all streets within the datasets. However, some streets had more measured distances than others due to their differences in length.

Chapter 4.2.2 Performance of VGI POI data against Positional Accuracy criterion

The evaluation of POI data consisted of measuring the distance between representations of the same point of interest collected by VGI techniques and by more accurate positioning (the control POI). As coordinates of respective points in both datasets were obtained, the distance between the same point in VGI and control representations was deemed to be the "error"– and the measured and calculated distance were compared. This comparison was done to detect errors that may result from manually measuring the distance between datasets. If there were differences between the measured distances and the calculated error, then it would indicate that the process of measuring was not accurate. Table 4.3 gives a summary of the distance measured; error calculated and statistics for the comparison of POI data.

Table 4.3 Summary of Distances measured from Mean VGI POI to Control POI

Feature	Measured DISTANCE (m)	Direction	ΔX (m)	ΔY (m)	Calculated Error (m)
Traffic Light1	2.569	S 64.1531 E	2.312	1.120	2.569
Traffic Light2	5.037	N 3.7566 E	0.330	5.026	5.037
Traffic Light3	2.668	N 78.0372 E	2.610	0.553	2.668
Lamp1	2.022	S 8.6471 E	0.304	1.999	2.022
Statue	2.028	S 85.8158 E	2.023	0.148	2.029
Public Toilets	3.785	N 14.5355 E	0.950	3.664	3.785
Flowerbed	3.747	N 36.2522 E	2.216	3.022	3.747
Museum	5.183	S 12.5936 W	1.130	5.058	5.183
Phone Booth	9.989	N 1.7268 W	0.301	9.984	9.989
City Hall	5.534	N 42.2974 E	3.724	4.093	5.534
Fountain	2.723	N 69.9144 E	2.557	0.935	2.723
Police Light Pole	6.963	S 83.7252 W	6.921	0.762	6.963
TL1	1.581	N 81.1574 E	1.562	0.243	1.581
TL2	3.030	N 70.3449 E	2.853	1.019	3.030
TL3	2.221	N 36.2095 E	1.312	1.792	2.221
TL4	1.311	S 35.3142 W	0.758	1.071	1.312
Statistics					
Mean	3.774		0.853	1.261	3.774
Standard Deviation	2.225		2.413	3.325	2.225
RMS	4.381		2.559	3.556	4.381

The VGI POI data complied with the positional accuracy criterion since the measured distance and calculated errors were less than 30m. The criterion declares that VGI must have a planimetric accuracy of no less than ±10m for roads (in clear unobscured areas), and ± 30m or less for other features (in clear unobscured areas). The table shows that "Measured Distance" values were the same as values for "Calculated Error" for all features. The Mean, Standard Deviation and RMS statistics for the calculated error and measured distance are also the same. The RMS error value, 4.381m, was expected to be higher because the variation in errors was expected to be larger. This is because, during data capture, it was difficult to obtain position fixes to relative accuracies of less than 10m in some cases (when close to buildings). In other situations, it was easier to obtain position fixes to relative accuracies of less than 5m (away from buildings). However, the VGI POI data collection was repeated at least four times for most features and an average value was calculated.

One of the main issues with both the evaluation of VGI street and POI data *and* the subsequent analysis of results is that it was assumed that the control data to which the VGI was compared was more accurate. The accuracy of the NRN and Control POI data affects the analysis of the results, and should be known. The NRN data originates from multiple sources (GPS, satellite imagery, and existing federal, provincial or municipal data) thus it is difficult to determine whether or not the accuracy of the data is uniform throughout the dataset. However, the planimetric accuracy aimed for the data is 10m or better (NRN Data Product Specification, 2007). The Control POI was obtained through Real Time Kinematic (RTK) GPS using New Brunswick Control Monuments (NBCM) as control for the survey. The RMS error for the NBCM used and those measured was 0.018m, which indicates that errors for the control survey varied to a degree of 18mm. The caveat to using control POI collected via RTK GPS is that the accuracy of the control POI is dependent on the accuracy of the NBCM. The monuments used were 24 years old and may have deteriorated over time or been disturbed, thus altering the accuracy of their recorded location.

Chapter 4.3 Performance of VGI datasets against Attribute Accuracy Criterion

Requirements for the attribute accuracy criterion were that VGI attributes should include a name or identifier and descriptive information, and the accuracy will be determined by comparing them with corresponding CGDI datasets attributes. However, only identification of VGI attributes was undertaken because making comparisons with NRN data for checking the accuracy of VGI street attributes was not possible. This is because the VGI streets and NRN datasets did not have similar attributes. The NRN had over 40 attributes compared to up to 5 attributes for VGI streets; and none of the attributes could be compared. Furthermore, the VGI POI attribute data had no descriptive information attached to it. The names of features, or the types of features were not included therefore comparisons to any CGDI data attributes could not be conducted. For these reasons it was determined that the attribute accuracy criterion was only partially fulfilled. It was partially fulfilled because the VGI data attributes did have identifiers and rather than descriptive information, time stamps and positional information were also recorded. For OSM streets, the attributes included users who contributed the data. This information is useful and can be used as metadata as points to where data came from. Table 4.4 gives a summary of attributes of the VGI datasets.

Table 4.4 Attributes of VGI Datasets

VGI Datasets	iPhone Streets	OSM Streets	eTrex Streets	GPSMAP Streets	POI
Attributes	Identifier	Identifier	Identifier	Identifier	Identifier
	Time Stamp	User	Time Stamp	Time Stamp	Coordinates
	Length	Time Stamp	Length	Length	Elevation
		Length			Time Stamp

Chapter 4.4 Performance of VGI datasets against Data Structure Criterion

The VGI datasets partially comply with the data structure criterion which required the data to have; Entity Name, Geometric Representation, Metadata for geometric representation, Descriptive Representation and Identifier. It was used to test the completeness of the datasets. All VGI datasets had geometric representation and some attributes (see Table 4.4). The values for the names of the datasets were changed; however, they originally had the date of their capture as the default name of the dataset. Default values for identifiers of entities within the dataset were also given. For example, in Figure 4.8 the structure of the OSM streets dataset is illustrated through its attribute table.

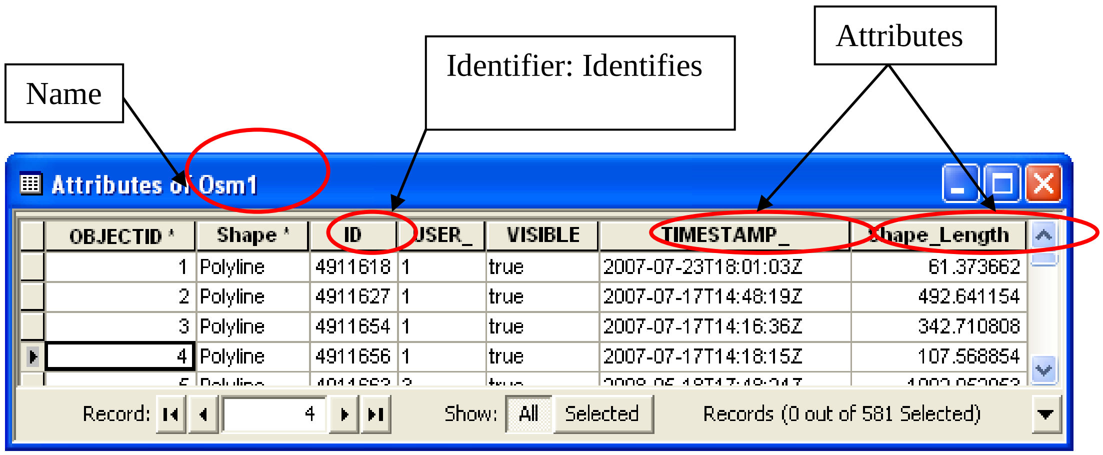


Figure 4.8: Attributes for OSM streets

The criterion was not wholly complied with because metadata for geometric representation existed only in part. Metadata should include the source of data and accuracy of data. However, for the VGI datasets; only the source of OSM streets was unknown and the accuracy of the data for all the datasets was unknown. Despite the lack of accuracy metadata the VGI datasets had most of the elements to be considered complete.

Chapter 4.5 Performance of VGI datasets against Metadata Criterion

As previously mentioned, there was no metadata qualifying the accuracy of data for the VGI datasets. The accuracy of the GPS and iPhone devices during data collection was not automatically recorded by the instrument because it tended to fluctuate and there was no inbuilt mechanism for recording it in the instruments. Therefore, during data collection the accuracy of the devices was noted when possible. Data that was downloaded from OSM did not contain any accuracy information. All datasets had a time stamp showing the date and time of data collection. The source for data collected in the field was known: hand-held GPS receivers. However, this was not the case for the OSM data, although OSM data had information on users who contributed the data included as attributes. This does not give the source of data but at least it can be known who provided data; if real names were used. This way, contributors could be questioned about how they obtained the data if there was need to do so.

It can be deduced that the metadata criterion was not fulfilled for the most part because the accuracy of the VGI data was unknown; even though accuracy during data collection was noted it was not recorded by the devices. Therefore a different investigator, using this data, wouldn't know this information. The source and accuracy of OSM data was unknown, however, the source of the other VGI datasets was known. The source of the other VGI datasets was also not recorded by the devices, but this information can be communicated to different investigators more easily than accuracy information through data documentation processes using software like ArcCatalog.

Chapter 4.6 Performance of VGI datasets against Uncertainty Criterion

The uncertainty criterion declares that the maximum acceptable proportion of errors of classification for attributes should be no more than 5%. As there were no descriptive attributes for the VGI datasets this criterion could not be measured, and it was not possible to measure the accuracy attributes that exist via comparative means as previously mentioned.

The maximum acceptable proportion of positional errors was to be no more than 10%. Estimating the percentage of errors in the VGI street datasets was done by: subtracting from 100, the values of percentage of VGI streets within 10m of NRN streets. The results, shown in Table 4.5, give the percentage of VGI streets that were NOT within the 10m NRN buffer, and are deduced as the errors. Factors contributing to the errors were discussed in Section 4.2.1. The OSM dataset had the least amount of errors as also discussed in Section 4.2.1. The magnitude of these errors is not known and would be difficult to obtain because the separation distance between the lines outside the buffer and the buffer would have to be measured. This would be a very time consuming process and would suffer from the same constraints as the manual comparison method for linear features discussed in Section 4.2.2.

Table 4.5 Percentage of Errors within VGI street datasets

VGI Street Data Source	**Percentage of Errors**
iPhone	17.14%
OSM downloads	5.95%
Garmin eTrex	9.53%
Garmin GPSMAP 76CSx (Edited)	10.19%
Garmin GPSMAP 76CSx (Unedited)	9.63%

The percentage of errors in the POI dataset could not be estimated using the same method for estimating percentage of errors in the VGI streets datasets because they are different data types. Also, since the calculated errors were well below the 30m threshold for all points in the POI dataset (see Table 4.3); it meant that there was no need to determine the percentage of errors exceeding the set limit. Instead the planimetric accuracy of the dataset was calculated using the Circular Map Accuracy Standard (CMAS) - which is used to express the planimetric accuracy for NRN and NTDB data products – and is derived from the equation below:

Standard circular error: $\sigma_c = 0.7071 (\sigma_x^{2} + \sigma_y^{2})^{½}$

σ_x: standard deviation in the X-axis

σ_y: standard deviation in the Y-axis

Circular Map Accuracy Standard: $CMAS = 2.1460\ \sigma_c$ (1)

The CMAS, hence planimetric accuracy for the POI data was 6.234m. This value is substantially lower than the required planimetric accuracy of 30m or less for features other than roads (in clear unobscured areas). It quantifies the overall accuracy of the data which is the goal of the uncertainty criterion. If there were errors in the POI data exceeding the threshold, then their magnitude and percentage could be calculated using the cumulative distribution function.

Chapter 4.7 Critique of Integration Framework

Previous sections examined the performance of VGI contributions against criteria in an effort to determine the reliability of the contributions. Now, it is important to critique the manner in which these volunteered contributions are integrated into authoritative mapping. This section briefly describes the proposed integration framework, developed in the previous chapter, and discusses its advantages and disadvantages, as well as how well it fulfilled the research objectives.

As discussed earlier in Section 3.6, the proposed integration framework features: (1) A wiki style web application for contributing and editing VGI by users; (2) A validation engine serving as a the backend of the web application and a temporary database for validating VGI before integration; (3) An integration phase performed after the validation. Figure 3.9 in chapter 3 shows an outline of the integration framework.

The last objective of the research was to design a framework which integrated VGI that satisfied the criteria and CGDI specification with CGDI datasets. This framework satisfies the objective in two ways:

(1) It is designed to determine whether or not VGI satisfies any criteria or CGDI specifications through the validation engine. The validation engine would employ rules based processing, thus making an assessment of VGI data using any set of rules or criteria set by a mapping agency. Different methods for validation can be selected by the agency and they depend on the type of data being validated. The type of data being validated also determines the extent to which automated validation methods, such Radius Studio™ validation software, can be used. For example, road data validation would be different from invasive species data validation. The latter may require more human intervention.

The second way the framework satisfies the objective is: (2) it includes peer review as part of quality control (through the web application) and it integrates VGI data after validation has taken place. This way, users can participate in editing/modifying and correcting some of the VGI data, however, the final validation is undertaken by mapping agency personnel and they decide what data is kept, modified or discarded. If the data complies with the criteria or rules it can be integrated with other datasets and committed to the main database using techniques like conflation that include rubber sheeting. If the data doesn't comply but is considered useful, it can be modified using the same integration techniques.

The advantages of this framework are that it provides a holistic view on integration of VGI with authoritative datasets like CGDI datasets. The framework also goes beyond integration of VGI with authoritative datasets. It includes how an organization seeking to use VGI would solicit it from the general public, and allow the public to edit/modify contributions as an interim step before actually validating the data themselves. This is achieved through the wiki style web application which provides a platform for users to contribute, edit/modify/review contributions and to view data once it has been updated with VGI contributions.

The framework also includes validation as part of the integration process, where validation occurs before VGI can be committed to the permanent database. Methods for validation and integration are also included, but the list is not exhaustive. Another advantage is that the framework is designed to outline the cyclic process that VGI would go through if it was to be utilized by an authoritative organization. The process starts from obtaining VGI, to validating and integrating it, to updating existing data with VGI, then finally to providing the updated data to users.

The limitations of framework are that it was derived from similar Map Insight and NES methods for validation and integration which used users' feedback and change reports for updating their databases. The only different approach reviewed was CTI's updating and validation method for its topographic database. However, the CTI method did not include VGI, but the decision tree validation method could be applied validating VGI. Assessing other examples of how VGI is utilized in government and industry could provide more insight on which approaches work best.

The framework could also be improved by including a mechanism to assess the credibility of contributors as part of the overall quality control (Nkhwanana, 2009). The framework illustrates a general process for integrating VGI with authoritative datasets; however, consideration for having a system that will be able to handle the dynamic and evolving nature of VGI needs to be taken into account. This means validation and integration techniques utilized should be mostly automated, and that the system should be flexible enough to accommodate the different types of contributions that were requested. It also means that mapping agencies should anticipate the types and amount of data they will receive in order to be better prepared and to avoid being inundated with data. They should also inform potential contributors about collecting data which would be most useful through guidelines.

Chapter 4.8 Summary

The performance of VGI against the set criteria was analyzed in this chapter. The results indicate that for the most part the VGI datasets conformed to the positional accuracy and data structure criteria. Measuring the performance of the VGI datasets against the attribute accuracy criterion was not possible because attributes for the VGI datasets could not be compared to the NRN dataset or any other CGDI dataset; due to the fact that they are different. The Metadata criterion was only partially fulfilled because the VGI datasets had time stamps and the source of data was known for all data except the OSM streets. However, the essential component of the metadata criterion -- the positional accuracy of data -- was not known for all datasets; thus the criterion was not wholly met. The uncertainty criterion was also only partially satisfied because quantifying the errors for attribute classification was not possible. Still, the overall accuracy of the datasets was quantified since the percentages of errors in the VGI street datasets were determined and the planimetric accuracy for the POI dataset was determined.

The advantages and shortcomings of the integration framework were discussed in this chapter, and the framework was briefly assessed in terms of how well it met final objective of this research -- to design a framework which integrated VGI that satisfied the criteria and CGDI specifications with CGDI datasets. It was determined that framework did meet the objective; however, improvements were suggested. They include incorporating a wider variety of validation and integration methods, to cater to different data types. This chapter sought to analyze the results and findings of the work done to ultimately answer the research question. The next and final chapter summarizes the degree to which the original research objectives were met and offers recommendations for future research.

Chapter 5 Conclusions

The overall goal of this research was to examine the suitability of VGI being used to augment SDI, i.e. Canadian Geospatial Data Infrastructure (CGDI), datasets. It investigated the extent to which VGI enabling technologies, including Location Based Services (LBS) positional and data acquisition techniques, ensure accuracy compliant with CGDI accuracy standards.

The characteristics of VGI and the influence of LBS positional techniques on the accuracy of VGI created via these means were examined in Chapter 2. Furthermore, the accuracy of VGI was evaluated in Chapters 3 and 4 to determine if it was compliant with CGDI accuracy specifications/standards. Methods that could be used to integrate VGI with CGDI datasets were also investigated, and a framework was developed to integrate VGI that satisfied CGDI data accuracy specification and standards with CGDI datasets (also in Chapters 3 and 4).

This chapter summarizes the outcomes of the research including how successfully the original research objectives were met and the research problem solved. The chapter concludes by discussing recommendations for future research.

Chapter 5.1 Research Outcomes and Issues encountered

The original research objectives, their outcomes and the conclusions drawn from the research investigation are provided below:

Objectives and Outcomes:

1. ***To identify and characterize the 3 leading LBS positional technologies and processes used for VGI generation 95% of the time, indicating their strengths and limitations***.

- Through the literature review it was found that GPS is the most common positioning technique employed, particularly in safety, navigation and tracking areas of LBS usage. The next most common technique employed in these areas of usage is Assisted-GPS (A-GPS) which employs other techniques like Wi-Fi and Cell of Origin (COO) when GPS is unavailable. It was not possible to quantify which technique was used for creating VGI most of the time because there are different LBS areas of usage and applications which use different positioning techniques which could be used to generate VGI. However it was estimated that navigation and tracking applications which employ GPS and A-GPS techniques would most likely be used for creating most VGI.
- The less-than ±10m accuracy of GPS and A-GPS techniques outdoors was one of their strengths. However, the low accuracy of these positioning techniques in built-up areas (up to ±50m) was a major limitation because most VGI is collected in such areas.

2. ***To identify the most common VGI data types, available 95% of the time on 5 well known websites with more than 1 million entries.***

- Points of interest, place names and linear features were the most common form of VGI available on websites which allow users to create and share spatial data and maps (see Table 2.1). However, it depended on the particular website being viewed and what its purpose was, because not all the sites had the same features. For example *OpenStreetMap* mostly had linear features since their goal is to build a map of the world's road networks and *Wikimapia* had polygons and place name because their goal is to describe the world.

3. ***To use the identified data types to develop criteria which assess VGI quality in terms of positional and attribute accuracy and uncertainty in order to ensure compliance with CGDI data accuracy standards 90% of the time.***

- Criteria which assess the following aspects of VGI quality were developed: positional accuracy, attribute accuracy, data structure, metadata and uncertainty. They were developed by comparing and selecting CGDI data accuracy standards and specifications for three CGDI data products with national coverage. Testing the criteria would also entail comparing VGI with CGDI datasets. Other factors that influenced the development of criteria included LBS positional accuracies and VGI characteristics.

4. ***To test a small sample of existing and newly created VGI at Canadian local, provincial and federal levels against that criteria and use them to compare with a small sample of existing CGDI data and accuracy standards.***

- VGI data collected from online sources and from the field using a LBS device and handheld GPS devices were tested against the criteria – which also included comparing with similar CGDI datasets. It can be argued that the data collected from the field is not strictly VGI because it has not been published or shared anywhere. However, the devices used and manner in which the data was collected were considered to be those which a volunteer could use to collect spatial data.
- It was found that four out of five (4/5) VGI street datasets conformed to the positional accuracy criterion and had less than 10% of positional errors. The required planimetric accuracy for CGDI datasets is no less than ±10m for roads in clear unobscured areas, and ±30m or less for other features (again, in clear unobscured areas). The accuracy of VGI street datasets was determined by comparing them with the National Road Network (NRN) – a CGDI dataset determined to have the required planimetric accuracy for roads. The percentage of VGI street centerline datasets within 10m of NRN streets was above 80% for all VGI datasets. And, the calculated planimetric accuracy for the VGI points of interest dataset was below 10m.
- The VGI datasets only partially fulfilled the data structure, metadata and uncertainty criteria, and the accuracy of VGI dataset's attributes could not be determined. This is because the attribute accuracy criterion and part of the uncertainty criterion required VGI dataset's attributes to be compared with similar

CGDI dataset's attributes. However, there were no VGI or CGDI datasets with similar attributes to compare. The metadata and data structure criteria were only partially fulfilled because there was no metadata regarding the geometric representation for the VGI datasets.

5. ***To establish a framework which seeks to implement the use of VGI in a SDI, using mechanisms that ensure data accuracy compatibility.***

- The framework was designed to determine whether or not VGI satisfies any criteria or CGDI specifications through a validation engine and it integrates VGI after validation has taken place. It provides a holistic view on integration of VGI with authoritative datasets like CGDI datasets. Automated methods for validating (e.g. rules based processing) and integrating (conflation) VGI are suggested in order to be able to cope with large amounts or evolving VGI contributions.

It can be concluded that Objectives 3 – 5 were satisfied, whilst Objectives 1 and 2 were only partially satisfied. Since the objectives were met or partially met in some cases, it can be concluded that the research problem was partially solved. This is because only the extent that VGI meets CGDI *positional* data accuracy specifications can be quantified. This is not the case for *attribute* accuracy, metadata, data structure or uncertainty specifications. Original research contributions emerging from the research outcomes include: discovery of suitability of VGI using criteria derived from existing CGDI standards; the criteria themselves, which were used to evaluate the accuracy compatibility of VGI with CGDI datasets; and demonstration of a holistic approach for integrating VGI with authoritative datasets, like CGDI datasets, which incorporates data validation and data sourcing.

Chapter 5.2 Recommendations for Future Research

1. Different LBS devices, apart from those using GPS, should be used in the collection of VGI. This will ensure that the influence of LBS positioning techniques on accuracy of VGI is actually investigated and that the investigation is more inclusive, because different positioning techniques and devices will be examined. In this research the devices employed GPS to obtain locations; therefore it can be argued that the influence of GPS on the accuracy of VGI was being investigated, even though it was not intended to be so.
2. Extensively reviewing more examples of VGI validation and integration processes in industry, government and those proposed by academia for different data types would provide better insight on integration methods that would be more suitable for VGI.
3. Implementation of the framework: Investigating whether or not it would be feasible for a mapping agency to collect VGI from public and use it to update their databases within their mandates and update cycles in a pilot study. This would require development of the web application that would allow users to contribute and edit VGI; and a system that uses automated methods to validate and integrate contributions after validation. Investigating the performance of the integration framework would determine how much effort (labour, cost, time) it would take for a mapping agency to obtain, validate and integrate VGI. It would also determine if indeed updating cycles could be improved by using VGI. Obstacles that may hinder

the use of VGI by such agencies, e.g. institutional structures and policies, would also be explored.

Chapter 5.3 Final Remarks

The suitability of using VGI to augment CGDI datasets was determined; because the general conclusion is that VGI generated from LBS which employ GPS positioning meets CGDI positional accuracy requirements. However, the research was limited because the main source of the VGI data was from devices using GPS to obtain location. It would have been ideal if other LBS positioning methods were investigated.

The research was also limited because the accuracy of VGI attributes could not be evaluated. This was caused by the fact that VGI and CGDI attributes were different therefore comparing them was not possible. A general approach for validating and integrating VGI with authoritative datasets, like CGDI datasets, was the main contribution of this research. However, further investigation of different methods that validate and integrate spatial data is required to improve the framework.

Bibliography

1Spatial, (2008). “Radius Studio Concept Guide Version 2.0”, 1Spatial Radius Studio™. Retrieved from *http://www.1spatial.com/pdf/rs_concepts.pdf*.

ANZLIC (1996). Spatial data infrastructure for Australia and New Zealand. *http://www.anzlic.org.au/asdi/anzdiscu.htm*, last access: December, 2008.

Armenakis, C., I. Cyr, and E. Papanikolaou (2002). **“**Change Detection Methods for the Revision of Topographic Databases”, *Proceedings of Joint International Symposium (ISPRS IV, SDH, CIG) on Geospatial Theory, Processing and Applications*, July 9-12, 2002, Ottawa, Canada.

Averkamp, J., (2007). “Finders Keepers”, *GPS World, Vol.* 18, no. 10, pp. 38-43.

Barnes, S. J., (2003). "Location-Based Services," *E-Service Journal*, Vol.2 no. 3, pp. 59-70.

Bernardos, A., Casar, J., and Tarrio, P. (2007). “Building a frameworkto characterize location-based services,” *InternationalConference on Next Generation Mobile Applications, Services and Technologies*,2007 (Cardiff, Sep 2007), IEEE, pp. 110-118.

Bishr, M. and Kuhn, W. (2007). “Geospatial Information Bottom- Up: A Matter of Trust and Semantics,” In *The European Information Society,* Springer, Berlin Heidelberg, pp. 365 – 387.

Bruns, A. (2006) ‘Towards produsage: Futures for user-led content production’, in F. Sudweeks, H. Hrachovec, and C. Ess (eds) Proceedings: Cultural Attitudes towards Communication and Technology 2006, Murdoch University, Perth, pp. 275-84. *http://snurb.info/files/12132812018_towards_ produsage_0.pdf*

Bruns, A. (2008). *Blogs, Wikipedia, Second Life, and Beyond: From Production to Produsage*. NewYork: Peter Lang.

Brannstrom, M., and Martenson, C. (2006). “Enhancing situational awareness by expoiting wiki technology,” Proceedings of Conference on Civil-Military Cooperation (CIMI) 2006.

Budhathoki, N.R., Bruce, B. and Z. Nedovic-Budic (2008). "Reconceptualizing the role of the user of spatial data infrastructure", *GeoJournal*, Vol. 72, no. 3-4, pp. 149-160.

Canada, Natural Resources (2007). “ CanVec Data Product Specifications 1st Edition,” Earth Sciences Sector, Centre for Topographic Information, Sherbrooke, Quebec, Canada, pp 16. Retrieved from *http://ftp2.cits.rncan.gc.ca/pub/canvec/doc/ CanVec_product_specifications_en.pdf*

Canada, Natural Resources (2007). “GeoBase National Road Network Data Product Specifications 2nd Edition”, Earth Sciences Sector, Centre for Topographic Information, Sherbrooke, Quebec, Canada, 18pp. Retrieved from *http://www.geobase.ca/doc/specs/pdf/GeoBase_DataProductSpecifications_NRN_EN.pdf*

Canada, Natural Resources (1996). “Standards and Specifications of National Topographic Database”, Earth Sciences Sector, Centre for Topographic Information, Sherbrooke, Quebec, Canada, Retrieved from *ftp://ftp2.cits.rncan.gc.ca/pub/bndt/doc/stdntdb3_en.pdf*

Christmas Bird Count, (2009). Christmas Bird Count History. *http://www.audubon.org/Bird/cbc/*, last access: August, 2009.

Chen C-C, Knoblock C. A, Shahabi C. (2006). “Automatically conflating road vector data with orthoimagery,” *Geoinformatica*, Vol. 10, no. 4, pp. 495–530.

Cobb M. A, Chung M. J, Foley H, Petry F. E, Shaw K. B, and Miller V. (1998). “ A rule based approach for the conflation of attributed vector data”, *Geoinformatica*, Vol. 2, no. 1, pp. 7-35.

Coleman, D.J., Sabone, B. and Nkhwanana, N. (2009). "Volunteering Geographic Information to Authoritative Databases: Linking Contributor Motivations to Program Effectiveness". Submitted for review to *Geomatica* Special Issue on Volunteered Geographic Information. June.

Dao, D., Rizos, C., and Wang, J., (2002). "Location-Based Services: Technical and Business Issues," *GPS Solutions*, Vol. 6, no. 3, pp. 169-178.

Department of Sustainability & Environment, (2008). “Notification & Edit System (NES) User Manual”, State government of Victoria, Australia, 99 pp. Retrieved from *http://nes.land.vic.gov.au/WebSite/help/NES%20User%20Manual.pdf*

D'Roza, T., and Bilchev, G., (2003). "An Overview of Location-Based Services," *BT Technology Journal*, Vol. 21, no. 1 pp. 20-27.

Elwood, S. (2008a). “Geographic Information Science: New geovisualization technologies – emerging questions and linkages with GIScience Research.” *Progress in Human Geography*, pp. 1-8.

Elwood, S. (2008b). “Volunteered geographic information: future research direction motivated by critical, participatory and feminist GIS.” *GeoJournal*, Vol. 72, no. 3, pp. 173 – 183.

ESRI Avenue Scripts, (2009). Osm2shape.avx. *http://arcscripts.esri.com/details.asp?dbid=15724*, downloaded, July 2009.

Flanagin, A. J., and Metzger, M. J. (2008). "The credibility of volunteered geographic information." *GeoJournal*, Vol.72, no. 3, pp. 137 – 148.

Gartner, G., Bennett, D., and Morita, T. (2007). "Toward ubiquitous cartography." *Cartography and Geographic Information Science*, Vol. 34, no 4, pp. 247–257.

GeoConnections, (2009). About the Canadian Geospatial Data Infrastructure. *http://www.geoconnections.org/en/aboutcgdi.html*, last access: July, 2009.

GeoConnections, (2004). "A Developers' Guide to the CGDI: Developing and publishing geographic information, data and associated services", Canadian Geospatial Data Infrastucture Technical Guide, GeoConnections, Ottawa, Canada, 253 pp. Retrieved from *http://www.geoconnections.org/publications/ - Technical_Manual/html_e/cgdiindex.html*.

GeoGratis, (2009). About National Topographic Database (NTDB) Canada, *http://geogratis.ca/geogratis/en/product/search.do?id=8147,* last access: May, 2009.

GIS-LAB, (2009). OSM to Shapefile conversion. *http://gis-lab.info/qa/osm2shape.html*, last access: July, 2009.

GLOBE: The GLOBE Program, (2009). About GLOBE. *http://www.globe.gov/,* last access: August, 2009.

Goodchild, M. F. (2007a). "Citizens as sensors: The world of volunteered geography." *GeoJournal*, Vol. 69, no. 4, pp. 211–221.

Goodchild, M. F. (2007b). "Citizens as voluntary sensors: Spatial data infrastructures in the world of Web 2.0." *International Journal of Spatial Data Infrastructure Research*, Vol. 2, pp. 24–32.

Goodchild, M.F. (2008). "Commentary: Whither VGI?" *GeoJournal*, Vol. 72, no. 3, pp. 239-244.

Goodchild M.F. and Hunter G.J., (1997). "A simple positional accuracy measure for linear features," *International Journal of Geographical Information Systems*, Vol. 11, no. 3, pp. 299-306.

Gouviea, C., Fonseca, A., Camara, A., and Ferreira, F. (2004). "Promoting the use of environmental data collected by concerned citizens through information and communication technologies." Journal of Environmental Management, Vol. 71, pp. 135 – 154.

Haklay, M. (2008). “How good is OpenStreetMap information? A comparative study of OpenStreetMap and Ordnance Survey datasets for London and the rest of England,” In *Environment & Planning* (under review).

Henry, B., and Temmink, R. (2008). “Transforming Digital Mapping with Community,” Tele Atlas® Presentation on Digital Mapping. Retrieved from *http://www.teleatlas.com/WhyTeleAtlas/Pressroom/PressKit/ssLINK/TA_CT020397*

Hunter G. J, (1999). “New tools for handling spatial data quality: Moving from academic concepts to practical reality”, *URISA Journal*, Vol. 11, no. 2, pp. 25–34.

IAB (2008). "User Generated Content, Social Media, and Advertising — An Overview". Interactive Advertising Bureau Platform Status Report. April. Retrieved from *http://www.iab.net/media/file/2008_ugc_platform.pdf*

Jensen J., Saalfeld, A., Broome, F., Cowen, D., Price, K., Ramsey, D., and Lapine, L., (1998). “Spatial data acquisition and integration”, University Consortium for Geographic Information Science (UCGIS) White Paper on Research Priority. Retrieved from *http://lab.geog.ntu.edu.tw/course/giss/paper/spatial%20data%20acquisition.pdf*

Kuhn, W. (2007). “Volunteered geographic information and GIScience,” In Workshop on Volunteered Geographic Information, University of California, Santa Barbara, pp. 86–97.

Labonte, J., Corey, M., Evangelatos, T., (1998). “Canadian geospatial data infrastructure (CGDI)—geospatial information for the knowledge economy,” *Geomatica*, (Canadian geoconnections web-site, accessed on June 2009: *http://cgdi.gc.ca/english/publications/index.html*.

Lathrop, R. G., P. Montesano, J. Tesauro, and B. Zarate, (2005). “Statewide mapping and assessment of vernal pool: A New Jersey case study”, *Journal of Environmental Management* 76, pp. 230–38. *http://www.dbcrssa.rutgers.edu/ims/vernal/viewer.htm*

Martin, M. E., Loubier, E. (2002). “Using landsat -7 images to update Canadian national topographic data”, *Proceedings of Joint International Symposium (ISPRS IV, SDH, CIG) on Geospatial Theory, Processing and Applications*, July 9-12, 2002, Ottawa, Canada.

Maue, P. and Schade, S. (2008). “Quality of Geographic Information Patchworks.” 11th AGILE International Conference on Geographic Information Science, University of Girona, Spain, 2008, pp. 1-8.

McDougall, K. (2009). "The Potential of Citizen Volunteered Spatial Information for Building SDI," *Proceedings of GSDI 11*, 15-19 June 2009, Rotterdam, Netherlands.

Natural Resources Canada, (2009). Canadian Spatial Reference System, *http://www.geod.nrcan.gc.ca/edu/index_e.php*, last access: August, 2009.

Nkhwanana, N. (2009). *Assessing Credibility of VGI Contributors and Trust in their Contributions*, M.Sc.E. thesis, Department of Geodesy and Geomatics Engineering, University of New Brunswick, Fredericton, N.B., Canada, Unpublished.

Notification & Edit Service, (2009). About Notification & Edit Service. *http://nes.land.vic.gov.au/WebSite/Login.aspx?ReturnUrl=/WebSite/default.aspx*, last access September, 2009.

OECD 2007. "Participative Web: User-Created Content: Web 2.0, wikis and social networking ". Report DSTI/ICCP/IE(2006)7/FINAL, Working Party on the Information Economy, Committee for Information, Computer and Communications Policy, Directorate for Science, Technology and Industry, Organisation for Economic Co-Operation And Development. 12 April. OECD Publishing. ISBN 9264037462, 9789264037465.

OpenStreetMap (2009). The OpenStreetMap stats report. *http://www.openstreetmap.org/stats/data_stats.html*, last access: March, 2009.

Perkins, C., and Dodge, M. (2008). "The potential of user-generated cartography: a case study of the OpenStreetMap project and Mapchester mapping party", *North West Geography*, Vol. 8, no. 1, pp. 19–32.

Platial: "Who and What is nearby", (2009). About Platial.com. *http://platial.com/*, last access: February, 2009.

Reed, J. H, Krizman, K. J., Woerner, B. D. and Rappaport, T. S. (1998). "An overview of the challenges and progress in meeting the e-911 requirement for location service," *IEEE Communication Magazine*, Vol. 36, pp. 30–37.

Rinner, C., Kessler, C. & Andrulis, S. (2008). "The use of Web 2.0 concepts to support deliberation in spatial decision-making", *Computers, Environment and Urban Systems*, 32, pp.386-395.

Saalfeld, A., (1988). "Conflation: Automated map compilation", *International Journal of Geographical Information Science*, Vol. 2, no. 3, pp. 217-228.

Seeger, C. (2008). "The role of facilitated volunteered geographic information in the landscape planning and site design process." *GeoJournal*, Vol. 72, no. 3, pp. 199 – 213.

Sieber, R. E. (2004). "Rewiring for a GIS/2." *Cartographica: The International Journal for Geographic Information and Geovisualization,* Vol. 39, no. 1, pp. 25-39.

Smith, J., Kealy, A., and Williamson, I., (2001). "Assessing Spatial Data Infrastructure Architecture for Integration with Wireless Location Services", *http://www.Scientificcommons.org/8635591*.

Shirky, C. 2008. "Here Comes Everybody: The Power of Organizing Without Organizations". New York: Penguin Press.

Stojanovic, D.H. and Djordjevic-Kajan, S.J. (2001). "Developing location-based services from a GIS perspective", In *5th International Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Service,* 19-21 Sept, Nis, pp.459-462.

Swedberg, G. (1999). " Ericsson's Mobile Location Solution," Ericsson Review.

Sui, D.Z. (2008). "The wikification of GIS and its consequences: Or Angelina Jolie's new tattoo and the future of GIS", *Computers, Environment and Urban Systems,* Vol. 32, no. 1, pp. 1-5.

Tagzania, (2009). About Tagzania. *http://www.tagzania.com/,* last access: February, 2009.

Tele Atlas®, (2009). Tele Atlas® Map Insight™. *http://www.teleatlas.com/PartnerPrograms/MapInsight/index.htm,* last access: September, 2009.

Tele Atlas®, (2009). Tele Atlas® Multinet®. *http://www.teleatlas.com/OurProducts/MapData/Multinet/index.htm,* last access: September, 2009.

TomTom Mapshare™, (2009). Get to know TomTom Mapshare™. *http://www.clubtomtom.com/general/get-to-know-tomtom-mapshare%E2%84%A2/,* last access: September, 2009.

Tulloch, D. (2008). "Is volunteered geographic information participation?" *GeoJournal,* Vol. 72, no. 3, pp. 161-171

Unni, R. and Harmon, R. (2003), "Location-Based Services: Models for Strategy Development in M-Commerce", In *Technology Management for Reshaping the World*, Piscataway, NJ: IEEE, pp. 416-424.

Virrantaus, K., Markkula, J., Garmash, A., Terziyan, V., Veijalainen, J., Katanosov, A. and Tirri, H., (2001). "Developing GIS-supported location-based services", *In Proceedings of WGIS'2001 – First International Workshop on Web Geographical Information Systems.* Kyoto, Japan (2001), pp. 423–432.

Wayfaring: “Follow you, follow me”, (2009). Explore and Connect. *http://www.wayfaring.com/,* last access: February, 2009.

Woodsford P. A, (2007). “ Spatial Database update and validation – Current advances in theory and technology”, ISPRS Workshop on Updating Databases with Imagery. Retrieved from *http://www.commission4.isprs.org/urumchi/papers/17%20P.%20A.%20Woodsford%20.pdf.*

Zandbergen P.A, (2009). “Accuracy of iPhone Locations: A Comparison of Assisted GPS, WiFi and Cellular Positioning”, *Transactions in GIS*, Vol. 13, (s1), pp. 5–26.

Zhang, J. F., and M. F. Goodchild. (2002). *Uncertainty in Geographical Information*. London:Taylor & Francis.

Appendix I

Location Based Services Components

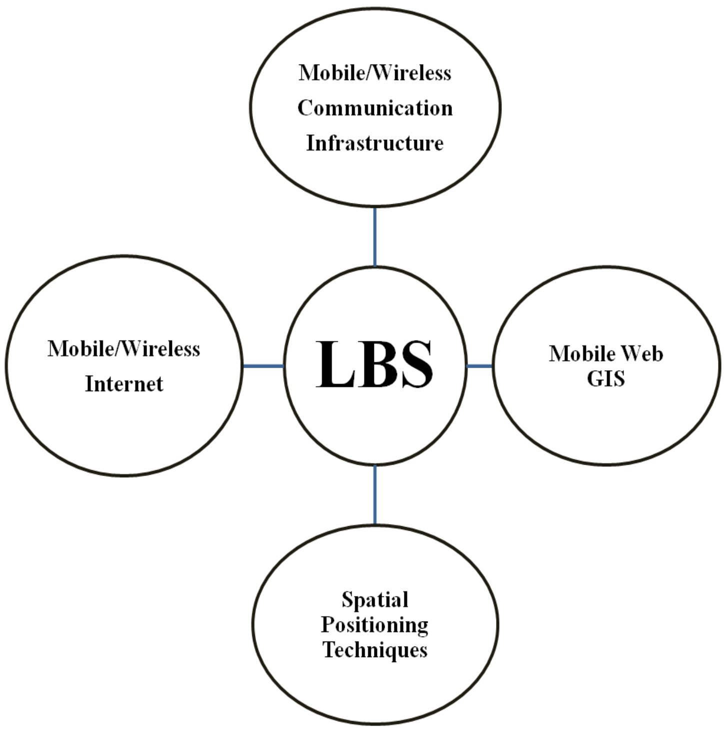


Figure I.1 Location Based Services components

Mobile/Wireless Communication

The mobile/wireless communication infrastructure plays a pivotal role for location based services for it provides the platform on which users can communicate, can send and receive data via internet, and can capture and store location. Mobile Networks have evolved from 1st generation (1G) analog cellular (voice-only, 300-1200 bps) to 2nd generation (2G) digital cellular (voice & data, 9600-19,200 bps) and on to 3rd generation (3G) digital cellular systems with much faster data communication speeds (144 – 500 KB/s) (Averkamp, 2007). 4th generation Internet Protocol (IP) networks follow next and provide even faster data speeds: up to 1 MB/s. 4th generation network that will be based on WiMax (Worldwide Interoperability for Microwave Access) standard have been deployed by mobile networks such as Sprint (Averkamp, 2007). WiMax is the next evolution in wireless broadband, offering ubiquitous internet coverage (WiMax.com, 2009).

The evolving mobile infrastructure present opportunities to enhance LBS capabilities and to overcome some of the existing LBS challenges associated with providing services to customers and data transfer. These issues will be discussed further at the end of the section.

Mobile/ Wireless Internet

Among the 4 billion mobile subscriptions worldwide, an increasing number are becoming internet enabled (GSM, 2008). For example, the Global System for Mobility (GSM) Association reported that there were 200 million 3G cellular phone subscriptions worldwide in 2007. These terminals allow users to surf wireless internet and upload and download data. Wireless mobile communication protocols such as GSM, employ the standards: Short Message Service (SMS) and General Packet Radio System (GPRS), which support LBS (Dao et al, 2002). SMS is used by LBS for providing mobile mapping information, such as turn by turn direction, in text format (Dao et al, 2002). GPRS can be used to overcome the limitations of SMS; primarily the ability to send only 160 characters of text. It is internet enabled and offers data transfer which allows mobile devices to connect to the internet, and data such as maps to be delivered to mobile devices (Dao et al, 2002). With the deployment of the WiMax standard, more internet coverage for mobile users is expected as well as better data transfer capabilities.

Mobile internet and mobile Web GIS are oftentimes grouped together due to the functionalities they offer. However, it is important to note the differences between the two. Mobile internet allows Web GIS applications and functions to be utilized in the mobile environment. Mobile devices are internet enabled to allow data transfer, therefore users can access data via the internet. Mobile internet also permits network carriers to add services to existing networks (Barnes, 2003). Web GIS provides spatio-temporal data and services across the Web (Stojanovic et al, 2002). Since users can access the Web through their mobile devices, they can also gain access to spatial data and Web GIS applications. The growth of LBS also suggests that more Web GIS applications are being designed specifically for mobile users.

Mobile Web GIS

Web-GIS is a shift away from traditional/pure desktop GIS. It provides GIS functionality through the Web and is available to wide audiences of non-expert users, possessing minimal browser technology for basic GIS functionality (zooming and panning of spatial data) (Stojanovic et al, 2002; Virrantaus et al, 2002). This next-generation GIS is composed of specialized Web services (components) that are self-contained applications which can be published, located and invoked across the Web using a wide spectrum of web enabled stationary and mobile devices (Stajanovic et al, 2002). However, developing Web-GIS or mapping applications for the wireless/mobile internet is challenging for several reasons. The major concern is the restricted display capability of mobile devices (Dao et al, 2002; Stajanovic et al, 2002). Apart from the limited map features which can be displayed, the speed of data transmission to mobile devices is much slower compared to wired networks (Dao et al, 2002). Moreover each device speaks a different wireless protocol and supports a variety of different Wireless Markup Languages (WML) (over 30 languages). WML is read and interpreted by the micro-browsers in mobile devices to display information (Dao et al, 2002). These different standards preclude developers from writing every application to individually support every single device available (Dao et al, 2002).

LBS Issues

The main issues faced by location based services are those of developing mapping applications for the mobile internet and interoperability between the converging technologies. Interoperability issues are the greatest and needs to be addressed across: (1) Wireless systems/networks (GSM, TDMA and CDMA), (2) Positioning technologies (COO, AOA, EOTD and A-GPS), and (3) Applications: different network and contents interfaces (SMS, GPRS and WAP); and different content formats (maps, languages and routes) (Dao et al, 2002). Achieving the full value for LBS depends on consistent communication across different regions, technology platforms, networks, application domains, and classes of products (Reichardt, 2001). With goals to address LBS interoperability issues, the Location Interoperability Forum (LIF), an organization dealing with LBS specifications, has the following main objectives:

- Reduce/limit the multiplicity of positioning technologies to be deployed.
- Promote common methods and interfaces for standards-based positioning technologies (COO, EOTD and A-GPS)
- Define common interfaces and methods between applications and the wireless networks irrespective of their underlying air interfaces and positioning technologies.
- Adopt common interfaces between applications and different types of content engines and databases (Location Interoperability Forum, 2001)

In developing mapping applications for the mobile environment the challenges encountered include; restricted display capability of devices; slow speed of data transmission; and each device speaks a different wireless protocol and supports a variety of Wireless Markup Languages (over 30 such languages) (Unni and Harmon, 2002; Dao et al, 2002; Stojanovic et al, 2001). For example Wireless Application Protocol (WAP) enabled devices support WML; Palm Operating System devices support TTML (Tagged Text Markup Language); and Voice-activated applications - Voice- XML and VoxML mark-up (Dao et al, 2002; Unni and Harmon, 2002). Mobile mapping requires standards which allow data to be easily transferred and displayed across the mobile internet to a wide variety of different mobile devices (Dao et al, 2002). Scalable Vector Graphics (SVG) and Geography Markup Language are two important standards for displaying spatial data on the standard (Dao et al, 2002; Virrantaus et al, 2002). In the mobile internet environment there is need to convert standard internet mark-up languages to languages which mobile devices understand, such as those previously listed.

Other issues include how to protect the privacy of customers whose location information is available to service providers. Improper and unauthorized use of this information threatens an individual's privacy. The impact of LBS issues on VGI relates to the services capabilities to offer high spatial data quality and an environment of information sharing. Minimizing the number of positional techniques employed by LBS will hopefully minimize spatial variability in VGI created via LBS. However, this is not a guarantee because not all VGI is generated by LBS; other data can be volunteered via different means. What it will do is give those utilizing the data a better idea of its accuracy, provided the source of data is included in metadata. Common interfaces and methods between applications and the wireless networks should allow more seamless data sharing, across different networks, and allow more users subscribe to services and to participate in VGI related activities.

VITA

Candidate's full name: Botshelo Onalenna Sabone

Universities attended: University of South Australia, Bachelor of Geographic Information Systems, 2006

Publication (s): Coleman, D.J., Sabone, B. and Nkhwanana, N. (2009). "Volunteering Geographic Information to Authoritative Databases: Linking Contributor Motivations to Program Effectiveness". Submitted for review to *Geomatica* Special Issue on Volunteered Geographic Information. June.

Sabone, B. (2009). "Assessing Alternative Technologies for Use of Volunteered Geographic Information in Authoritative Databases," GEOIDE 41 Working Paper xx, University of New Brunswick, Fredericton, N.B.

Conference Presentations: "Assessing Contributor Motivations and Enabling Technologies in Volunteered Geographic Information", Session: Oceans Accountability: Linking Science and Technology, at Ocean Management Research Network 2009 (OMRN) National Conference, 21-24 October, Ottawa, ON, Canada.