

Volunteered Geographic Information:
A Review of the OpenStreetMap.org Project
and Factors Relating to its Reliability

by

Scott George

A thesis submitted to the
Faculty of Graduate and Postdoctoral Affairs
in partial fulfillment of the requirements for the degree of

Master of Science

in

Geography with Specialization in Data Science

Carleton University
Ottawa, Ontario

© 2017

Scott George

ABSTRACT

In the field of Volunteered Geographic Information usage, understanding the historical background from which information came can be just as important as the content. This study reviews the development of the OpenStreetMap.org dataset and its potential to provide the most current information about a region following natural disasters. First, a review of the related studies into the use of this type of geographic information and description of the types and motivations of contributors is conducted. Second, contribution activity surrounding four major earthquake events and the community's response are reviewed. A process workflow is derived for comparing the spatial variance of OpenStreetMap.org data to that of a trusted dataset with the further goal of supporting the data's use in operations. A scenario is then used to test the workflow with the final output of cartographic products geared towards using OpenStreetMap.org data to prioritize data collection during ongoing operations.

DEDICATION AND ACKNOWLEDGEMENTS

I would like to first and foremost dedicate this undertaking to my wife Monica who has provided the faith and support needed to complete this work and be my partner and best friend through many great life trials. The birth of and endless care for our son during this time have been a true testament to her steadfast dedication to family and the tireless effort she provided to supporting my efforts in this work.

Next, it gives me great pleasure to acknowledge the skillful support and guidance of my thesis supervisor, Dr. Scott Mitchell, for his invaluable constructive criticism and advice during the course of this research. His insightful feedback has steered this project and ensured that it meets rigorous academic standards. I would also be remiss to fail to mention the inputs of Dr. Andrew Davidson. His initial coaching on the strengths and limitations of the path chosen for this study helped set the stage for what is contained within.

I would also like to thank OpenStreetMap.org/The OpenStreetMap Foundation and its affiliates along with the Provincial Government of New Brunswick, GeoNB (www.snb.ca), for the data used in much of this work as well as my appreciation to the previous researchers, whose work with VGI and the OpenStreetMap project has inspired me.

I share the credit for this undertaking with the Canadian Department of National Defence and the Canadian Armed Forces for their belief in my abilities and for the time and funding provided to me to complete this project

Lastly, I would like to give special thanks to all of those in the Department of Geography and Environmental Studies at Carleton University for the experience of

studying and learning with them. The faculty and students have enriched this experience on so many levels that words cannot alone describe nor properly praise them.

TABLE OF CONTENTS

ABSTRACT	ii
DEDICATION AND ACKNOWLEDGEMENTS	iii
TABLE OF CONTENTS	v
LIST OF FIGURES	vii
LIST OF TABLES	x
LIST OF KEY WEBSITES	xi
ACRONYMS/ABBREVIATIONS USED	xii
Chapter 1 : Introduction and Overview	1
1.1 Introduction	1
1.2 Thesis Context and Objectives	3
1.3 The OpenStreetMap.org Project.....	5
1.4 The Research Question.....	11
Chapter 2 : A Review of the Literature on VGI and the Current State of OSM Data Analysis	15
2.1. The Current State of Voluntary Geographic Information Research.....	15
2.2. New Information Collection and Use Concepts.....	29
2.3 Use of OSM in Natural Disaster (Earthquake) Scenarios	37
2.4 Related Works and Proposed Usage of VGI/OSM Data for This Study and Beyond.....	41
2.5. Summary.....	45
Chapter 3 : Methods and Study Area	47
3.1 Introduction	47
3.2 Study Area Selection	49
3.3 Analysis of the Historical Data Associated with the OSM Dataset	57
3.4 A Process for OSM Dataset Usage in Natural Disaster Response	66
3.5 Thematic Analysis	83
3.6 Rapid OSM dataset Review.....	89
3.7 Final notes on the development of the Idealized Process Workflow	92
Chapter 4 : Results and Discussion	96
4.1 A Review of the Historical OSM Contributions Surrounding Natural Disasters	96

4.2 Spatial and Temporal Analysis and Use of OSM Data	117
4.3 A Final Note on the Results Obtained.....	124
Chapter 5 : Summary and Conclusions	125
5.1 Research Summary	125
5.2 Conclusions and Future Research	128
REFERENCES.....	130
KEY DEFINITIONS	135
ANNEX A – Workflow for Historical Review of the OSM Dataset	136
APPENDIX 1 – Number of Active Contributors and Total Volume of Contributions	137
APPENDIX 2 – Actions Being Carried Out on the OSM Dataset.....	140
APPENDIX 3 – Entity Types Contributed to the OSM Dataset	142
APPENDIX 4 – Unique Contributor Activities for a One Month Period Following a Major Earthquake	146
ANNEX B – Fictional Scenario Generated to Test Thesis Workflow.....	147
APPENDIX 1 – Thesis Trial Process Workflow	148
APPENDIX 2 – Workflow for Rendering of Temporal Differences	149
APPENDIX 3 – IDW Temporal Analysis.....	150
APPENDIX 4 – IDW Version Analysis.....	151
APPENDIX 5 – AOI used for Thesis Trial Scenario	152

LIST OF FIGURES

Figure 1.1 – Growth of OpenStreetMap users since its inception	5
Figure 2.1 – Volunteered Geographic Information: the Nature and Motivation of Producers (Coleman et al., 2009).....	16
Figure 2.2 – Levels of Participation and Engagement in Citizen Science Projects (Haklay, 2013)	18
Figure 2.3 – Rinkwatch GUI (www.rinkwatch.org).....	23
Figure 2.4 – An example of the buffer method used by Haklay et al. (2010)	33
Figure 2.5 – ITN of London, England with data sample areas utilized for the validation of OSM data (Haklay et al., 2010)	33
Figure 2.6 – The 25 Intrinsic Quality Indicators utilized by iOSMANalyzer	36
Figure 2.7 – Maps from OSM data showing the Haiti capitol of Port-Au-Prince before the earthquake, two days after earthquake and how it looked 1 year later.	38
Figure 2.8 – Comparison of OSM contribution volumes in the first month following a severe earthquake in Developed and Developing Infrastructure countries.....	39
Figure 2.9 – Potential Workflow for OSM use/inclusion in a Natural Disaster Aid Operation.....	43
Figure 3.1 - Locations of magnitude 8.0 or greater earthquakes ever recorded	50
Figure 3.2 – Timeline associated with the most recent contributions to the OSM dataset for Haiti (OpenStreetMap Analytics <small>beta</small>).....	52
Figure 3.3 – 2008 New Brunswick Flood Extent	55
Figure 3.4 – Images of the 2008 Fredericton, New Brunswick Flooding (Source: 4 ESR)	56
Figure 3.5 – A new road (red) being created and connected to an existing one (blue)	60
Figure 3.6 – Conceptual Process for the Analysis of osh.pbf Data for employment in the regular review of the OSM Dataset’s growth with regards to contribution reliability	62
Figure 3.7 – Process Workflow for Extracting OSM History data utilizing the VGI Analytics Framework and Microsoft EXCEL	63

Figure 3.8 – Quadtree structure representing the spatial indexing of the operations generated from a September 2015 OSM Full History File (Rehrl and Gröchenig, 2016)	65
Figure 3.9 – Spatial Variance Analysis Method used in reviewing .osm data from Geofabrik.de.....	69
Figure 3.10 – 1m interval buffers extended out from the Geobrik.de OSM data roadways	70
Figure 3.11 – Geofabrik.de Spatial Variance from the NBRN Analysis Results	72
Figure 3.12 – Fictional CAF Orders Generated for Thesis Test Scenario	74
Figure 3.13 – Process workflow for utilization of raw OSM data in spatial variance analysis.....	75
Figure 3.14 – AOI utilized for Thesis Process Workflow Testing	76
Figure 3.15 – Spatial variance of Overpass data to that of the NBRN analysis results....	78
Figure 3.16 – Process Workflow for Temporal Review of Contributions to the OSM dataset.....	80
Figure 3.17 – IDW interpolation results for Timestamps and Version Numbers within the thesis test scenario, New Brunswick AOI.....	82
Figure 3.18 – Carleton University athletic complex, OSM data vs. CanVec+ data	83
Figure 3.19 – Buffer Analysis of building (hospital) centroids between OSM (Test) data and Trusted (True) data (Zheng and Zheng, 2014).....	89
Figure 3.20 – KeepRight.at Error Count Breakdown Map.....	91
Figure 4.1 – OSM Contributions in the month following the 2010 Haiti earthquake	97
Figure 4.2 – Comparative review of the process results from Section 3.3(Annex A, Appendix 1).....	99
Figure 4.3 – OSM contributor activities for a one month period following a major earthquake	103
Figure 4.4 – OSM contribution activity for the province of New Brunswick surrounding the 2008 floods.....	105
Figure 4.5 – Current Road Contribution timestamps for the province of New Brunswick	106

Figure 4.6 - AOI contributors and their volumes of contributions to the current OSM dataset.....	107
Figure 4.7 – Analysis of Unique Contributors (Ranked) to the OSM dataset for the Thesis Test Scenario Study Area and their Volume of Contributions Made	108
Figure 4.8 – Data sources in the AOI and the volumes of Contributions by Contributor	110
Figure 4.9 – highway_Line contributions and their related data sources for the AOI....	115
Figure 4.10 – Thesis Test Scenario Buffer Analysis Results.....	118
Figure 4.11- IDW Interpolation of Timestamps for the Thesis Test Scenario AOI	120
Figure 4.12 – Magnified view of Thesis Test Scenario Temporal Interpolation Output (Scale ~ 1:50,000)	122
Figure 4.13 – Temporal Analysis of Carleton Campus OSM data (Scale ~ 1:10,000) ..	123

LIST OF TABLES

Table 2-1 – Regions of interest which were struck by a high magnitude earthquake event	40
Table 2-2 – Related OSM Accuracy and Reliability Studies	44
Table 3-1 – Data sources used and time periods reviewed in the Historical Analysis of the OSM History (.osh.pbf) for the different regions.....	58
Table 3-2 – Executable Operation Types on the Different OSM Data Types (Rehrl and Gröchenig, 2016).....	59
Table 3-3 – Concepts of Activity Theory applied to VGI mapping activities (Rehrl et al., 2013).....	Error! Bookmark not defined.
Table 3-4 – OSM executable operation sequence used for the mapping action of adding a road.....	61
Table 3-5 – Data Sources used in Spatial Comparison Process Development	66
Table 3-6 – OSM Attribute Fields for the buildings in Figure 3.21	84
Table 3-7 – CanVEC+ Polygon Attributes	85
Table 3-8 – Automatically Generated OSM Attributes	85
Table 3-9 – User Generated OSM Attributes	86
Table 3-10 – Risk Relationship between Number of Errors and Volume of Contributions	92
Table 3-11 – MS-DOS/Windows based OSM processing tools trialed.....	93
Table 4-1 – Breakdown of AOI Contributors and their Volumes of Contributions	108
Table 4-2 – Detailed breakdown of the Contribution Data Sources for the Current AOI dataset.....	111
Table 4-3 – Volumes of AOI highway_Line Contributions and their Associated Version Numbers	114
Table 4-4 – Breakdown of Total highway_Line segments and their Contribution Data Source for the AOI	116

LIST OF KEY WEBSITES

<u>NAME/ACRONYM</u>	<u>WEBSITE</u>
FEMA	https://www.fema.gov/
GeoGratis	http://www.nrcan.gc.ca/earth-sciences/geography/topographic-information/free-data-geogratias/11042
GeoNB	http://www.snb.ca/geonb1/
HOT	https://www.hotosm.org/
ISO	https://www.iso.org/home.html
LINZ	http://www.linz.govt.nz
NOME	https://www.geointservices.io/index.html
OSMF	https://wiki.osmfoundation.org/wiki/Main_Page
QGIS/OSGeo Project	http://www.qgis.org/en/site/
TIGER	https://www.census.gov/geo/maps-data/data/tiger.html
SIGAME	http://www.sigame.es/SIGAME/
UN/WFP SDI-T	https://www.wfp.org/logistics/blog/first-test-version-sdi-t-geoportal-launched-wfp
USGS	https://www.usgs.gov/

ACRONYMS/ABBREVIATIONS USED

<u>Abbreviation</u>	<u>Meaning</u>
4 ESR	4 Engineering Support Regiment
AOI	Area Of Interest
API	Application Programming Interface
CAF	Canadian Armed Forces
CS	Citizen Science
.csv	Comma Separated Variable file format
DB	Database
DIN	The Canadian Forces Defence Internal Network
DTG	Date Time Group (i.e. Date of Incident: 10JAN17 T00:00:00)
DWG	Data Working Group
ESRI	Environmental Systems Research Incorporated
EXCITES	EXtreme CITizEn Science
FEMA	Federal Emergency Management Association (United States)
FME	Feature Manipulation Engine
GCS	Geographic Coordinate System
GeoGratis	Natural Resources Canada Free Data – GeoGratis Portal
GIS	Geographic Information System(s)
GoC	Government of Canada
GPS	Global Positioning System(s)
GST	A Canadian Armed Forces Geomatics Support Team
GUI	Graphical User Interface
highway_Line	OSM XML tagging feature most commonly associated with linear features such as roads and paths
HOT	Humanitarian OpenStreetMap Team
IDW	Inverse Distance Weighted data interpolation method
IMD	Index of Multiple Deprivation
IoT	Internet of Things

ISO	International Standards Organization
LINZ	Land Information New Zealand
MCE	The Canadian Forces Mapping and Charting Establishment
NBRN	New Brunswick Road Network dataset provided by Geo NB
NOME	National Geospatial Intelligence Services Open Mapping Enclave
.osh	OSM Planet History File
OSM	OpenStreetMap.org
.osm	OpenStreetMap XML file format
OSMF	The OpenStreetMap Foundation
.pbf	Protocolbuffer Binary Format file system
PCS	Projected Coordinate System
PM	Participatory Mapping
SIGAME	SI stema de G estión de Ayudas con Medios en Emergencias
TIGER	Topologically Integrated Geographic Encoding and Referencing
USGS	United States Geological Survey
VGI	Volunteered Geographic Information
WGS 84	World Geodetic System 1984
XAPI	Extensive Application Programming Interface
XAPI	XML DB Application Programming Interface
XML	Extensible Markup Language

Chapter 1: Introduction and Overview

1.1 Introduction

The phenomenon of Volunteered Geographic Information (VGI) as described by (Goodchild, 2007) is a rapidly expanding field in which geographic information and data are being generated by non-professionals, both uncompensated and unbranded entities, and being shared with the greater Geographic Information Systems (GIS) community. The generation of VGI is conducted by several types of contributors involved in multiple levels of collection involvement and is a low cost, if not free, solution for generating geographic information and data; however, the current state of the field of GIS development holds broad distrust in its accuracy (both spatially and thematically) as well as the reliability of the individual contributors who provide it. The current community of GIS professionals carries accreditation and formal training in the field, and as such, has garnered the respect of the information consumers as being the source for what would be considered “Trusted Information and Data”.

Several current initiatives such as the United States National Geospatial-Agency Gateway’s NSG Open Mapping Enclave ([NOME](#)) initiative or the United Nations/World Food Program Spatial Data for Infrastructure ([UN/WFP SDI-T](#)) are being conducted into enhancing the functional use of VGI for various humanitarian aid reasons. These initiatives look to realize and utilize the openly available and temporally current information and data which are collected by large groups of interested contributors in a variety of ways. The development and fundamental basis of these projects indicates that the international audience is seeking to use the concept of multiple source sensors for mapping as a viable resource in areas where little to no professionally generated mapping data exists.

One of the most successful and robust project datasets utilizing VGI comes from OpenStreetMap.org (OSM). This dataset design was first created and released by Steve Coast in 2004 (Ramm et al., 2011) and as of 2006 has been steered under the direction of the OpenStreetMap Foundation (OSMF), a United Kingdom-based international not-for-profit organization supporting, but not controlling, the OpenStreetMap Project. The OSMF is dedicated to encouraging the growth, development and distribution of open geospatial data for anyone to use and share. The information and data contained within the OSM dataset are the direct result of contributors who are encouraged (but not compensated) to provide local knowledge in order to ensure that the dataset maintains its temporal currency and spatial accuracy. Alongside a multitude of national level government agencies which have moved to provide open access to data, the dataset contributors themselves use portions of these open government data sources along with commercial grade Global Positioning System (GPS) receivers, satellite/aerial imagery (which is provided both gratuitously in times of need and openly when available), as well as simple paper maps, all combined with local knowledge to achieve the currency and accuracy goals. These types of information collection and sharing practices have led to questions such as how many contributors indicating the same information would make it reliable and/or credible without a professional accreditation; how do we properly define criteria which would question whether or not the information contributions are valid; how do the spatial (i.e. “where it is”) and thematic (i.e. “what it is”) contents of the professionally generated data coincide with the non-professional contributions provided by the crowd; are we as consumers of these data willing to have a trade-off between the completeness of semantic/source data for improvements in areas such as the temporal

currency of what is available; and what could be accepted as a minimum quality threshold for this trade off?

1.2 Thesis Context and Objectives

The primary focus of this thesis is the development of a process workflow to be used to maintain the temporal currency of standard datasets with a view to compare and analyze the spatial and thematic accuracy of OSM data with a professionally generated standard dataset assessing the reliability of the contributed information. Analysis will then be conducted to determine what levels of spatial and thematic variance are experienced between the OSM data and a professionally generated standard dataset in order to provide decision makers with empirical support on whether or not the OSM data can be used to replace, augment, or verify the existing information held by the professionally generated standard dataset.

This thesis reviews OSM contribution responses to natural disasters, primarily following earthquakes, to determine what requirements within the spatial and thematic accuracy domain should be considered when using OSM data in emergency management scenarios. The review of this event-related data is conducted due to its worldwide commonality, to gain an understanding of how we can rely on the contributions of the crowd following a natural disaster, and what factors might affect their overall input is sought.

The process to be developed is framed around a response to natural disasters involving a common natural event in Canada (flooding) and a simple/repeatable approach which can be used by the Canadian Armed Forces (CAF) in conjunction with other aid agencies as a guide for responding to a disaster relief situation.

There have been several different approaches to the communal information collection and sharing process such as the multi-organizational adoption of emergency management information systems. For example, Aedo et al. (2010) documented the all too common silo effect in which the participation of a variety of functionally independent agencies with their own developed processes of conducting operations can malfunction. These studies have clearly shown that multiple relief agencies operating in adjacent regions due to a collective emergency can cause a non-fluent communication flow, leading to a breakdown of organizational task responsibility and an overall diminished coordination of relief effort. From this, a consideration of what it is that would unite operational intent with multi-organizational harmony in times of need should be made. There have been several initiatives sparked with this consideration in mind such as the Semantic Web Ontologies created by Zhang et al. (2015), and the overarching multi-agency approach to aid relief provided by SIGAME developed by Aedo et al. (2010). These efforts have illustrated that through the Internet of Things (IoT), where computing devices which send and receive data are embedded in everyday objects, multiple contributors of information with regards to a specific crisis can be beneficial to the greater cause of care provision and life saving measures.

This thesis focuses on the use of an OSM dataset to provide reliable assistance to initial responders in an emergency management situation. The final outcome of this research aims to illustrate that VGI can be a source of viable and credible contributions to the “Map of the World,” given an understanding of the contributions and contributors behind it.

1.3 The OpenStreetMap.org Project

1.3.1 The Project Intent

The OpenStreetMap Project was originally founded by Steve Coast in the United Kingdom in 2004 (Ramm et al. 2011). It was originally intended to be a collaborative project to create a free on-line editable map of the world. The project itself was inspired by the success of Wikipedia's approach to collaborative knowledge sharing and how this could be applied to solving issues of restrictions on use or availability of map information across much of the world. Although the map itself is useful, the background data generated by the OpenStreetMap project are considered to be the project's primary output. As we will discuss further in this research, the quality of data input widely varies on the worldwide scale due to the heterogeneous density of contributions from region to region.

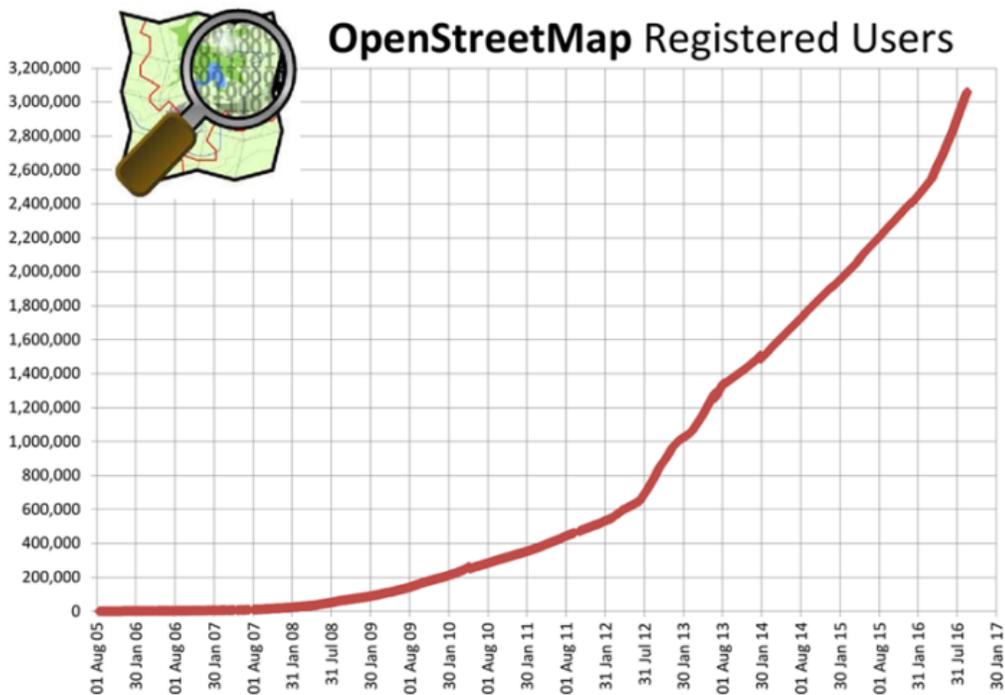


Figure 1.1 – Growth of OpenStreetMap users since its inception
<http://wiki.openstreetmap.org/wiki/Stats>

The exponential growth of registered contributors to the OSM dataset, shown in Figure 1.1, has caused much interest in the larger GIS community. The project itself is currently directed by the OSMF which is charged with ensuring the continuing growth of the initiative. The OSMF is composed of an international committee of members who elect a board of directors to steer the efforts of information/data collection as needed, and provide the legal and technical framework which ensures its continued usability and relevance. Fundraising for the projects future success is also handled by this foundation and the continued efforts have greatly lengthened the lifespan of the projects initially intended usability for research into the crowdsourced information it provides.

The project dataset itself began through contributions which were simple and small in scale. These included the basic outlines of the land features found around the globe and whatever the contributors could provide. The contributors would upload such things as GPS tracks and simple plotting of features derived from free imagery. Over time, and with the development of open source GIS software packages, the contributions themselves became more complex. The initial forms of contribution still remained; however, the contributor audience was now able to bulk upload data from open data sources such as the Government of Canada's data.gc.ca portal. These bulk uploads of data now account for a large majority of the information held by the current OSM dataset, with revisions constantly being conducted by the active users of the project.

1.3.2 Currently Understood Limitations and Constraints of OpenStreetMap Data

In order to properly conduct this research, an understanding of the limitations within the OSM dataset itself is required. When the structure for the dataset was first

being conceptualized, it was decided that any values contributed which were representative of elevation would not be included. This was due to the fact that commercial over the shelf global positioning systems (which the average contributor would potentially be using) were not properly calibrated to provide an accurate enough measurement of elevation. From this technologically dominant constant, the dataset framework was intentionally designed to not be third dimensionally (magnitude or Z) aware at any of the locations within the map dataset outside of attributional notations.

Another currently understood limitation of the dataset is that the data contributions are not strictly held to a proper set of standards, thus naming and categorical attribution conventions used in different regions make it difficult to ensure the data collected for analysis are really comparable. For example, “highways” in Canada are often referred to as “routes” in the United States. In order to combat this effect, the dataset instead uses a parent “highway_Line” attribute class which encompasses all forms of trafficable linear features such as trails, roads, paths, etc. This approach has its merits when dealing with regional differences; however it complicates analysis of the dataset and reduces thematic value.

1.3.3 The “Concept of Completeness” As It Pertains To Open Data Sources

In a workshop presentation on VGI, Steve Coast (2007) noted that “It is important to let go of the concept of completeness” (Zheng and Zheng, 2014). This thesis research was conducted under that premise. When considering the concept of what is a true “Gold Standard” or “Professional” dataset, there must be a benchmark for setting the encompassing factors which make it so. In the case of a dataset which pertains to spatial accuracy/correctness and thematic (attributional) accuracy, the argument for precision is

a closely fought battle. The regularly, and assumed periodically, collected data provided by a trusted source utilizes at least some form of standardization based on pre-established rules for data collection such as those set out in the International Organization for Standardization (ISO) frameworks. More specifically, agencies often look to ISO 19157:2013 which establishes the principles for describing the quality of geographic data and seeks to:

- Define components for describing data quality;
- Specify components and content structure of a register for data quality measures;
- Describe general procedures for evaluating the quality of geographic data, and
- Establish principles for reporting data quality.

As in many cases related to this research, ISO standardization is a widely used framework for the creation and utilization of data within industry. Unfortunately the proprietary nature for the provision of ISO accreditation services must be then viewed as a ready-made tool to assist in the streamlining of data collection and review processes, rather than the actual de facto standard for which operations should be conducted. As pointed out in the ISO 19157:2013 outline:

- *ISO 19157:2013 defines a set of data quality measures for use in evaluating and reporting data quality. It is applicable to data producers providing quality information to describe and assess how well a data set conforms to its product specification and to data users attempting to determine whether or not specific geographic data are of sufficient quality for their particular application.*
- *ISO 19157:2013 does not attempt to define minimum acceptable levels of quality for geographic data.*

Although the ISO does make an effort to define the criteria for the actual thematic completeness of a dataset, issues such as societal interests, the sharing of proprietary (and often private) data and security concerns can cripple the details available to completely describe the region for which the dataset focuses. Due to these considerations, when it comes to determining whether or not the so called standard dataset is the truth for comparison to the OSM dataset, analysis of the data available should always consider the initial need for data rather than the root specifications of the data available itself. The review of OSM data to be conducted in this research should indicate whether or not its spatial accuracy is comparable to what is held as the standard for many professionally mapped areas, and that thematically, it can contain many informational annotations not covered by its openly available trusted counterpart datasets.

The process development in this research seeks an initial understanding of a targeted Area of Interest (AOI), with a view that further information and data will become available and added to the dataset as operations are conducted. Upon reaching that stage, specific information gathering requirements can be spatially determined, depending on the spatio-temporal limitations of what is currently on hand. For the case of a natural disaster, the most current information available for the affected region would most likely be sought in order to determine our resource (aid) allocations. As a final stage in this process, if inclusion or sole use of the OSM dataset is to be considered for operations, a method for reviewing the contributors responsible for the data collection and information provision with regards to their overall interests should be retained.

As I will review below, much of the contributions to the OSM dataset fall into the 90:9:1 rule (Neilson, 2006) when it comes to social media type contribution behaviour

and the immediate response mapping needs following a crisis. This concept of social media usage indicates that in most online communities, 90% of users are individuals who never contribute, 9% of users contribute a little, and 1% of users account for almost all of the actual contributive action. From this, a basic understanding and acceptance of how VGI and data are provided to the OSM project can be established and an assessment made for its utility in supporting relief/recue needs in the context of a natural disaster.

As a final note on the concept of overall dataset spatial and thematic completeness, considering the original purpose for an information or data inquiry, the musings of Coast (2007) should be remembered in that practitioners need to forget about having a complete understanding of any AOI from the onset of operations, and look to whatever is readily available. This conceptualization will form the basis for sourcing what will be required in the data collection process once an aid operation has commenced. The material reviewed in this thesis has its founding in the desire to deploy mapping professionals on an operation with as much information as initially possible, with a view to continually update the dataset once operations are under way and new information becomes available.

For this study's review of OSM data, it was decided that instead of viewing the dataset in terms of "how close" it was to the "Trusted Standard" dataset, methods which seek to define "how much" information could be added to what was previously understood as the official record were developed. The synthesis of studies reviewed in this research are primarily based on developing a cohesive and common approach to providing aid to those in need in times of crisis by harmonizing best practices of the professionals in the field. Thoughtful planning for emergencies is a universally accepted

best practice; however, one should carefully consider the fact that a complete understanding of the information requirements and contributive activities to the OSM dataset for a region following a natural disaster will never be fully understood until such time as one has occurred.

1.4 The Research Question

This thesis develops methods to show that there are measures which can be taken to review the accuracies of the contributions within the OSM dataset. The VGI and open data which are contained within the dataset are scrutinized in order to determine any significant patterns of contributions, and an exploration of how they can be interpreted is conducted. The research developed a repeatable process workflow, which if automated, could be used as a basis for OSM data inclusion and use in the face of a natural disaster response. It is in this context that the research question is asked:

“What are the quantifiable characteristics that can be used to determine the spatial and thematic accuracy and the reliability of the contributors of VGI, in particular, OpenStreetMap.org (OSM) data?”

This question forms the culmination of the research focus which is then broken down into four sub-questions aimed at enriching the overall outcomes of this thesis.

1.4.1 What limitations exist within OSM which must be further understood?

Within the context of OSM and the greater VGI community, a movement towards a common practice of data collection is still within its growth phase. The OSM dataset does attempt to provide contributors with a suggested format for their inputs, but still lacks a formal policing of what exactly is uploaded to the dataset in order to maintain its truly open nature. The structure of control which the OSMF’s monitoring of the project conducts provides a gatekeeping aspect to the contributions provided; however, actual vetting of individual contributions by other contributors at the micro scale remains the

subject of which the group involved must be viewed as conducting their own iterative form of peer review. When considering this fact, it then becomes the data consumers' responsibility to clearly define how much the data provided can be viewed as trustable and reliable once an understanding of the contributions source has been determined.

Available tools for analyzing OSM data include a myriad of approaches yielding similar yet unique results. Many of the outputs rely on a translation of the OSM data structure (NO-SQL), composed of Nodes, Ways, Closed Ways, and Relations, to more commonly used generic Geographic Information Systems (GIS) vector data objects (composed of relational dataset objects (SQL) such as points, lines, and polygons). There are several drawbacks to this approach when we consider the in depth analysis of the contribution input itself. Several agencies such as Environmental Systems Research Incorporated (ESRI) and the OSGeo Project (QGIS) have developed their own methods of translating OSM data into a more ingestible standard GIS format. Exploration and understanding of the reasoning into the limitations of these translations will also be discussed further in this study.

The data contributed to the OSM project itself have no limitations on sources. The dataset is composed of a varied collection of inputs ranging from amateur geocache collectors to open but professionally generated data from governmental agencies such as the province of New Brunswick's GeoNB, or New Zealand's Land Information New Zealand (LINZ) agency. It is due to this fact that the nature of the sources found in the constantly evolving OSM dataset itself often cause debate as to whether or not it should be trusted as a data source. On one hand, the currency of the dataset is viewed as providing the most up to date information available for a specific spatial/thematic query;

however, the methods used to generate the information vary and as such call into question the validity of what has been provided. As a result of this consideration, further review of the effects of data collection and contributions over time have also been included in this study.

1.4.2 How do we quantify contributors and their contributions for use in the determination of a reliability index or measure?

The identities of the individual contributors to the OSM dataset are kept anonymous through the nature of the OSM project; therefore, a specific review of any one particular contributor cannot readily be accomplished through a single targeted approach. This process development seeks to determine factors related to the growth of the OSM dataset over time. By analyzing the history behind the creation of new versions of entities (also referred to as OSM changesets), it will be shown that contributor activities surrounding natural disasters can vary greatly both in the period leading up to an event as well as immediately following one.

1.4.3 What spatial and thematic variance thresholds must be established in order to determine the viability of applying the reliability index or measure before the OSM data are used?

When considering the differences, both spatially and thematically, between OSM data and a professionally-generated dataset, it is important to consider the driving factors which were/are in play in order to have data collected in the first place. For the case of OSM data usage following a natural disaster, what would be considered an acceptable average spatial difference between features found in a trusted dataset to those same features in OSM? This study seeks determine what underlying factors could be used to define what should be considered if OSM data is to be used as, or included in, an initial dataset in the case of an emergency management scenario.

1.4.4 Can OSM data be used to augment existing datasets once the spatial and thematic accuracy of the data as well as the reliability of its contributors has been ascertained?

In order to properly ascertain whether or not to employ the use of OSM data, it must be understood that any dataset will contain omission errors beyond our control. The professionally-generated datasets mentioned in sub-questions 1.4.1, are an amalgamation of sub-datasets merged together in order to yield a final product. This can be problematic when considering that items such as railway network data may have been collected some 30 years prior to current road network data, and collected under completely different standards, yet the date of publication for the dataset maintains that it is current as of its release date. Understanding this, it is important that decision makers in an emergency management scenario are provided with the proper background into a datasets composition before moving ahead with its usage. From this perspective, a review of the tools for analyzing the attributes associated with OSM dataset should be conducted to ascertain a best, based on expediency, approach to providing a thorough analysis of the information at hand. Once this background analysis has been conducted, a process for determining other factors such as spatial accuracy should then be conducted to determine approximately how much risk there may be associated with the usage of the data.

The synthesis of the knowledge gained through these sub-questions should provide a greater understanding of the quantifiable characteristics of the VGI, or more specifically OSM, data being reviewed and what factors should be considered in establishing the reliability of its contributors.

Chapter 2: A Review of the Literature on VGI and the Current State of OSM Data Analysis

2.1. The Current State of Voluntary Geographic Information Research

In today's growing field of Open or Crowd Sourced data, the trend towards the maintenance of temporal currency is an ever present concern. The data itself can provide almost instantaneous feedback as to events which are unfolding in different geographic regions. As multiple agencies such as the United States' Federal Emergency Management Agency (FEMA) and the Government of Canada (GoC) sub-units such as the CAF strive to ensure preparedness in the face of emergencies such as natural disasters both domestic and abroad, the costs of maintaining temporally current and complete datasets can often be untenable. This has led such agencies to look to the larger geoinformatics community to exploit information collected from non-professionals. A doctrinal approach to establishing policy in the field of VGI usage is already under development from several international perspectives, and studies have strived to show that information which is provided voluntarily can be trusted and reliable given an understanding of how the data itself was collected (Haklay 2015, Cipeluch et al. 2010, Zulfiqar 2008). In many cases, the mass availability of technology for use in the collection and sharing of geographic information and data has led to a paradigm shift in its employment and use. Several nations considered to be lacking in socio-economical and physical infrastructure development have been able to skip large portions of the internal growth and development challenges faced by other nations. Frameworks for the foundation of policies which could properly incorporate the reliable use of VGI as a primary data source in the future have been developed (Haklay 2015, Cipeluch et al. 2010, and Zulfiqar 2008). These frameworks seek to identify the different types of VGI

contributors, understand the scope of the contributions themselves, and finally determine the appropriateness for the use of the VGI available.

2.1.1 The Contributors

As a starting point, it is important to consider the very sources of the contributions of VGI. There are in fact several types of non-professional data collection activities which are currently being reviewed for their possible usage and inclusion in the generation of trusted geographic information. From several angles it can be seen that initiatives such as [NOME](#) or [SIGAME](#) are being developed with an aim to harness the contributions of multiple or crowd sourced data sources. In the case of [NOME](#), the concept of every soldier being a sensor provides greater support that the crowd can provide an increased situational awareness. It is in this regard that a definition of the contributors themselves is sought. The information and data provided are commonly defined under terms such as Volunteered Geographic Information (VGI), Citizen Science (CS), Participatory Mapping (PM), and Crowdsourcing (Lauriault and Mooney, 2014). These activities can be further broken down based on the individual types of contributors who provide the information and data (Coleman et al., 2009). The categorical range described by (Coleman et al., 2009) is outlined in Figure 2.1 and further defined below.

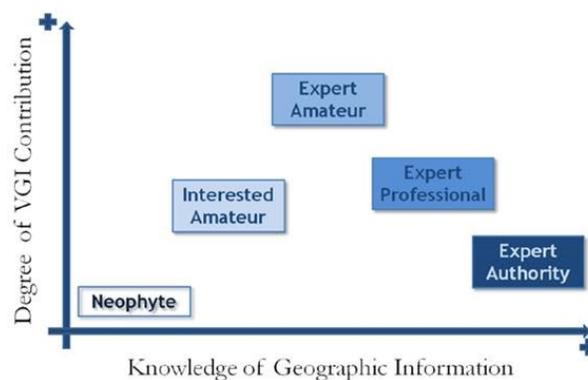


Figure 2.1 – *Volunteered Geographic Information: the Nature and Motivation of Producers* (Coleman et al., 2009)

Neophyte: is an individual without a formal background in a subject, but who possesses the interest, time, capability and willingness to offer an opinion;

Interested Amateur: has discovered an interest in a subject, has begun reading the background literature, consulted with other colleagues and experts about specific issues, is experimenting with its application, and who is gaining experience in appreciating the subject;

Expert Amateur: is a person who may know a lot about a subject, practices it passionately on occasion, but still does not rely on it for a living;

Expert Professional: someone who has studied and practices a subject, relies on that knowledge for a living, and may be sued if their products, opinions and/or recommendations are proven inadequate, incorrect or libellous; and

Expert Authority: a person who has widely studied and long practiced a subject to the point where he or she is recognized to possess an established record of providing high-quality products and services and/or well-informed opinions, and stands to lose that reputation and perhaps their livelihood if that credibility is lost even temporarily.

This breakdown of contributor types provides a basic framework for which a further distinction of the relevant differences between the types of professional and non-professional contributors of information or data can be made and further refinement of methods for viewing the reliability of the information and data they provide can be accomplished. The most important point to note here is that the boundary between professional and non-professional data collection and analysis is becoming less clearly defined and should be seen as a varied range between highly unreliable to completely reliable contribution provision.

The aim of this review will be to further discuss the processes currently in practice for the collection of open information and data, and how these contributions are being measured for various attributes such as spatial or thematic accuracy and overall reliability. The ongoing efforts being made to encourage the use of the greater community of VGI in organisations which are traditionally strictly in the professional framework of study will also be reviewed. Finally, an understanding of the possible ways forward with regards to the use of this type of information and data for the purposes of emergency responses.

2.1.2. The Division of Levels in Participation for the Collection and Analysis of VGI

Along with the different classes of contributors outlined earlier, it is also important to further understand the categories currently accepted in the field of VGI study for the types of activities employed in the actual collection of VGI. Haklay, (2013) developed a typology consisting of four separate levels which takes into consideration the formal scientific knowledge, engagement and cognitive input of collection participants as shown in Figure 2.2.

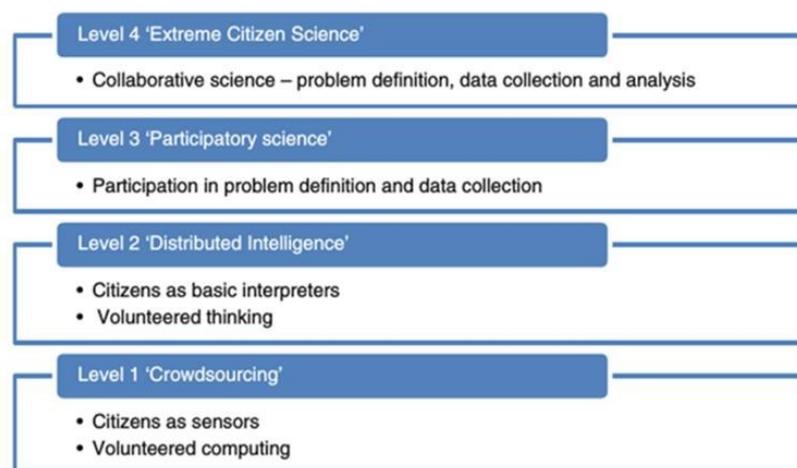


Figure 2.2 – *Levels of Participation and Engagement in Citizen Science Projects* (Haklay, 2013)

Level 1 - Crowdsourcing: participation is limited to the provision of resources, and overall cognitive engagement is minimal. The initial concept of utilizing the individuals in the greater community, which are non-professional in nature, as sensors for data collection and basic information processing, arises at this level. In participatory sensing, or citizens as sensors approaches to data collection, participants are asked to carry sensors around, collect data and report/feed these back to an experiment organiser. From a scientific perspective the characteristics of the instrumentation are known (e.g. the accuracy of a specific GPS receiver, data collection intervals, etc.) and the experiment is controlled to some extent as assumptions about the quality of the information can be and are made.

When a review is made of the first level “Crowdsourcing” typology described by Haklay, (2013), there are several widely known examples of this activity. In the case of citizens as sensors, companies such as [Garmin©](#), [Magellan©](#), and [TomTom©](#) allow users of their sensory devices to share georeferenced data collected through activities such as cycling or geocaching. The collected data are often then shared through various projects and initiatives such as OSM or OpenCycleMap.org. This example requires minimal cognitive engagement from the individual (sensors) and allows for the generation of non-professional data which can be seen as an additive to an existing dataset as it becomes available.

In reviewing the activity of voluntary computing, there have been projects such as U.C. Berkeley’s *SETI@home* (<http://setiathome.berkeley.edu>) initiatives in which users are provided with analytical software to run on their computers at home which analyzes data obtained from large scale sensor networks available through the Internet. There are

numerous other examples of crowdsourcing initiatives which can be found; however, the example provided here is an initiative which has been maintained for over a decade of proven use within the non-professional community. This heritage provides potential reliability and credibility of the information which should not be overlooked when assessing the viability of VGI use.

Level 2 - Distributed Intelligence: here the cognitive ability of the participants is considered a resource. Participants are offered some basic training and collect data or carry out one or more simple interpretation types of activities such as the filling out of surveys or checklists. Quality assessments are made by the project/experiment designers and continued attention is given to volunteer contributors to facilitate ongoing learning.

For this second level of participation, “Distributed Intelligence”, basic training in the collection of data is provided to the greater audience interested in the findings. Simple data collection procedures are developed by a professional entity and the interested amateur is engaged in the collection of data through several differing forms ranging from a paper checklist to record observations, to the conscious use of GPS during such activities as geocaching. The *Coastwatch Survey* (<http://coastwatch.org/europe/survey/>) involves the entire spectrum of individual contributors from Neophyte to Expert Authority in a project designed to monitor European coasts in an effort to develop a database which can then be used as a platform for environmental policy change and development by government (Lauriault and Mooney, 2014). Due to the fact that this data collection process is designed by a professional entity, collection is performed by the general public, and the data collected is reviewed by the professional entity prior to promulgation, it would be considered Distributed Intelligence. It is therefore clear to

observe that this is the level in which a formal version of quality control in VGI generation and use begins to emerge. As research in this area continues to expand, a push towards online interoperability is witnessed in examples such as the framework designed by Zhang et al. (2015) for use in disaster response. In the field of GIS, professional entities are developing online base geographic models which can be updated rapidly by non-professional users if observations of extreme events (i.e. natural disasters) are observed. This is done in order to aid in speeding up disaster response both for responders and the general public alike. These on-line response systems are commonly tailored for very specific geographic regions; however, the use of computer programmed gateways being used indicates that efforts to quality control the flow of VGI being ingested are in the works, in order to enhance the reliability of the information or data provided.

Level 3 - Community Science: in this case participants help develop the problem to be investigated and in consultation with scientists and experts, the collection methods are designed. Experts provide analytical expertise and interpret the results. Volunteers may become experts in the data collection and analysis through their engagement and may suggest new research questions which can be explored by the data they have collected.

When analyzing the participatory level of Community Science, it is important to understand that it is at this level where the contributor also becomes one of the primary consumers of the information generated. At this level of participation, it is often seen that community groups have observed a particular area of interest and wish for it to be understood and exploited. It is also important to note that at this level, it is the community which defines the problem and not the professional (scientist, agency, etc.);

wherein, most of the initiatives are primarily concerned with societal changes (i.e. monetary dispersal, provision of services, etc.). To serve as an example of this, Community Science (www.communityscience.com) employs a collaborative and participatory approach to problem solving through the use of targeted data collection from communities which have the self-identified need to quantifiably approach qualitative problems and serves as a marker for this level of VGI. The problems approached through this type of VGI collection are topics such as the awarding of funds for infrastructure development projects based on communal (want vs. need) requirements to empower and involve the contributors to ensure that the information provided is not only current but correct in nature.

Level 4 - Collaborative Science or EXtreme CITizEn Science (EXCITES): this is an integrated activity where professional and non-professional, ideally expert amateur types of contributors, group together in order to prioritize and decide which scientific problems to work on. Collaborative decisions are also made on the nature of the data to be collected to ensure there are valid results being sought and that they adhere to scientific protocols while also matching the motivations and interests of participants. The participants may also be involved in the analysis and publication or utilisation of the results. In this domain, scientists act as facilitators as well as experts.

This method of collection and analysis of VGI shows the most promise for enabling trustable and reliable sources of VGI. In this case, one of the clearest outlines given in order to define citizen science is *“the scientific activities in which non-professional scientists volunteer to participate in data collection, analysis and dissemination of a scientific project.”* (Cohn, 2008; Silvertown, 2009). The professional community works

hand-in-hand with the broader community to not only develop the studies to be conducted but also attempting to ensure that the research and its results are beneficial to both the professional and non-professional communities. An excellent example of this approach would be the Rinkwatch (www.rinkwatch.org) initiative brought forth by the geographers at Wilfred Laurier University (see Figure 2.3). This initiative allows non-professionals to access an interface which communally maps skating rinks in an area and allows users to simply update the site with whether or not the rink is usable on a day-to-day basis. This interface not only allows users to also see what other rinks in their area are useable, but also provides temporal temperature data to the geographers who formed the site for use in climate change mapping and modelling.

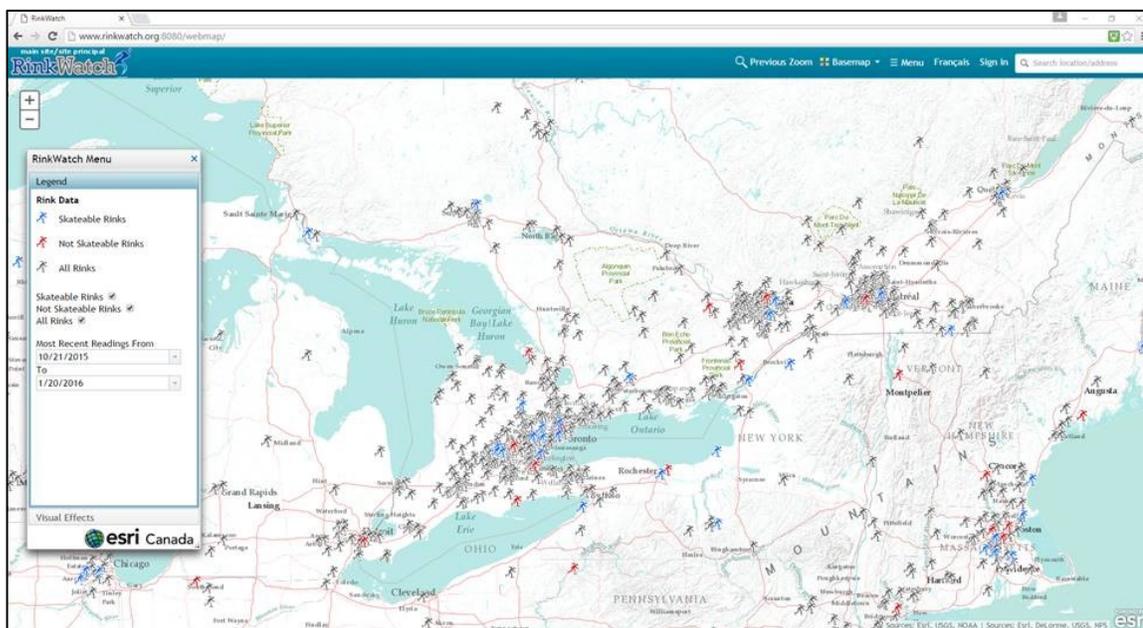


Figure 2.3 – Rinkwatch GUI (www.rinkwatch.org)

Much of the current research into the trusted use of VGI is targeted at the EXCITES level. The “Doing It Together” (DIT) initiative serves as an example of how EXCITES is currently expanding. The initiative itself is geared around empowering citizens to

explore, measure and experiment with the world around them in efforts to address the challenges faced when looking to a sustainable future. This initiative is currently being run by the pan-European network, European Citizen Science Association in conjunction with universities, science galleries, museums and arts organizations, and NGOs. The initiative is a four year (2016 – 2019) commitment to increase public participation in scientific research and innovation across Europe through the use of their travelling “Science Bus”. This project has even gone so far as to provide Ten Principles of Citizen Science to outline its “Sharing best practice and building capacity”. The fact that this initiative has garnered as much support as it has, serves as clear indicator that research in the field of VGI generation and collection through EXCITES will continue to grow over time.

When similar projects which involve information sharing vice crowdsourcing are reviewed, it is seen that they do share similar interface design challenges such as those described by the work of Aedo et al. (2010). This study reviewed the response to large-scale emergencies as a cooperative process that requires the active and coordinated participation of a variety of functionally independent agencies operating in adjacent regions. A review of IT-mediated cooperation among Spanish response agencies was conducted concerning the challenges which were faced in the adoption of the SIGAME interface, information sharing, communication flows, and coordination among agencies that do not share a unity of command. The study empirically detailed how participating/collaborating agencies tend towards working within their individual solutions to problems rather than readily accepting input from external forums or

agencies. Overall this individualism leads to a silo effect in which inefficiencies or lost opportunities occur more frequently.

From the current state of projects in this field it is clear to note that the evolution of VGI generation and collection practices, and its trusted use, will only gain credibility as the approaches reach their maximum efficiencies. These new methods of data or information collection and collaboration are being iteratively developed and will eventually be fully established. Continued monitoring of these initiatives should reveal trends which verify the effectiveness of the collection practices, and serve to exemplify that the VGI collected is in fact of a reliable and accurate nature.

2.1.3. Current Approaches for Validating Data and Determining Reliability

Once we have passed the initial gateway of considering the different methods in which VGI is generated, we must now consider the methodologies which could be employed to validate the data and determine the reliability of the sources before they are promulgated as a trusted output product. The current research into this suggests the focus should be on the contributors themselves and how to properly quantify the contributions. There have been calls for the creation of a registry in which contributors and their contributions are logged by expanding on the tagging process currently employed by the OSM project, however the contradiction between maintaining the anonymity of the contributors and the ability to label the information as unbiased, has led to this tracking and verification method not yet being largely explored (Coleman et al., 2009).

Goodchild and Li (2012) proposed three possible alternative approaches to VGI validation. The three approaches offered in their study involved non-privacy invasive methods for validating and ensuring VGI reliability:

2.1.3.1. The Crowd-Sourcing Approach: Crowd-sourcing in this sense refers to the ability of a group to validate and correct the errors that an individual might make.

This ideally would take the form of Linus's Law, named by (Raymond, 1999) in honor of Linus Torvalds, the originator of Linux, which states that "*given enough eyes, all bugs are shallow*"; in the context of software engineering, the bugs in a piece of software are most likely to be found and corrected if a large number of software engineers have an opportunity to review the software. As the original meaning of the quote was given by Raymond (1999) a tangential view must be taken in that by accepting the fact that with multiple contributors, as in the case of VGI, we could adopt a surrogate view of the information provided as being peer reviewed by multiple reviewers. Haklay et al. (2010) adopted this approach and quantitatively showed that in the case of VGI, or more specifically the OpenStreetMap.org project, no statistically significantly further improvement in spatial accuracy within a region is seen above 13 contributors. With this in mind it can be surmised that review is being conducted and corrections to the OSM dataset are being made on an ongoing basis. This approach is however noted as having its own set of flaws, in that it is reliant on a significant number of contributors being available, and that the socio-economic conditions of the region under study (e.g. access to the proper tools to collect and contribute geographic information) were not considered as variables. The effects of this will be examined later in this study and an attempt to review the background of the contributions being made in response to a natural disaster will be conducted.

2.1.3.2. The Social Approach: This approach is defined by its hierarchy of trusted individuals who act as moderators or gate-keepers. As mentioned previously, tracking

contributions by individuals allows for the calculation of metrics of commitment and reliability and could provide a quantitative basis for determining a level of trust. As an example of this, within the OSM project there are two tiers in the hierarchy of contributors: ordinary users, and the Data Working Group (DWG). The DWG is currently composed of eight members, who deal with vandalism, copyright violation, disputes, geographic locking, etc. Every registered user of the OSM project can add and edit the geographic features contained in the dataset; however, if there is any dispute, the DWG decides the solution and thus serves as both a moderator and a gate-keeper. In this case, if VGI provision is viewed in comparison to that of information generated by a professionally credentialed mapping agency, an understanding that the OSM project also provides hierarchies of trust based provision of information is evident. These are observed in the form of review and oversight to ensure the credibility and reliability of the information and data provided. Thus, by maintaining a proper control via moderator who is a subject matter gate-keeper, the VGI sources for the OSM project are in fact openly displaying their level of reliability.

2.1.3.3. The Geographic Approach: In this approach a comparison of a purported geographic fact with the broad body of geographic knowledge conducted, or otherwise stated, how it is possible to know whether a purported geographic fact is true or false. Consider the first law of geography “All things are related, but nearby things are more related than distant things” (Tobler, 1970), and taking into account the accuracy goals of VGI reporting within the OSM project, the argument is strengthened that if several local sources close to an event are reporting the same or similar facts about a specific

area, it is more likely that the information is reliable than several non-local sources reporting the same facts about dispersed and heterogeneous areas.

2.2. New Information Collection and Use Concepts

In the development of VGI-incorporated cartography, there have been several examples brought forward to help bridge the currently existing trust gap. As with any emerging capability, the professionally accredited founders of the field are often reticent to accept changes to the current paradigm in which they operate. In many cases, as seen within the information sharing studies by Buscher and Wood (2013) and Aedo et al. (2010), the established methods for performing tasks by an organization internally are extremely robust. Iterative approaches to constant improvement tend to be used over significant amounts of time, and the idea of “why should we re-invent the wheel?” comes into play. In the field of GIS application, a long-standing common practice has been to only rely on data collected by trusted (professionally generated) sources. These sources usually carry an accreditation of some sort which elevates them to the level of professionals who receive compensation for their goods, like data, and provision of services, like the collection of data. This paradigm causes a rift when considering that VGI could be used to augment or replace the accepted data sources and the culture of distrust increases. As time passes, VGI research continues to expand, and the developers must remain cognizant of the currently accepted sources of data and the standards gateways which must be passed before its use will be more readily accepted.

2.2.1. Quantifying Collection

An easily quantifiable aspect of VGI (specifically GIS or OSM) quality which has arisen is the reliability of the sensing apparatuses of commercially available GPS receivers or smart phones. Traditionally, much of the geographic information collection was performed by professional agencies through the use of land based terrestrial survey

methods, or even further back historically, through the use of input such as surveys or monitoring tables filled out by trusted members of communities. With the advent of satellite technologies and cellular tower emplacements, cartography has moved further and further towards the use of remote sensing applications and satellite based terrestrial survey. The apparatuses used for this have differing levels of accuracy, and as with most scientific instruments, the availability of them to the masses is limited to the consumer market which can afford them. If one chooses to trust professionally generated data over that provided through VGI (OSM), an understanding of whether or not the professionally generated data was collected using a terrestrial, land-based observation survey, or collected through the use of a professional grade GPS sensor not available to the general public should be made. In the case of the land-based data collection, a higher level of accuracy and therefore reliability would most likely be observed; however, if we consider the functional characteristics of the GPS itself, which relies on a moving object (satellite) triangulating the position of another object while the surface on which the target object is located is also moving, the complexity of this system alone calls into question its accuracy. Considering VGI and its generation using common commercial off-the-shelf grade, dedicated GPS receivers (such as those provided by GARMIN, TOM TOM, etc.) and their well-documented and known positioning errors, the fact arises that in some cases, both forms of data collection (professional and VGI) rely on the same basic premise and as such should be treated similarly.

Many of the VGI collection systems discussed in recent literature suggest the use of other forms of positioning systems whose functioning mechanisms are commonly misunderstood. The most notable of these would be contributors using smartphones to

capture and share geo-referenced information and data. These smartphone based positioning systems actually use the position of land based transmission towers (not satellites) to triangulate a user's geographic position. It is in this way that the inaccuracy which could be caused by a moving satellite positioning function is removed, and allows the geographic positioning of the smartphone to conceivably attain a higher level of spatial accuracy. The collection and sharing of geographic information using smartphones can be readily seen in examples such as the British Trust of Ornithology's BirdTrack (www.bto.org). This initiative utilizes geo-referenced images to capture bird species in order to trace their emigrational patterns and the contributions to the dataset have been seen to span continents. With local VGI projects continuing to emerge, examples such as the iSpot (www.ispotnature.org) initiative have also taken a similar approach. Contributors are encouraged to submit images of the flora and fauna around them in return for descriptions of what species they are. These data are then collected and mapped by the iSpot team in order to develop an understanding of the geographical extent covered by the different species.

2.2.2. Qualification of VGI Use in a Quantitative Way

When attempting to quantitatively verify the information which is being contributed, it is apparent that by having different types of contributors, which are contributing several different standards of contribution type, confusion would arise by way of disorganisation. Antoniou et al. (2010) viewed this as a matter which could be more correctly dealt with by streamlining the requirements of the contribution itself. The current XML file schema utilized by the OSM project is functional, but does not closely adhere to what would be considered a historically standard GIS protocol. To remedy this, Antoniou et al. (2010)

suggest modifications of the existing file system into layout methodologies which would bring the data being collected up to a more commonly accepted International Standard Organization (ISO) documented standard. Although efforts have been made to do this, further development into this approach of data intake management would be beneficial for the development of a long standing ability to utilize and maintain the integrity of VGI datasets as they evolve.

Several studies have been carried out to determine the positional variance associated with VGI in comparison to various standard datasets. Girres and Touya (2010) provided a comparable analysis of French OSM data with regards to the number of contributory transactions the service received and, Cipeluch et al. (2010) assessed the reliability of Ireland OSM data compared to Google Maps and Bing Maps to determine the associated accuracies and spatial offsets. In an attempt to quantify the number of VGI contributions and contributors required to validate GIS data in an area, Haklay et al. (2010) further reviewed the concept of Linus' Law in application to 1 km² sample areas in the London U.K. developed area. In their analysis, a model to analyze VGI (OSM) positional variance from an accepted standard, in this case the U.K. Ordnance Survey Meridian 2 dataset was developed. They utilized the process developed by Zulfiqar (2008), in which spatial accuracy in sample areas or grid squares, in the vicinity of London, England was used as the basis for their review. This study highlighted the percentage overlap (as seen in Figure 2.4) between the two vector datasets when realistic buffers, set around the professionally generated dataset, were included to account for the relative flow of traffic through an area and the measurement of spatial accuracy from the contributions used in the collection of the two different data types differ. They found a

high level of spatial accuracy, and a statistically significant (80% or higher) average overlap was observed between the datasets at relatively low buffer limits. Using this information, they were then able to determine the actual number of contributors in the sample grid squares, as seen in Figure 2.5, and quantitatively show that the sample areas with the decreased spatial offset between the features of the two datasets was directly linked to an increased number of unique data contributors in the sample area.

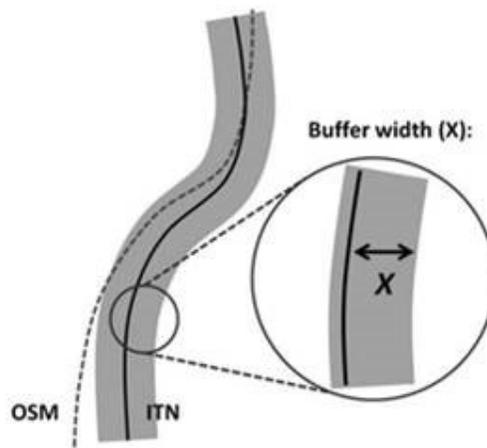


Figure 2.4 – An example of the buffer method used by Haklay et al. (2010)

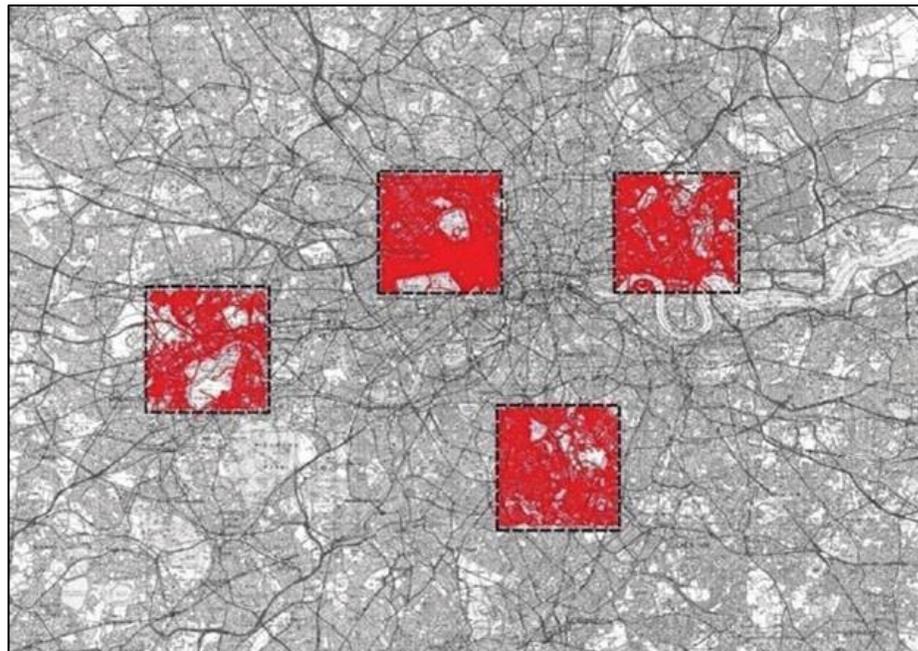


Figure 2.5 – ITN of London, England with data sample areas utilized for the validation of OSM data (Haklay et al., 2010)

The authors do admit that this relationship was not without fault in that the sample areas which were under study contained a large amount of social and economic heterogeneity which could possibly have led to variability in data quality. Haklay (2008) presented different study areas, with varying Indices of Deprivation, such as varied levels of income, described in the Oxford University's Index of Multiple Deprivation (IMD), which potentially affected the root OSM data collected. Although these variables were identified, further study which accounted for average social and economic status in an area was not conducted. As the authors admitted, the consideration of the additional factors covered in the IMD could provide a broader understanding of limitations regarding contributor densities in the areas studied with a view to confirm the Linus' Law correlation of contribution efforts which were observed in their work.

2.2.3. The Determination of VGI Use Requirements

If the type of open data analysis conducted in Haklay (2010) is considered, and its potential for factors such as those found in the IMD to cause bias towards VGI not being treated as reliable and trusted, it is apparent that there is a need to clearly define the scope of VGI involvement within a project or operation. This definition should be viewed as either direct or indirect on its influence on the output products. Haklay et al. was based around a direct role in which a product built solely on VGI was reviewed with regards to the number of contributors which were actively involved in updating and correcting the contributions for a given region. If a consideration is made of the possible indirect roles for the use of VGI in GIS and Digital Cartography production, a different set of questions and potential uses arises. In this case VGI could be treated as the standard for comparison and verification rather than the sole input itself. See et al. (2015) worked on

the problem of overlapping several standard datasets, used for determining land cover types at different spatial resolutions and applying heuristics or geographically weighted regressions to the maps in order to compare them at the pixel level. They identified areas where all of the standard datasets disagreed significantly. Following this the sampling scheme of the global validation dataset created by Zhao et al. (2014) for collecting data via Geo-Wiki was employed and lead to a crowdsourced dataset being generated. Finally, a quality assured subset of data was used to determine the actual groundcover type and determine which of the datasets (standard vs. crowdsourced) was more accurate. Although this methodology would be useful in theory, the datasets used were research datasets, and assumed that the datasets were free of bias (ownership or other interests) which could potentially cause difficulties if this process were to be applied commercially using datasets collected by competing professional agencies. This research did, however, provide a solid example of how small pockets of VGI could be used to properly augment an existing professionally generated dataset

Barron et al. (2014) provide a comprehensive framework for analyzing the OSM dataset with regards to the quality of spatial data as outlined in ISO 19113. The framework proposes spatial data quality can be evaluated using the following elements of ISO 19113:

- “*Completeness*”: describes how complete a dataset is. A surplus of data is referred to as “Error of Commission”, a lack of data in contrast is “Error of Omission”.
- “*Logical Consistency*”: declares the accuracy of the relations manifested within a dataset. This element can be further subdivided into “intra-theme consistency” and “inter-theme consistency”.

- “*Positional Accuracy*”: defines the relative and absolute accuracy of coordinate values.
- “*Temporal Accuracy*”: the historical evolution of the dataset.
- “*Thematic Accuracy*”: describes the accuracy of the attributes assigned to the geometric elements.

Looking to apply the framework, Barron et al. (2014) point out that it must be remembered that OSM data quality is heavily dependent on the purpose for which it is to be used, and refer to this as a “Fitness for Purpose” assessment, previously defined by Veregin (1999). The analysis framework itself is based around a tool known as iOSMAnalyzer which is used to review the history of OSM data for specific regions based on 25 intrinsic quality indicators as outlined in in Figure 2.6.

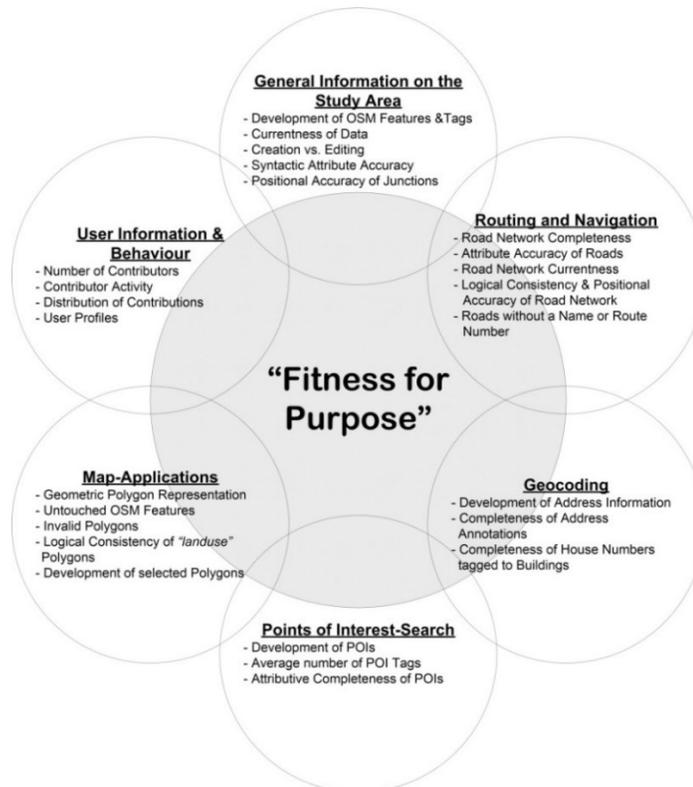


Figure 2.6 – The 25 Intrinsic Quality Indicators utilized by iOSMAnalyzer (Barron et al., 2014)

As the framework indicates, by reviewing the history of OSM data for a specific region, an analysis of the 25 growth characteristics outlined over time can be used to determine the region's data quality with regards to the previously discussed ISO 19113 elements. In their work, this framework was employed to assess the amount of road network contributed, the dataset's positional accuracy, house number attribute inclusion and the overall geometric representation of natural polygons. Their results were consistent with several other studies (Girres and Touya, 2010; Haklay et al., 2010; Neis and Zipf, 2012) and supported the hypothesis that high numbers of contributors within a specific region will lead to a stable high quality OSM dataset which is more likely to be kept up to date.

2.3 Use of OSM in Natural Disaster (Earthquake) Scenarios

In order to form a better understanding of the reactions of the crowd when a natural disaster occurs, analysis of natural disasters in different regions around the globe should be conducted. Considering the destructive nature of earthquakes and how they affect very large areas all around the world, an understanding of the contributions to the OSM dataset following such events serves to illustrate the potential for the data's utility in relief operations. Neis et al. (2010) studied the disastrous earthquake in Haiti in 2010 and documented a surge in OSM contributions for the region which ultimately lead to this being one of the most successful OSM projects to date. Figure 2.7 shows how the OSM dataset for the region surrounding the capital of Haiti grew extensively almost overnight and is now one of the most well mapped regions of the world in OSM.

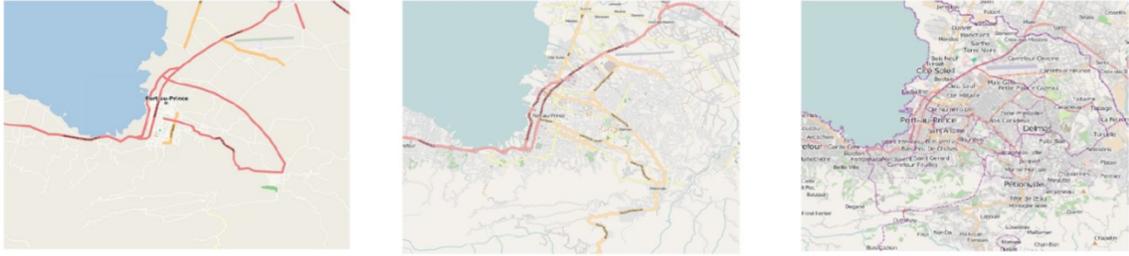


Figure 2.7 – Maps from OSM data showing the Haiti capital of Port-Au-Prince before the earthquake, two days after earthquake and how it looked 1 year later. (Neis et al., 2010)

Neis et al. (2010) heavily credit the Humanitarian OSM Team (HOT) and their introduction of new OSM tags which were incorporated to provide additional attribute information to the OSM dataset related specifically to disasters. This additional tagging schema allowed those mapping the region to provide information on potential obstacles and areas to avoid due to earthquake related damage. This schema was developed rapidly and was based on the United Nations Spatial Data Infrastructure (Transport-Related) or UN SDI-T by OSM community members familiar with the field. It should also be noted that social media and web 2.0 technologies were utilized to spread the word regarding the earthquake and resulted in proprietary entities such as the Satellite Imaging Corporation's GeoEye and DigitalGlobe's high resolution satellite imagery being made openly available and free of charge to the OSM community. Because of this, OSM contributors were able to use the available imagery to rapidly update the OSM dataset, as can be seen in Figures 1.2 and 2.8.

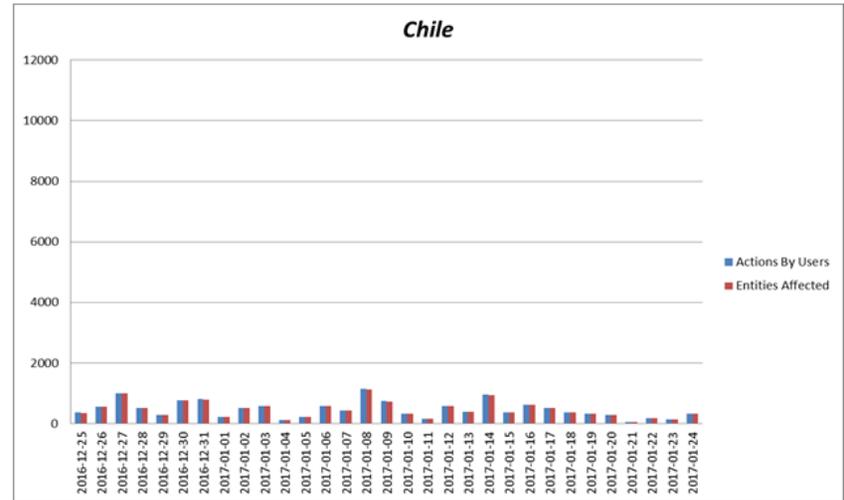
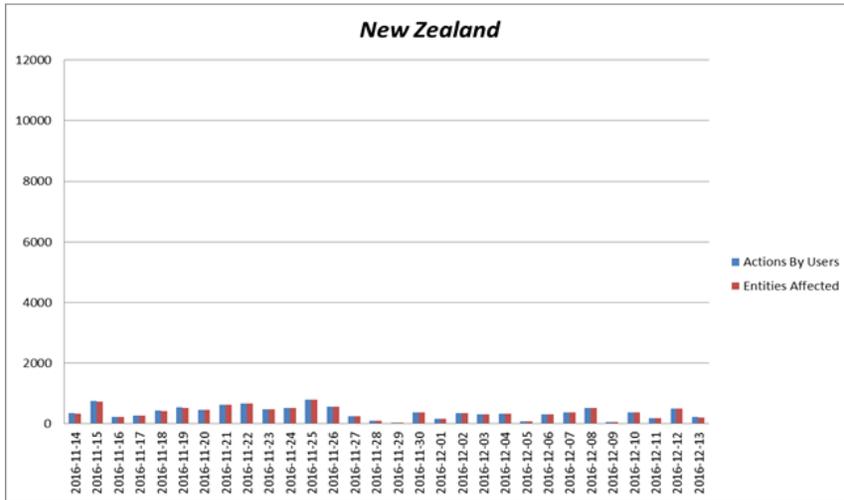
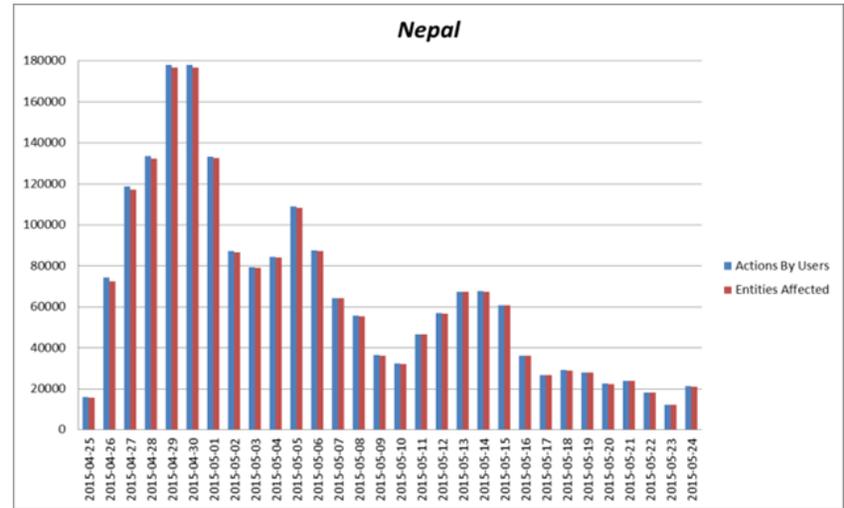
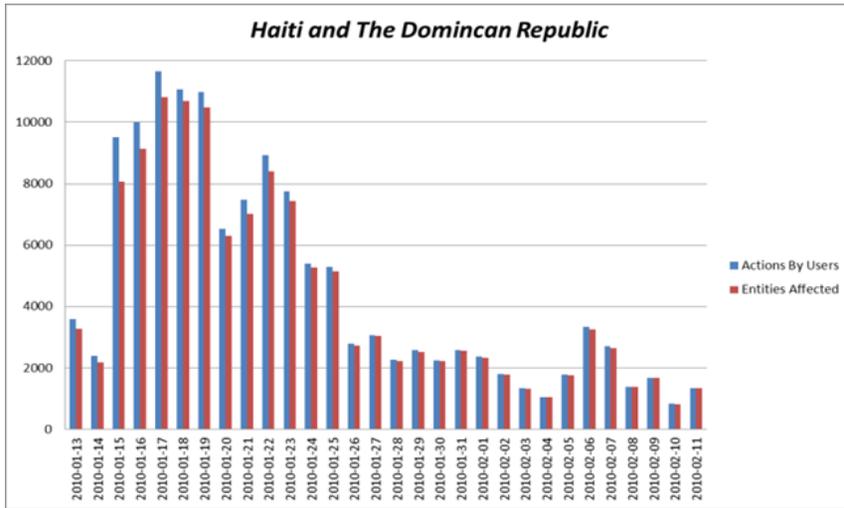


Figure 2.8 – Comparison of OSM contribution volumes in the first month following a severe earthquake in Developed and Developing Infrastructure countries

Table 2-1 – *Regions of interest which were struck by a high magnitude earthquake event*

Country	Date	Epicenter ¹	Magnitude ²	Total Number of OSM Contributions in the Month Following the Earthquake
Haiti	12 JAN 2010 1653H	Lat 18.457 Lon -72.533	7.0	135,711
Nepal	25 APR 2015 1156H	Lat 28.147 Lon 84.708	7.8	1,986,346
New Zealand	14 NOV 2016 0002H	Lat -42.757 Lon 173.077	7.8	11,275
Chile	25 DEC 2016 1422H	Lat -43.416 Lon -77.880	7.6	14,567

Due to the occurrence of a surge of contributions to the OSM dataset following a worldwide recognized natural disaster, analysis should be conducted to determine if there are in fact similarities to this response following other highly destructive, high magnitude earthquakes. All of the events in Table 2-1 were similar in their natural destructive power; however, by analyzing the response of the OSM community, a dramatic difference in terms of how actively involved they were following the incidents is seen in Figure 2.8. Much of this could be credited to the efforts of HOT in the time period immediately following the disasters, but there are also other factors which must be considered. A significant difference in OSM dataset contribution efforts following an earthquake can be seen. In the cases of Haiti and Nepal contributions spiked in volume shortly after the events occurred. This could be possibly attributed to the work of the HOT; however, it cannot be overlooked that socioeconomic factors in these regions could have also contributed to the reactions of the OSM community (Haklay, 2008; Haklay et al., 2010).

¹ Coordinates provided are in decimal degrees based on the standard OSM WGS 84 GCS

² Magnitude is based on the Moment Magnitude (M_w) scale provided by the USGS

Regions such as Haiti and Nepal are generally viewed as developing countries based on their individual political and economic infrastructure development, whereas the incidents in New Zealand and Chile can be used as contrasting examples where socioeconomic development has led to the provision of open data (through their open government data portals LINZ and Datos.gob.cl respectively). As discussed earlier, the OSM project has evolved to include any and all available open data sources to augment its data stores. The availability of these services could be viewed as a possible link to the actual need, if any, for additional OSM contribution responses in regions which would be currently seen as fully developed in terms of open data availability. It is clear that there are different magnitudes of responses within the OSM community to different events which should be evaluated based on their historical precedence. As described in Neis et al. (2010), significant amounts of information (i.e. contributions) can be generated for a region in a relatively short time given the proper circumstances and support of the crowdsourced mapping community. As a whole, this trend towards community assistance of any form to those in need should not be overlooked.

2.4 Related Works and Proposed Usage of VGI/OSM Data for This Study and Beyond

In the material related to VGI usage, there are several potentially viable avenues for its inclusion in professionally generated datasets. This thesis seeks to assess the potential to utilize VGI (in the form of OSM) for the purposes of assisting in natural disaster aid provision. Three options are considered:

1. The OSM dataset could be utilized as a complete substitute in the face of a relatively recent, professionally generated dataset being unavailable for a given region (i.e. the case of the Haiti earthquake in 2010 as reviewed in Neis et al. (2010));

2. A hybrid dataset could be generated by merging the OSM dataset with a professionally generated dataset through some form of data manipulation, or
3. The OSM dataset could be used to prioritize professional data collection needs based on the temporal currency of data in a specific Area Of Interest (AOI).

The research in this study concentrates on a workflow to accomplish option 3; however, the tools which were reviewed could be used in any of the three options for determining the spatial accuracy of OSM data when compared to a professionally generated dataset.

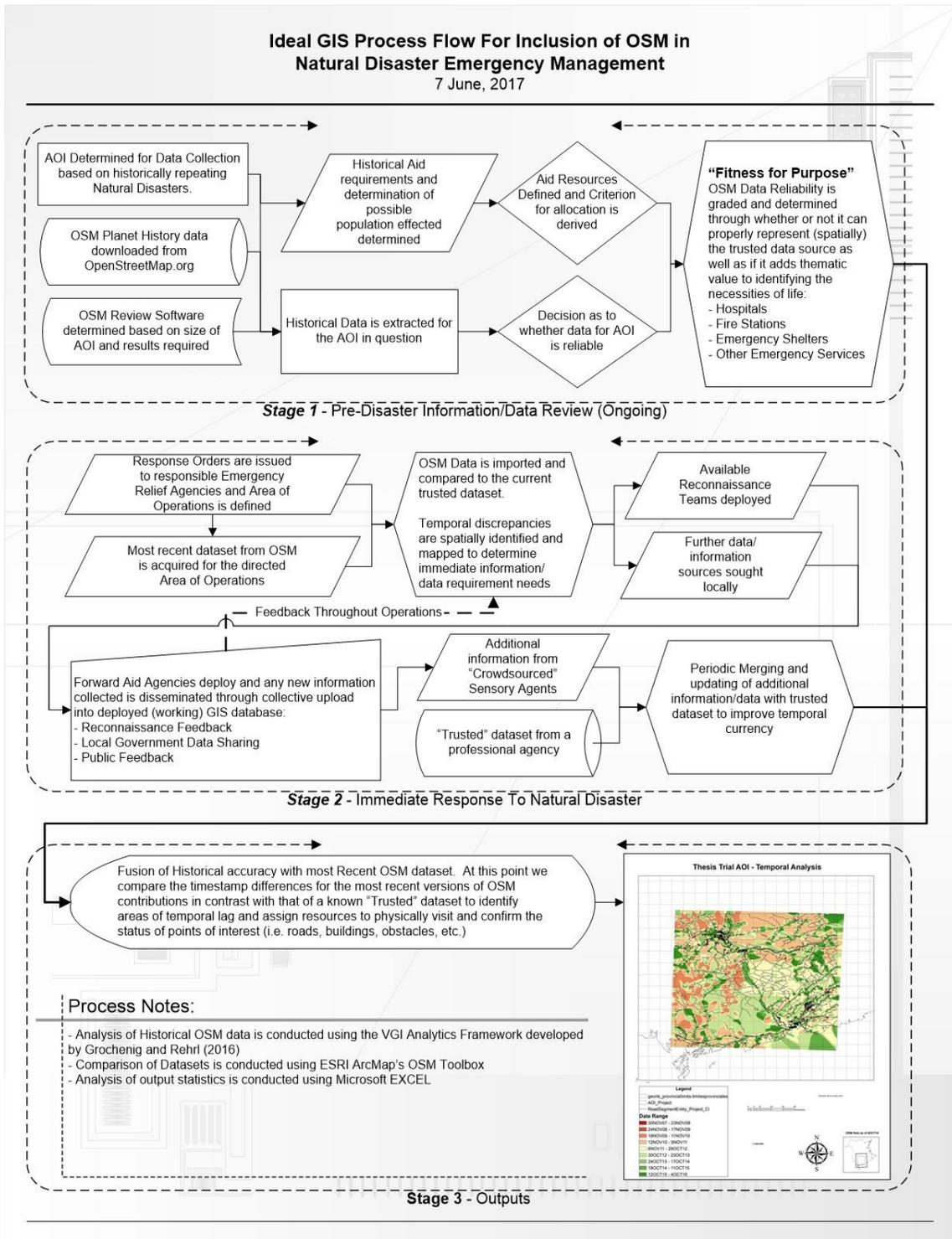


Figure 2.9 – Potential Workflow for OSM use/inclusion in a Natural Disaster Aid Operation

Figure 2.11 presents an idealized workflow which was designed at the proposal stage of this thesis research, and then served as a guideline for the further development of

processes utilized in Chapter 3. The key point is that methods to not only compare the differences between the OSM dataset and a professionally generated dataset were being sought, but also efforts to show how the OSM dataset could be included to provide useful information to augment a dataset in an emergency management scenario.

Table 2-2 – Related OSM Accuracy and Reliability Studies

Studies	Contribution	Spatial Scope	Tool/Framework	Analysis Scope
Haklay et al. (2010)	Application of Linus Law to OSM Testing of positional accuracy to determine relationship between number of contributors and quality of road data	Four 1 mi ² study sites within London, UK	Unknown Tool ITN Road Network vs. OSM dataset	Contribution Profiling, Completeness Estimation
Neis and Zipf (2012)	Contributor analysis. Number of changes conducted in an area, when and where are contributors active	Global	Java Application Open Source Libraries	Contributor / Contribution profiling
Mooney and Corcoran (2012)	Analysis of the history of OSM entities which have been editing to at least version 15; Identification of disagreements between contributors on specific entities	UK, Ireland	OSM History Splitter PostgreSQL	Change Detection Contribution Profiling
Neis et al. (2013)	Contributor and data analysis in relation to area and population density in selected cities; more contributors locally found in European cities vs. higher non-local contributors for cities outside Europe	Global (Twelve different urban areas)	Java Application Open Source Libraries	Contribution Profiling
Rehrl et al. (2013)	First conceptual model for analysis of VGI activity data	Global	Activity Model	Contribution Profiling
Barron et al., (2014)	Analysis of OSM data based on a 25 intrinsic quality indicator framework for fitness of OSM data for purpose; generates statistics, diagrams and maps	Global	iOSMAnalyzer	Contribution Profiling, Completeness Estimation, Change Detection
Mooney and Corcoran (2014)	Review of contributor behaviours and interactions between them; mapping behaviour is not predictable; contributors show clustering and tendency to work on the same entities	Global (Seven Cities)	Unknown Tool	Contributor / Contribution Profiling
Rehrl and Gröchenig (2016)	Analysis of OSM global history file; Framework allows for the analysis of contribution type volumes during a specific timeframe in day/month/year formats	Global	VGI Analytics Framework	Contribution Profiling, Change Detection, OSM Dataset Growth Monitoring

Table 2-2 provides a collective timeline of related research into the various aspects of reliability associated with OSM data and the tools used perform the analyses. The tools created specifically for OSM data analysis have been evolving since before 2010

(Zulfiqar, 2008); however, published results are not widely available. This evolution of analysis methods has seen the approaches grow from the simpler dataset parsing tools such as OSMConvert, OSMosis and OSMembrane into more advanced frameworks for data visualization and interpretation such as iOSMAnalyzer, or the VGI Analytics Framework, used in the research method of this thesis.

2.5. Summary

The concepts and case studies discussed in this section frame the current movement forward into the future use of VGI as a trusted and reliable resource, dependent on the intended use of the information. Current distrust of incorporating VGI in the process of generating trusted GIS based mapping products centers on the fact that the concept of VGI in the form of OSM data credibility is still relatively new and in the process of having standards established to streamline its contributions. This situation brings about the concern that it cannot be properly validated or reviewed through traditional methods. Many of the approaches brought forward in this review show that certain aspects of standardized data can be let go, i.e. the concept of spatial and thematic completeness as discussed by Coast (2007), in a form of provision of data under the pretense that occasions arise when it simply must be asked, “Is having unreliable data better than having no data at all?”

In an ideal world, the quality of geographic data provided by OSM would strictly follow the internationally recognized standards set forth by the ISO/TC 211 Geographic Information / Geomatics Organisation (ISO 19113:2002) established in 2011. Although this standard for data collection was established, most VGI is still plagued by the fact that individuals are not held accountable for the accuracy and comprehensiveness of their

contributions and as such cannot be considered reliable when viewing their individual additions to the efforts of a VGI project. As a counter to this, most studies in the field have focused on methods which ensure that VGI use should not be employed as a sole provider of information but as a measure to bolster current information generation processes or simply just be used to verify what is generated by professional scientists and their affiliated agencies as the common truth associated with the region under study. Although these could be construed as secondary roles, the importance of continuing to generate more and more examples of proven results using VGI only strengthens its credibility and robustness for inclusion in the information gathering process.

Chapter 3: Methods and Study Area

3.1 Introduction

This chapter describes the methods used to address the research questions set forth in Section 1.4. Sections 3.1 and 3.2 provide basic information about the study areas and the data which were collected and analyzed, most of which were provided through the OpenStreetMap.org project and its affiliates, or GeoNB. Section 3.3 describes the methods used to review the historical data associated with the OSM dataset, which are directly associated with answering questions 1.4.1 and 1.4.2. Section 3.4 then describes the methods developed and tested for addressing the matters of research questions 1.4.3 and 1.4.4, framed around a test case scenario for the process workflow. The test case scenario was developed to review the different software platforms used in generating the results, and determine how to best utilize them in an expedient review process for the available OSM data. Finally, Section 3.5 discusses the approach made in the thematic review of OSM data.

This research used data from professional sources such as GeoNB and the Government of Canada's GeoGratis CanVEC data repositories in comparison to the OSM dataset. The professional datasets under scrutiny used varying projection schemas, so in order to enable consistent analysis and comparisons, the ESRI ArcGIS 10.3 software platform was used to re-project all of the data into the standard World Geodetic System (WGS) of 1984 (USGS, 2017b), which is used by the OSM project. The documentation of these professionally generated datasets does confirm that they are in fact amalgamations of several different sources, from varying time periods, and as such are only accredited as being current as of their release dates.

For the OSM data reviewed in this research, mixtures of the most current dataset (in the case of the process outlined in Section 3.4) as well as the historical extracts found in Section 3.3 were reviewed. It is critical to remember that since the OSMF took over the project, the contributions to the OSM dataset are guided not bounded by the guidelines set forth by the OSM project at http://wiki.openstreetmap.org/wiki/Map_Features. These guidelines form the basic framework for the dataset; however, they rely on the contributors to properly adhere to them and as such attributional information for contributions is often found to contain a large degree of heterogeneity.

The data processing for this study was all conducted on an ASUS© model UX305F Notebook PC with an Intel® Core™ M-5Y10c CPU @ 0.80 GHz 998 MHz with 8GB of RAM. The computer had an internal 256GB solid state hard drive and an external 3TB Hard Disk Drive. The operating systems used for the methods discussed in Section 3.3 were a mix of the Linux-based Ubuntu© 16.04 LTS and Microsoft Windows 10©. The process development discussed in Section 3.4 was performed in Microsoft Windows 10 with ArcGIS Desktop 10.3©, and the ArcGIS Editor for OpenStreetMap tools extensions. The analysis and graphical representations for the resulting data from both of these sections were conducted using Microsoft Office EXCEL 2010© Home and Student edition.

The primary reasoning for the choice to limit the majority of the analysis methods to a Microsoft Windows-based platform was driven by the fact that the end goal of this study would be to provide methods for deployed Geomatics Support Teams (or GSTs) within the CAF to utilize existing hardware and software at their regular and usual disposal rather than searching for other means such as those openly available through the

Internet. Although the CAF, and more specifically MCE, are constantly updating their electronic hardware and software capabilities, the process workflow developed in this thesis should allow for them to use their existing resources, only with a view to speed up the process as capabilities improve. Ideally the analysis process developed in Section 3.4 would be able to be applied when not connected to the Internet in a forward operations situation. This is due to the fact that, as we have seen, when a natural disaster occurs there are often outages in critical infrastructure such as telecommunications. The workflow itself allows for rapid extraction of OSM data for a targeted area before deploying to an affected region. Once deployed, comparative analysis to a trusted dataset, if available, and spatial interpolation of the temporal currency of the OSM data for the purposes of assigning data collection and reconnaissance needs could be conducted.

3.2 Study Area Selection

This portion of the research methodology discussion seeks to outline the various geographic foci of the research conducted and why it would be relevant to form an understanding of the VGI efforts made both historically and in a current context. In the global context, absolute confidence was placed on the legitimacy of the OSM data available due to the lack of requirement for a comparative dataset. In the context of the Canadian perspective, a data provider commonly used by the CAF was available for the comparative approaches to be utilized and as such the region was ideal for the purposes of this research.

3.2.1 Global Analysis of the OSM Dataset

Section 3.3 reviews the responses of the OSM community immediately following a significant natural disaster. Earthquakes are a globally occurring common phenomenon

(e.g., see Figure 3.1) which has warranted numerous studies into their effects and likelihood to occur again. The analysis conducted in Section 3.3 is based around the four major earthquake events described in Section 2.3 (Table 2-1). These earthquake events were all similar in magnitude; however, the reaction of the OSM community varied greatly. It is with that consideration in mind that a method was sought to derive a snapshot of the actual contributions made in the different regions leading up to, and immediately following the earthquakes.

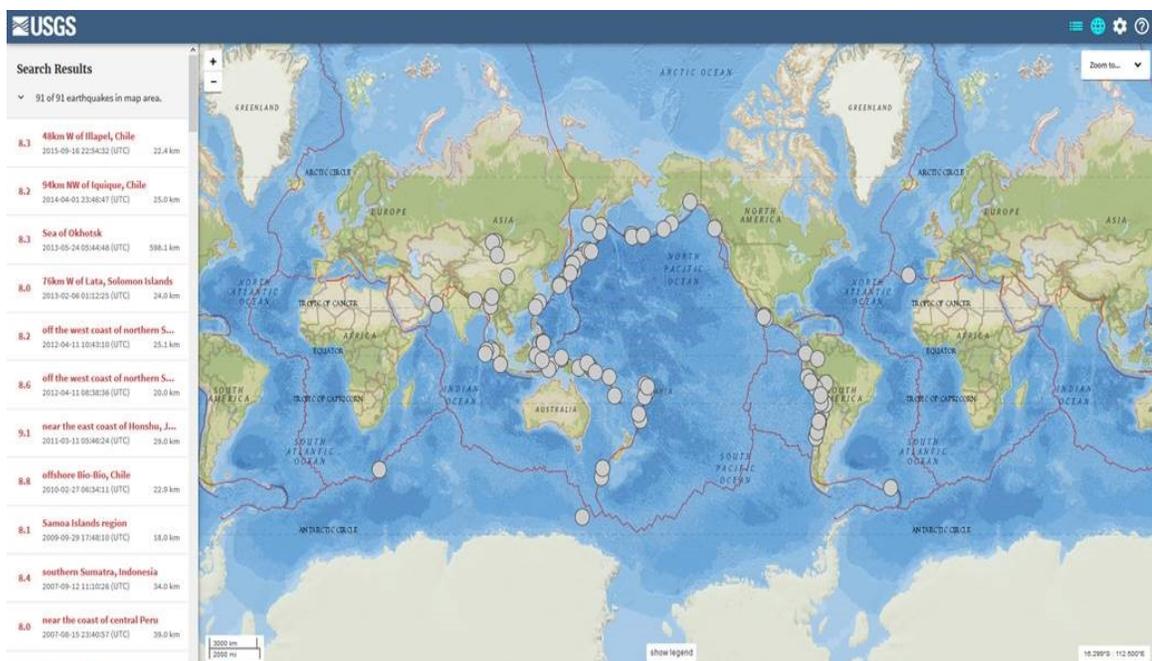


Figure 3.1 - Locations of magnitude 8.0 or greater earthquakes ever recorded (USGS, 2017a)

The OSM project provides its data in two different formats, the commonly used .osm and history log .osh. Within the context of the .osm format, only the most current values of the data can be found (Gröchenig, 2012; Gröchenig et al., 2014a; Rehrl et al., 2013). This format is smaller in size and thus ideal for use in immediate response to natural disaster scenarios; however, the replacement/removal of preceding versions of the data in this format diminishes the ability to temporally review the dataset immediately

surrounding a specific event. The .osh format of the dataset is maintained to provide a complete log of all valid edits (i.e. versions) made to the dataset. This log contains all of the previous values, since October 2007, associated with the nodes, ways and relations contained in the dataset and provides the data required to accurately analyze and understand how the OSM dataset looked at a specific point in time for use in the review of the reactions of the OSM community surrounding a disaster.

Among the tools which have been developed to review the timeline of contributions to the OSM dataset, one example is [OpenStreetMap Analytics_{beta}](#) (Figure 3.2) which provides a view of the current data values globally for the .osm dataset.

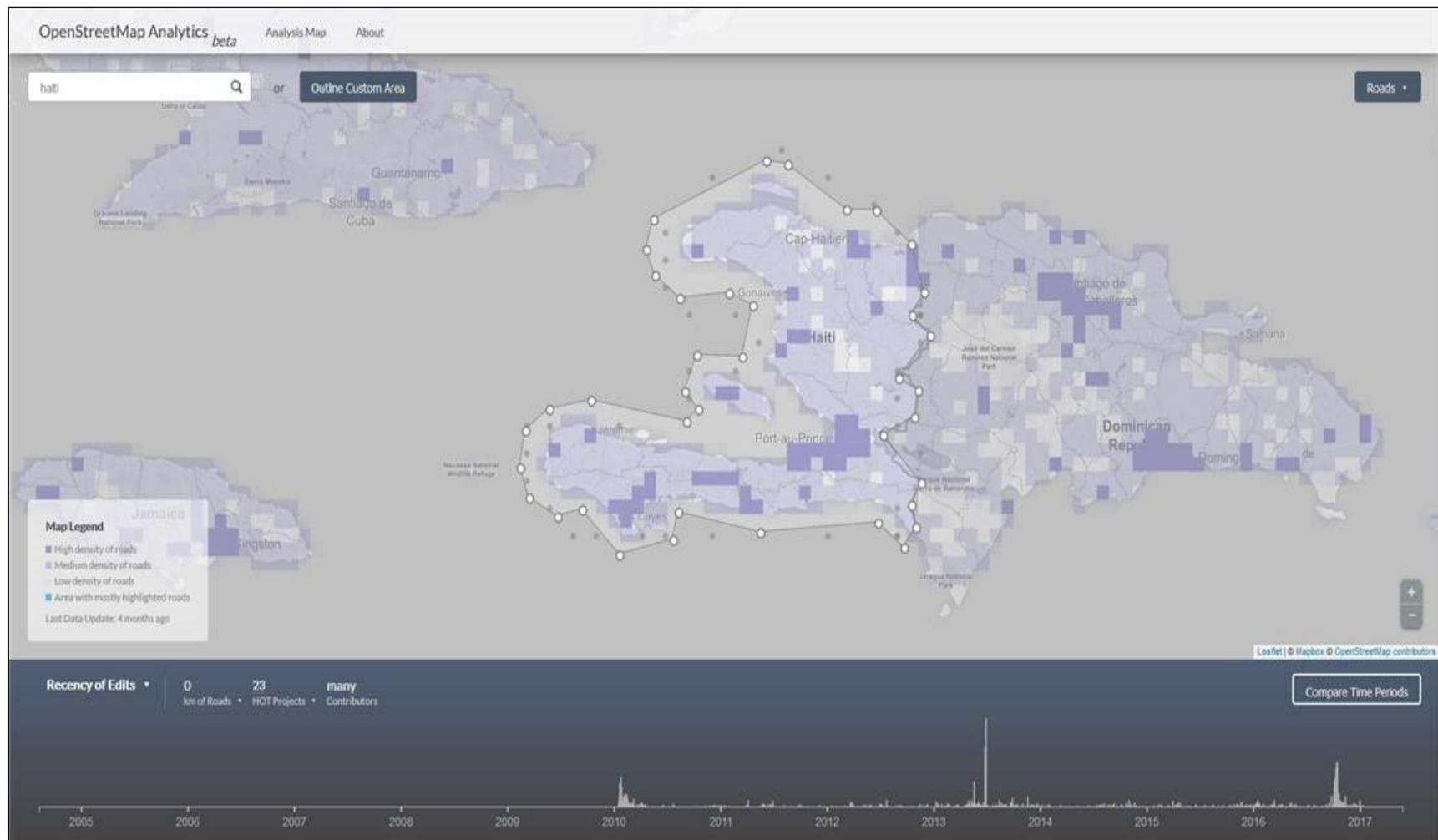


Figure 3.2 – Timeline associated with the most recent contributions to the OSM dataset for Haiti ([OpenStreetMap Analytics beta](#))

The existence of these types of tools is indicative that automated processes for reviewing the OSM dataset are already in the development stages; however, as the documentation for this tool admits in discussing its limitations, the interpretation of the provided graphs and statistics is limited to current contributions found within the .osm data and not the full history (.osh) dataset. As an example of the effects of these contribution version replacements within the .osm format of the dataset, Figure 3.2 would indicate that the spike in the 2010 VGI inputs to the dataset is quite small in comparison to the large volume of contributions made in the mid-2013 and late-2016 timeframes. Through the lens of the OpenStreetMap Analytics *beta* tool, the more recent spikes in contributions would in effect cause the relative volume of entities, with a timestamp from 2010, to decrease as they are replaced by their more recent versions. As was discussed in the contribution activity review by Neis et al. (2010), the region of Haiti under scrutiny saw an almost exponential surge in contributions following the disastrous earthquake in 2010. Based on this observation, I decided that in order to review the historical contribution (.osh) data for a specific region, a tool such as the VGI Analytics Framework developed by Rehrl and Gröchenig (2016), would be employed. For the purposes of this research, it was also decided to remain focussed on the contributor behavior rather than the specific spatial changes and the maturity of the spatial dataset in order to gain a further understanding of how the community would react following a natural disaster.

3.2.2 Reviewing the OSM Dataset with a Canadian application perspective

For the purposes of testing the process workflow derived in Section 3.4, an area which would most accurately capture repeated natural disaster effects in Canada had to be chosen. This decision was based on information security considerations for the

availability of highly spatially accurate data which could be accessed and analyzed openly within the public domain. As we have witnessed historically, the Canadian landscape encounters a broad spectrum of environmental conditions which have caused events such as forest fires (e.g. the 2016 Fort McMurray Wildfire), ice storms (e.g. in the St. Lawrence Valley in 1998), and in the case of this study, severe overland flooding, which has been repeatedly experienced in the province of New Brunswick (Government of Canada, 2013).

2008 New Brunswick Flooding Extent

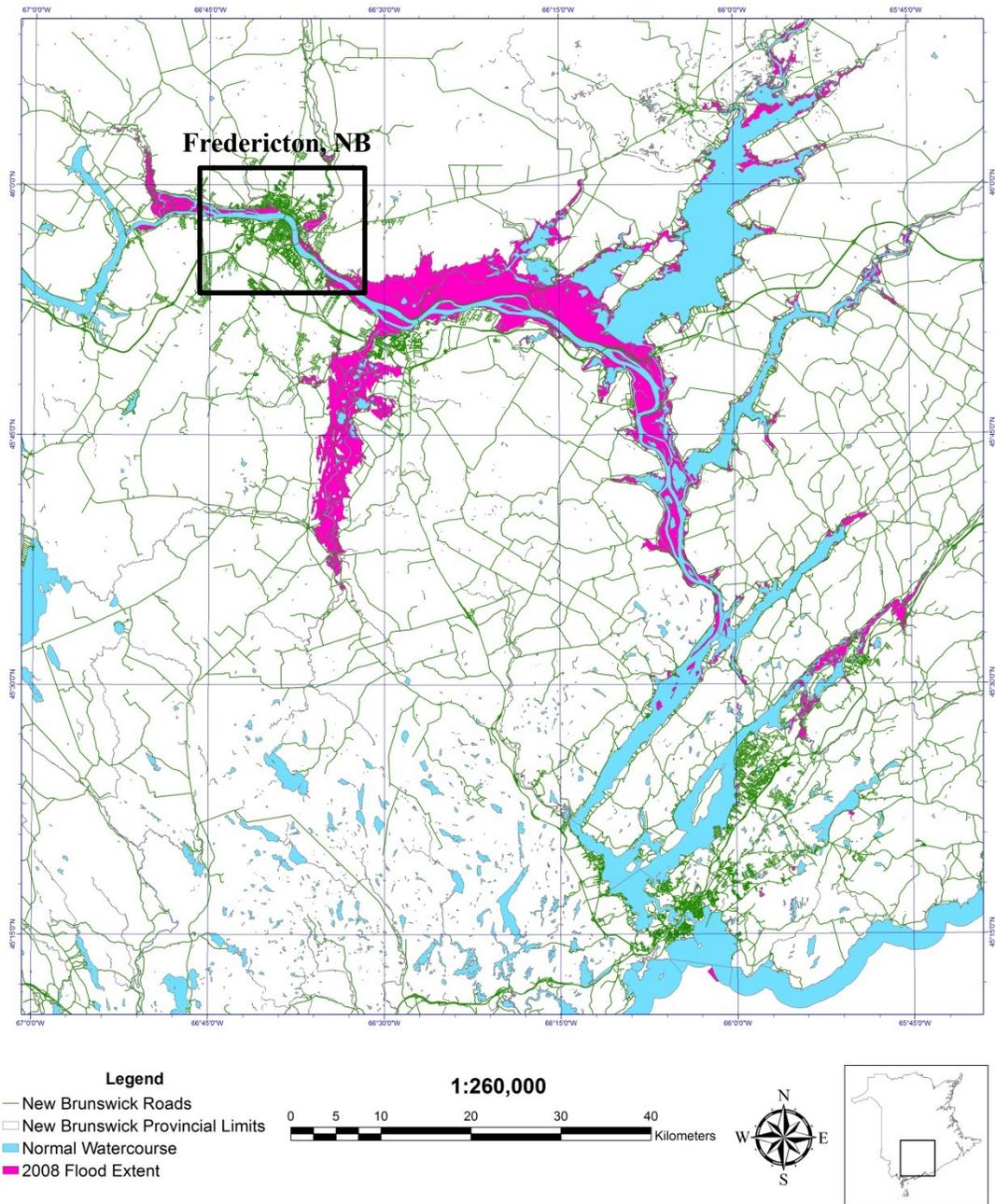


Figure 3.3 – 2008 New Brunswick Flood Extent

In consulting with MCE, it was determined that the openly available data provided by GeoNB would be most appropriate for use as the “Gold Standard” dataset to be used in comparison with the OSM dataset in order to determine the OSM data’s spatial

accuracy. This was accomplished using a similar process to that developed by Goodchild and Hunter (1997) and applied by Girres and Touya (2010), Haklay (2010), Haklay et al. (2010), and Zheng and Zheng (2014).

As we see in Figures 3.3 and 3.4, the flooding in New Brunswick in 2008 caused extreme damage to the local infrastructure, resulting in aid being requested from the DND/CAF. This flooding was attributed to the rapid spring snow melt associated with the record setting snowfall experienced the winter before, and was responsible for the worst recorded overland flooding in New Brunswick history. Since the CAF was involved with assisting in flood relief efforts in the province of New Brunswick, or more specifically the region surrounding the flood plain of the Saint John River, it was appropriate to scrutinize this as an Area Of Interest (AOI) for testing the workflow derived in Section 3.4.



Figure 3.4 – Images of the 2008 Fredericton, New Brunswick Flooding (Source: 4 ESR)

3.3 Analysis of the Historical Data Associated with the OSM Dataset

In this section a repeatable and measurable process for reviewing the history of OSM data for a specific region, during a specified period of time is sought. The process developed was primarily based on the VGI analytics framework tools designed by Rehr and Gröchenig (2016). The outputs of the framework which can be reviewed through this process are quite dramatic and could help to ascertain what further considerations should be made before a level of uncertainty could be accepted when using OSM VGI in regions where little to no other current or professionally generated geographical data exist

3.3.1 Data Used

The data used in this research for the historical analysis surrounding the four earthquakes outlined in Table 2-1 were downloaded directly from the OSM Project affiliate, Geofabrik.de. This was done in order to speed up the analysis process by only using pre-extracted portions of the OSM Planet History (.osh) file itself at <http://planet.openstreetmap.org/pbf/full-history/>. As a note on the data used for this portion of the research, the .osh file in Protocolbuffer Binary Format³ (or .pbf) extracts were bounded by the .poly⁴ files available for each region at their respective download site listed in Table 3-1.

³ Protocolbuffer Binary Format is primarily intended as an alternative to using an XML format and was designed to support future extensibility and flexibility.

⁴ The Osmosis polygon filter file format is supported by Osmosis, mapsplit, osmconvert, osmchange and pbftoosm as a way of defining regional geographic information extraction polygons.

Table 3-1 – Data sources used and time periods reviewed in the Historical Analysis of the OSM History (.osh.pbf) for the different regions

Earthquake Region	Open Data Download Location	DTG of Earthquake	Timespan Reviewed ⁵
Haiti and Dominican Republic	http://download.geofabrik.de/central-america/haiti-and-domrep.osh.pbf	12 JAN 2010 1653H	Analysis Start Date = “2009-01-12 T00:00:00Z” Analysis End Date = “2010-02-12 T00:00:00Z”
Nepal	http://download.geofabrik.de/asia/nepal.osh.pbf	25 APR 2015 1156H	Analysis Start Date = “2014-04-25 T00:00:00Z” Analysis End Date = “2015-05-25 T00:00:00Z”
New Zealand	http://download.geofabrik.de/australia-oceania/new-zealand.osh.pbf	14 NOV 2016 0002H	Analysis Start Date = “2015-11-14 T00:00:00Z” Analysis End Date = “2016-12-14 T00:00:00Z”
Chile	http://download.geofabrik.de/south-america/chile.osh.pbf	25 DEC 2016 1422H	Analysis Start Date = “2015-12-25 T00:00:00Z” Analysis End Date = “2017-01-25 T00:00:00Z”

Understanding that the scale of the geographical areas covered by the different .poly regions varies in size, which made the volume of contributions for the time periods analyzed differ in magnitude, the overall behaviour of the contributors themselves remained similar when other factors such as the availability of an open data infrastructure are considered. In order to address the issue of regional size differences, the initial intent of this research was to download and analyze user defined areas of interest surrounding the earthquake epicenters which would be of the same size. Attempts to accomplish this were made using several different tool packages; however, limitations in both software and computing power along with the availability of physical memory space for unzipping and reviewing the Full Planet History file prevented this from being a viable option. These capability restrictions caused further analysis to be performed within the parameters outlined in Table 3-1.

⁵ These are in the format used by the VGI Analytics Framework’s settings_batch.xml to extract the time period of interest surrounding the earthquakes.

Since the process being developed here was designed to be used with the tools available to a forward deployed GST, the potential workflow/process outlined in Figure 2.9 for Stages 2 and 3 of disaster management needed to be constrained. The OSM Full Planet History file in its entirety could only be considered for use on the larger and more powerful computers at the MCE HQ. In the context of this research, that option was ruled as untenable, and the pre-processed regional extracts from Geofabrik.de were deemed acceptable alternatives.

3.3.2 Process Steps Involved in the Analysis and Related Outputs

The idealized workflow/process developed in the proposal stage for this portion of the thesis research can be found in the flowchart at Stage 1 of Figure 2.9. The design of the VGI Analytics Framework by Rehr and Gröchenig (2016) are now reviewed to provide an excellent breakdown for analyzing mapping activity in the context of OSM and forms the blueprint for the methods used here. It is important to remember that the XML Schema used by the OSM dataset is composed of Node, Way, and Relation data types and that all operations performed on the dataset are captured and stored within the .osh.pbf dataset. An overview of the executable operation types which could be conducted in a change to the OSM dataset can be seen in Table 3-2.

Table 3-2 – Executable Operation Types on the Different OSM Data Types (Rehr and Gröchenig, 2016)

	Node	Way	Relation
Create	Op Create Node	Op Create Way	Op Create Relation
Update (geometry)	Op Update Coordinate	Op Add Node	Op Add Member
		Op Remove Node	Op Remove Member
		Op Reorder Node	Op Reorder Member
		Op Update Coordinate	Op Update Coordinate
Update (attribute)	Op Add Tag	Op Add Tag	Op Add Tag
	Op Update Tag Value	Op Update Tag Value	Op Update Tag Value
	Op Remove Tag	Op Remove Tag	Op Remove Tag
Update (entity)	Op Recreate Node	Op Recreate Way	Op Recreate Relation
Delete	Op Delete Node	Op Delete Way	Op Delete Relation

As pointed out by Rehrl and Gröchenig (2016), OSM contribution activity needs to be thought of as a structured VGI mapping activity in order to appropriately form approaches such as those sought in this thesis research. Table 3.3 applies the concepts of Activity Theory (Kuutti, 2006) to VGI mapping activities, and provides comparative OSM editing activity types.

Table 3-3 – *Concepts of Activity Theory applied to VGI mapping activities (Rehrl et al., 2013)*

	Activity Theory	Applied to VGI Mapping Activity	OSM Dataset Example
Activity	A chain of associated actions following the same motive.	A set of mapping actions driven by a certain goal (e.g. improving accuracy)	Emergency mapping of a specific region.
Action	A goal-driven part of an activity. An ordered sequence of operations.	A sequence of operations by a single volunteer within a certain timespan manipulating a single feature to reach a certain goal (e.g., editing a feature geometry)	The series of operations related to creating a single line feature.
Operation	Atomic. Being part of an action. Being executed automatically in order to reach a certain goal.	Atomic Change to a single feature by a unique volunteer and at a unique timestamp (e.g. adding a node)	Each of the individual executable operations carried out in the creation of a map feature.

As a simple example, which was discussed in Rehrl and Gröchenig (2016), consider the construction of a new road which is connected to an existing one. Depicted in Figure 3.5, the series of executable operations outlined in Table 3-4 would be representative of the OSM dataset update.

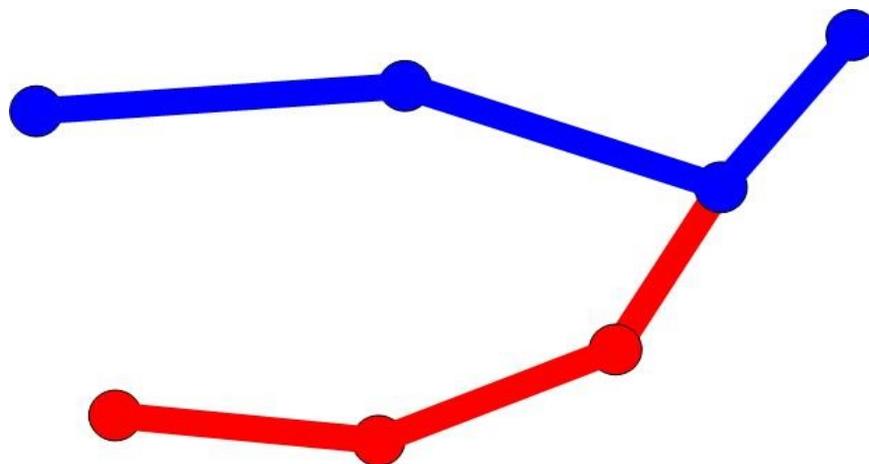


Figure 3.5 – *A new road (red) being created and connected to an existing one (blue)*

Table 3-4 – OSM executable operation sequence used for the mapping action of adding a road

Entity ID	Operation Type	UserID	Timestamp
Node 5	Op Create Node	Map Cont 123	2010-01-12 16:53:00
Node 6	Op Create Node	Map Cont 123	2010-01-12 16:53:00
Node 7	Op Create Node	Map Cont 123	2010-01-12 16:53:00
Way 2	Op Create Way	Map Cont 123	2010-01-12 16:53:00
Way 2	Op Add Node (Node 3)	Map Cont 123	2010-01-12 16:53:00
Way 2	Op Add Node (Node 5)	Map Cont 123	2010-01-12 16:53:00
Way 2	Op Add Node (Node 6)	Map Cont 123	2010-01-12 16:53:00
Way 2	Op Add Node (Node 7)	Map Cont 123	2010-01-12 16:53:00
Way 2	Op Add Tag (highway = rural)	Map Cont 123	2010-01-12 16:53:00

Designing an analysis process to review the OSM Planet History File required an understanding that the VGI Analytics Framework was designed to provide its results based on the Create, Read, Update, Delete (CRUD) paradigm of database management (Martin, 1983). This paradigm formed the basis by which the entities defined in the OSM data model (Nodes, Ways and Relations) may be created, read, updated or deleted within their host data structure, continuously over time. This meant the framework designers had to define specific functions for reading and interpreting the executional operations contained within the Planet History .osh.pbf file structure. The functionality requirements which were developed in the framework with respect to this could ideally, allow for the .osh.pbf data to be extracted for a distinct AOI, at a distinct time interval with regards to a natural disaster and review the response of the OSM community in that AOI. Following this line of thought, a potential process for extrapolating and utilizing the outputs of the framework was conceptualized as seen in Figure 3.6.

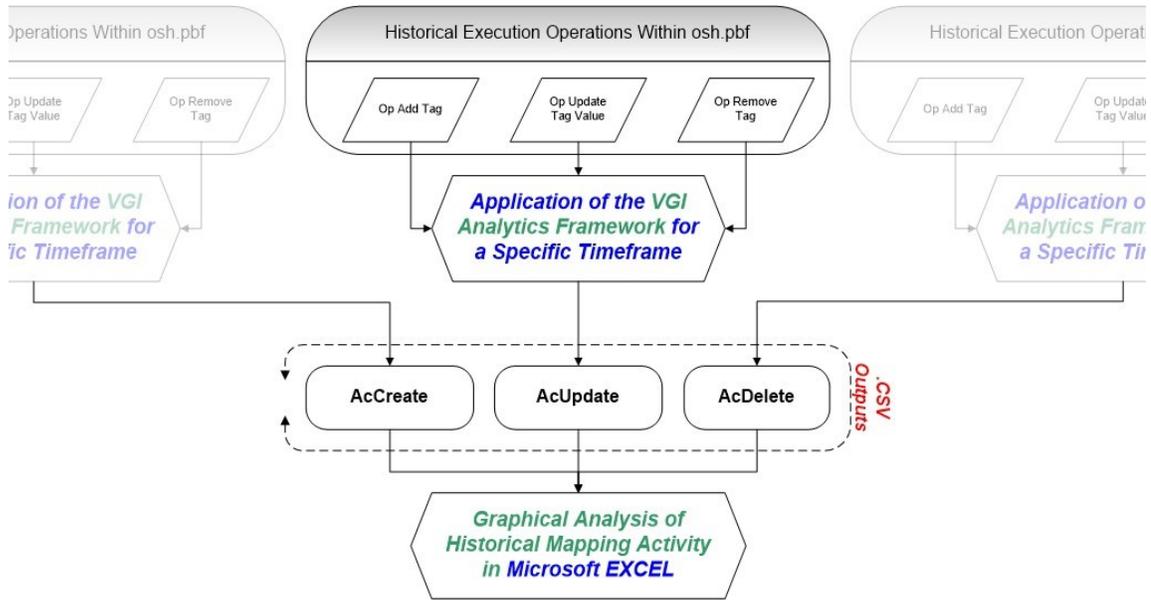


Figure 3.6 – Conceptual Process for the Analysis of osh.pbf Data for employment in the regular review of the OSM Dataset’s growth with regards to contribution reliability

This conceptual model for analysis was then tested to determine a best workflow for the inclusion of the VGI Analytics Framework. This process involved several iterations which will not be included in this discussion; however the final idealized process workflow is detailed in Figure 3.7 (enlarged at Annex A). The figure itself provides a record of the exact steps and commands used to extract and analyze the data used in the research process.

VGI Analytics Framework use for Temporal Review of OSM Contributions

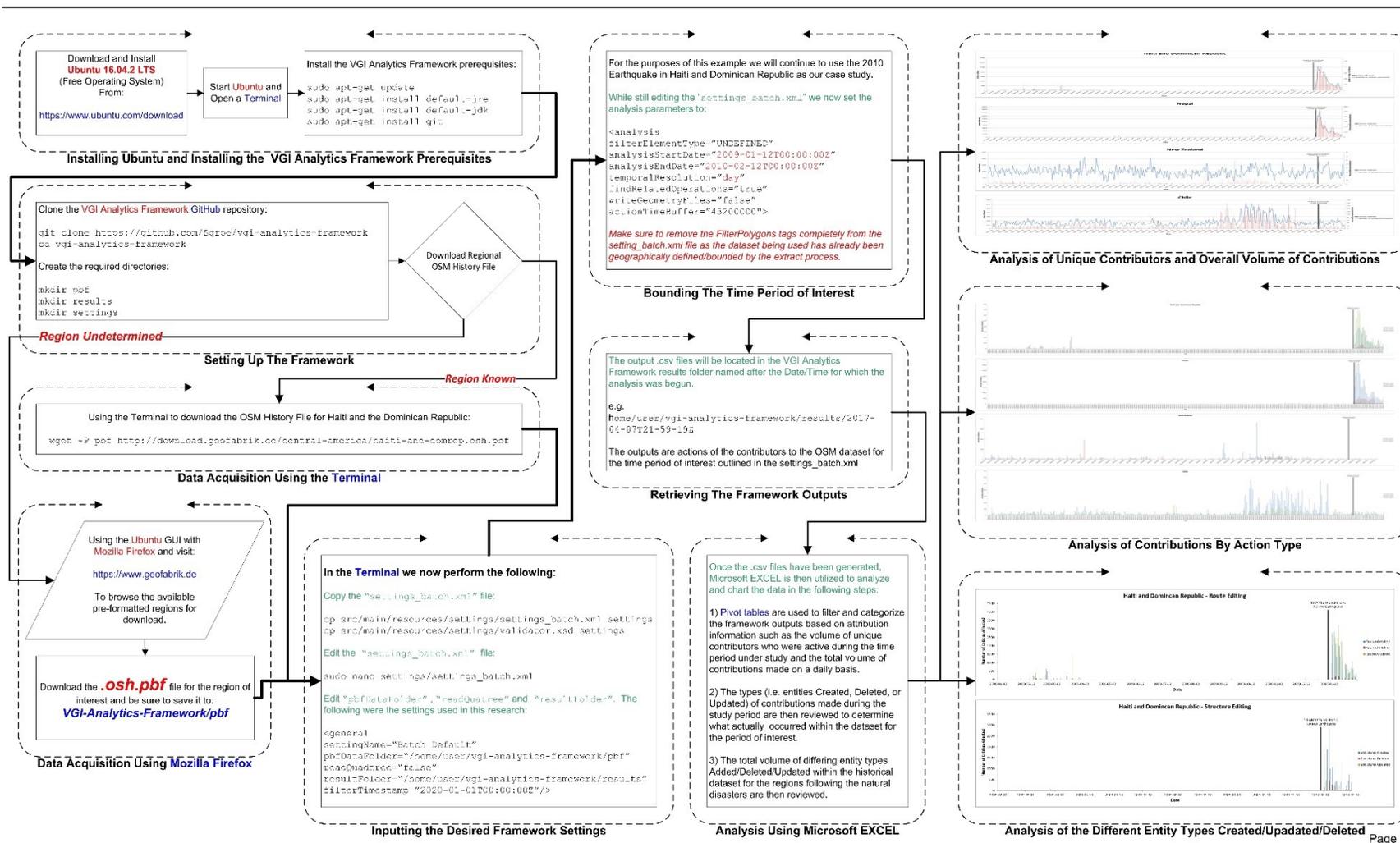


Figure 3.7 – Process Workflow for Extracting OSM History data utilizing the VGI Analytics Framework and Microsoft EXCEL

A general overview of the steps used to analyze the dataset follows:

As discussed in Section 3.3.1, the AOI used for historical review was determined by the extracts available from Geofabrik.de:

1. The OSM History File (.osh.pbf) was downloaded from Geofabrik.de for the four main AOI listed in Table 3-1.
2. Once the data had been downloaded, the following steps were then applied to each of the separate AOI:
 - (a) Using Linux/Ubuntu 16.04 LTS, The VGI Analytics Framework (Rehrl and Gröchenig, 2016) was utilized to extract the historical data for a time period of 1 month immediately following each of the earthquakes, and a period of one year preceding each of the disasters.
 - (b) The resulting outputs were delivered in comma separated values (.csv) format which then allowed for simple data analysis using pivot tables and graphs generated in Microsoft EXCEL.
3. The counts of contributions/edits to the OSM dataset for the time period under scrutiny surrounding the earthquakes were then charted and reviewed. (Annex A, Appendix 1).
4. The contribution actions themselves (i.e. entities being Created, Updated, or Deleted) conducted by the OSM contributors were then isolated and reviewed (Annex A, Appendix 2)
5. In light of the work by Gröchenig et al. (2014b), the creation of linear entities (i.e. routes) and closed relation entities (i.e. structures) were then separated in order to test

if the development of the road networks for the regions held true to patterns observed by Gröchenig et al. (Annex A, Appendix 3).

At this stage of the OSM analysis process it should be noted that other analysis metrics can be derived using the VGI Analytics Framework. These metrics and representations of the data contained within the dataset were deemed as unnecessary for the purpose of this thesis, but do merit mention. For example, the VGI Analytics Framework provides such functionalities as the ability to generate a Quadtree representation of the dataset as seen in Figure 3.8. This representation of the dataset, where the colours indicate the number of features contained within a cell, could be used to visually understand what volume of feature contributions have actually been made to an AOI, and form an understanding of whether or not data for the region could potentially be trusted as properly peer reviewed.



Figure 3.8 – Quadtree structure representing the spatial indexing of the operations generated from a September 2015 OSM Full History File (Rehrl and Gröchenig, 2016)

3.4 A Process for OSM Dataset Usage in Natural Disaster Response

The process derived and tested in this section focuses on a regional analysis of the current OSM dataset in an area which a natural disaster has recently occurred and aid is about to be deployed. The analysis assumes that the OSM dataset for the region has been recently reviewed using the process outlined in Section 3.3, and it has been determined that it contains appropriate data which are representative of the region.

The process can be used to target an area of geographic interest and identify the temporal deficiencies of the data for that area. Inverse Distance Weighted (IDW) spatial interpolation was applied to the magnitude value derived from the time stamp and version information regarding the individual OSM contributions for the region to provide a visual reference of potential areas of data collection interest. These results would then be used to prioritize information collection requirements following a natural disaster event such as an earthquake or flood.

3.4.1. Reference Data

In this portion of the process workflow a determination of the spatial accuracy of the OSM dataset is sought. For the purposes of the process development for what would be considered Stage 2 of the ideal process workflow (Figure 2.9), the open data sources listed in Table 3-5 were used.

Table 3-5 – Data Sources used in Spatial Comparison Process Development

Data Provider	Open Data Download Location	Date of Data Download	Size of Download
GeoNB	http://geonb.snb.ca/downloads/nbrn/geonb_nbrn-rrnb_shp.zip	25 FEB 16 ⁶	150 MB
Geofabrik.de	http://download.geofabrik.de/north-america/canada/new-brunswick-latest.osm.bz2	16 MAR 16	85 MB
Overpass	http://overpass-api.de/api/map?bbox=-67.0000,45.1000,-65.7500,46.1250	24 MAY 16	504 MB

⁶ The actual date of release for the Trusted Data used in this analysis was 20 MAR 15 and the link provided in the Open Data download location will, at the time of publication, provide the more recent 31 MAR 17 open data from GeoNB.

The open data from GeoNB was used as the trusted benchmark or “Professionally Generated” dataset for the basis of spatial accuracy comparison. As detailed in Table 3-5, there were two providers of OSM data reviewed. For the process developed in this section, the current OSM data for the region were used rather than the Full History data utilized in Section 3.3. This was done in order to ensure that only the most recent updates (versions) to the dataset were considered when generating the output products.

The data from GeoNB came as a readily ingestible ESRI ArcMap shapefile which required re-projection into WGS 84 in order to be used in the comparison process. In the case of the Geofabrik.de dataset, the data were pre-processed by Geofabrik.de to convert the OSM data into a standard GIS shapefile which greatly aided in expediting the analysis process. For the use of the .osm data provided by Overpass, the ESRI ArcMap OSM toolbox was used to download, extract, and symbolize the OSM data for a bounded area of interest rather than a pre-defined geographic region such as that provided by Geofabrik.de. The pure OSM dataset from Overpass is quite a bit larger than the others, due to its more detailed set of attributes.

3.4.2 Process Steps Involved in the Analysis and Related Outputs

As a starting point to this portion of the thesis research, a spatial comparison for linear features (roads) similar to that developed by Goodchild and Hunter (1997) was employed. Due to the fact that there is no de-facto standard convention for contributions to the OSM dataset, the highway_Line feature class is usually observed as having the greatest amount of volume and length contributions of linear features (when converted from Nodes, Ways, and Relations) within the dataset. Upon acquiring the initial datasets (trusted and OSM) for the entire province of New Brunswick, a quick summation of the

linear feature lengths they contained was conducted. The OSM dataset contained a far greater summed feature length. In order to further develop this process based on operational requirements, I assumed that any and all available routes to a destination would be considered as viable for use in emergency aid provision. Due to the fact that the OSM data contained a far greater summed linear feature length, the decision was made that instead of using a spatial comparison based on the trusted dataset, which was the common practice in the previous studies, the spatial buffer analysis could also be applied using the OSM dataset. For the purposes of this research, the spatial buffer comparison process developed and discussed in Goodchild and Hunter (1997) and used by Haklay (2010) was instead applied to the OSM highway_Line features vice the trusted dataset. For this analysis buffers were placed at 1m intervals (out to 20m) around the OSM highway_Line features and then used to clip the trusted dataset. This analysis then provided an empirical measure of the average spatial variance between the two linear networks. This measure could then be plotted and reviewed to gain an understanding of what distance can a decision maker be $X\%$ assured that the OSM features are within $\pm Y$ (in m) of their trusted counterparts. This form of the spatial analysis method was then conducted to determine the average spatial variance between OSM data and the trusted data provided by GeoNB.

3.4.2.1 Spatial Accuracy Analysis of OSM Data Provided by Geofabrik.de

In order to determine a proof of concept for this portion of the process workflow design a method of expedient spatial comparison between .osm data and a trusted dataset was sought. The regional .osm shapefile from Geobarik.de was downloaded and used as the comparative dataset against the trusted New Brunswick Road Network (NBRN)

dataset. This was partially due to fact that the data provided by Geofabrik.de are the primary source of OSM data used by MCE. The process for this comparison is detailed in Figure 3.9.

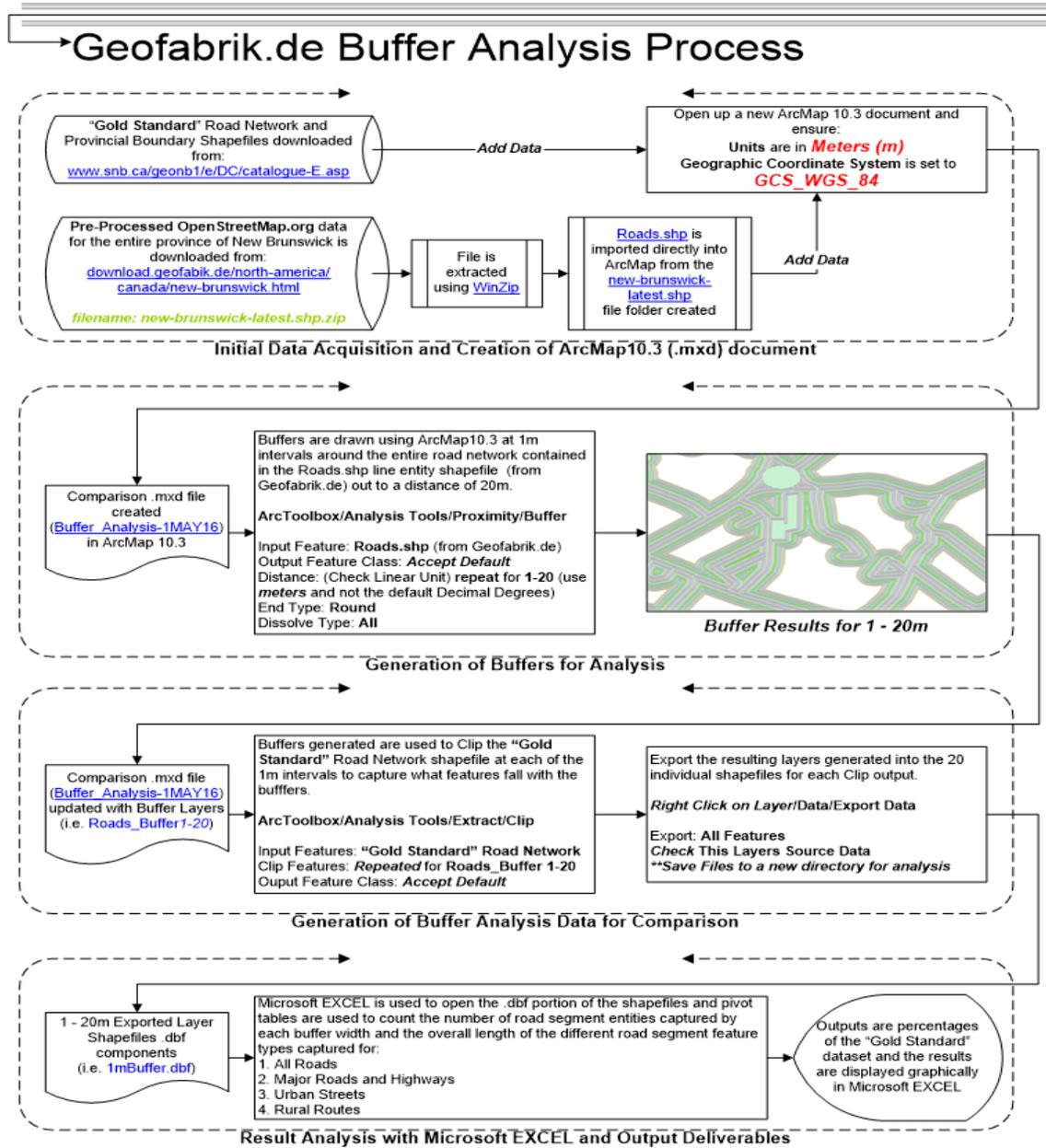


Figure 3.9 – Spatial Variance Analysis Method used in reviewing .osm data from Geofabrik.de

The following is a general overview of the steps used to perform the spatial variance analysis between the data provided by Geofabrik.de and the NBRN provided by GeoNB:

1. The OSM data for New Brunswick and the NBRN data were added to a new ESRI ArcMap map document.
2. Buffers at 1m intervals⁷ were added around the OSM highway_Line feature class for distances of 1 – 20m using the Buffer Analysis Tool in ESRI ArcMap toolbox (see Figure 3.10).



Figure 3.10 – 1m interval buffers extended out from the Geobrik.de OSM data roadways

3. The NBRN was clipped using the buffers generated for each of the 1 – 20m intervals.
4. The initial length and number of roadways contained within the NBRN dataset was compared to the roadways contained within each of the interval clipped files to determine a percentage of overlap and volume of roadways captured between the datasets at each interval distance using the relationship:

$$C_{Routes} = \frac{\sum Trusted\ Dataset_{Clipped\ Length}}{\sum Trusted\ Dataset\ Total\ Length} \times 100$$

⁷ The buffers were generated individually due to the fact that the current Multiple Ring Buffer tool in the ArcMap toolbox would crash when used on such a complex network dataset

5. The attribute tables (.dbf files) for the clipped datasets were opened using Microsoft EXCEL, and the entities were broken into the classifications of Major Roads, Urban, and Rural based on the labelling conventions used by GeoNB;
6. The entity lengths as well as the numbers of entities captured in different classifications were summed, and graphed, to show the mean offset distance between the two types of data at the individual buffer intervals (Figure 3.11).

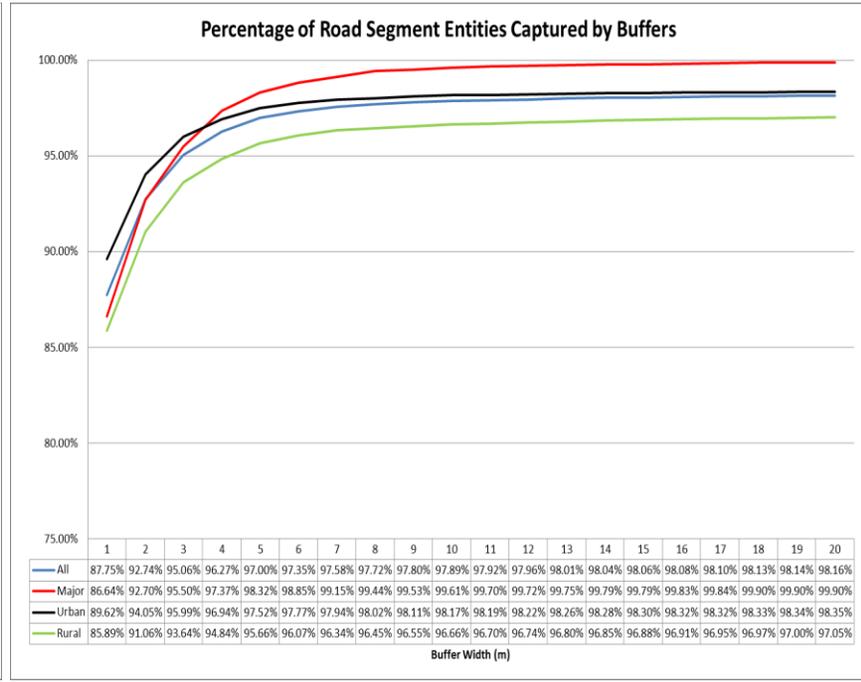
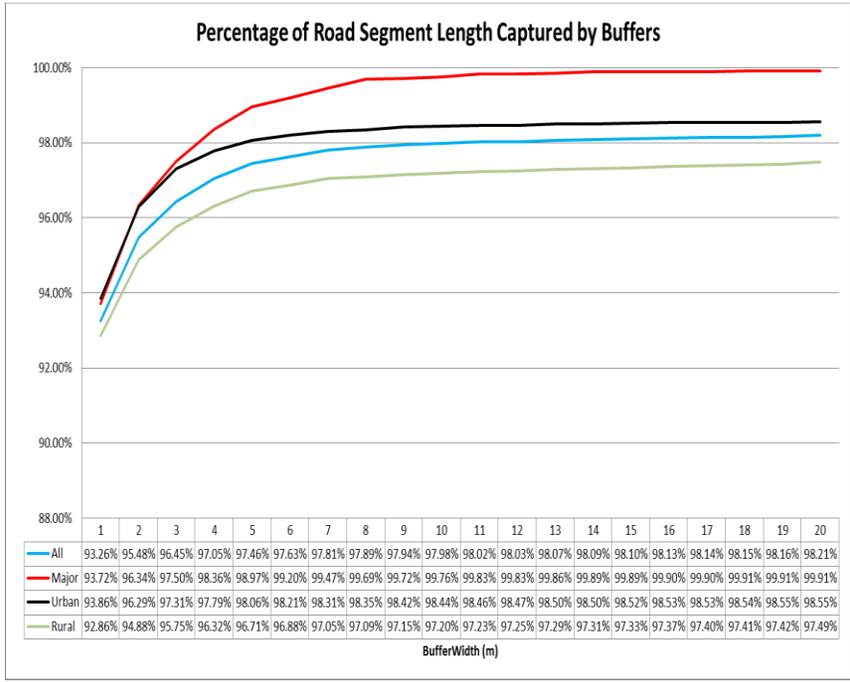


Figure 3.11 – Geofabrik.de Spatial Variance from the NBRN Analysis Results

When a more comprehensive review of the OSM data used for this analysis was conducted, it was noted that the pre-processing conducted by Geofabrik.de removed most of the attribute data in the conversion of the raw .osm data into shapefile format. Therefore, while, a proof of concept for analysis was achieved, only spatial variance could be analyzed and this process was deemed as only appropriate for expedient comparison of the spatial aspects of the datasets rather than the more in depth process developed in the next portion of this research.

3.4.2.2 Spatial Accuracy Analysis of OSM Data Provided by the Overpass API through ESRI ArcMap

The workflow was further refined to include factors such as the ability to target a specific area of interest rather than a limited pre-processed regional breakdown, and to ensure that the OSM data used included its attribute fields such as source, UserID, timestamp, and version number. The ESRI OSM toolbox add-on was reviewed and through trial and error, the parameters of the final process workflow were formed. The missing factors identified above were addressed and the resulting process could theoretically be automated should it be adopted into CAF doctrine for the use of OSM data.

A fictional scenario (dubbed “Operation EASTERN LYRE”) was designed to test the fine-tuned workflow based on the CAF doctrinal approach to the issuance of orders to a forward deployed GST. The one page document, depicted in Figure 3.12 and enlarged at Annex B, provided the specific details which would potentially be given to the GST in order for it to employ OSM data as their primary data source when deploying on a natural disaster response due to flooding in New Brunswick, or more specifically the Saint John River flood plain. These orders were devised under the premise that in some situations,

Thesis Process (XAPI) Trial Run Based on Fictional Operation EASTERN LYRE

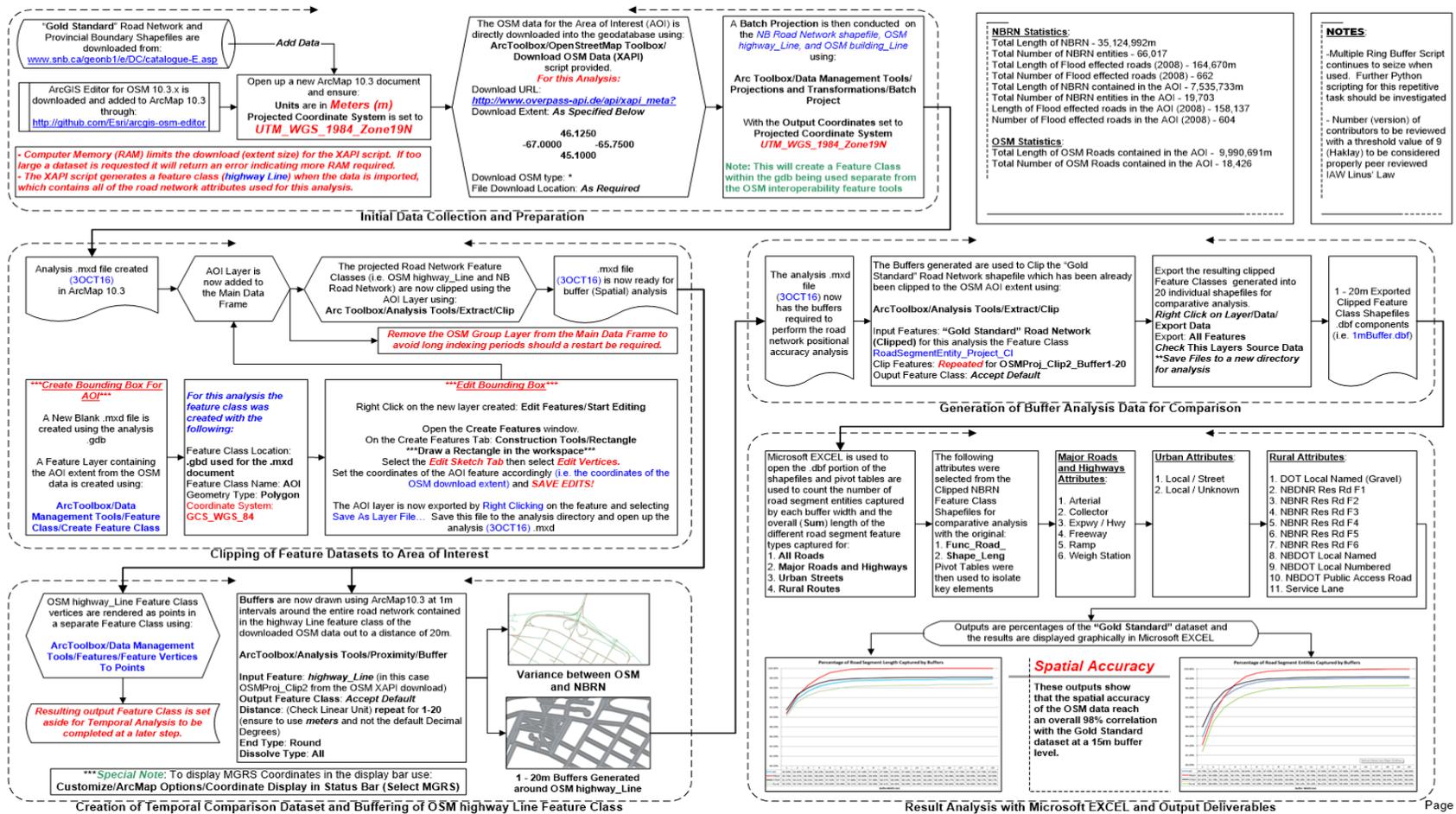


Figure 3.13 – Process workflow for utilization of raw OSM data in spatial variance analysis

The process steps utilized and related outputs for this analysis were as follows:

1. The analysis data was acquired from two sources:
 - a. The OSM data for the flood AOI (from http://www.overpass-api.de/api/xapi_meta) were directly added to the analysis map document using the bounding parameters in laid out Figure 3.14 , and

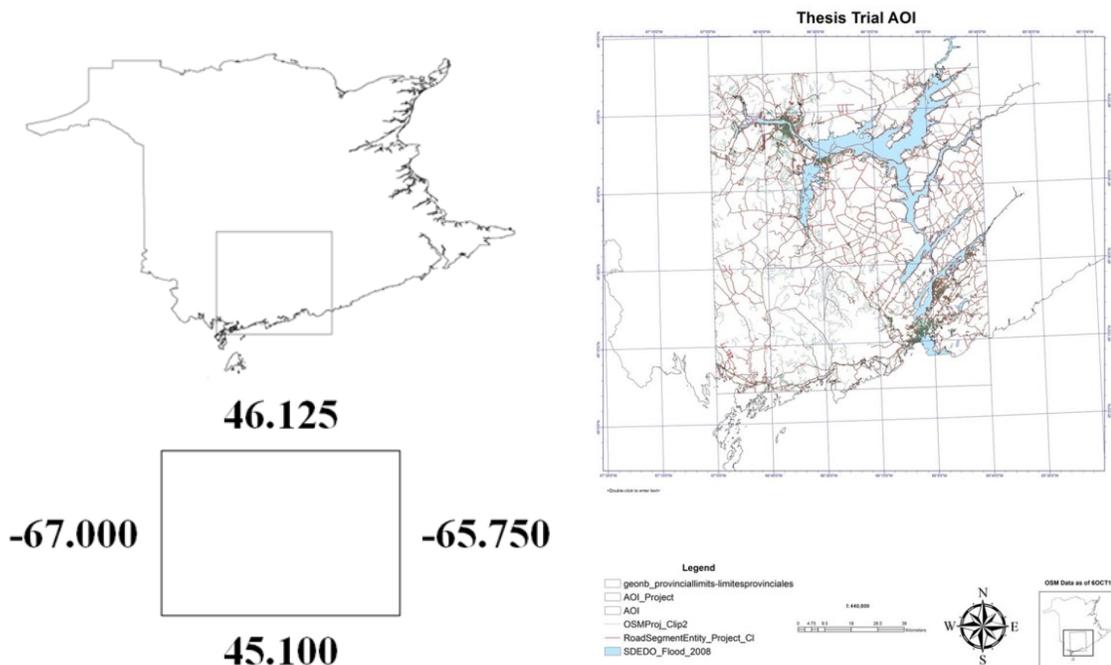


Figure 3.14 – AOI utilized for Thesis Process Workflow Testing

- b. The NBRN data provided by GeoNB (as depicted in Figure 3.14 and enlarged at Annex B, Appendix 3) were then added, and re-projected into WGS 84.
2. Both of the datasets were then clipped to the geographical boundaries of NB as well as the bounding box to ensure that the spatial variance analysis would be formed around the same geographic parameters.

3. A copy of the imported OSM dataset from the Overpass-API for the bounded region was then made and set aside for a later step in the analysis process.
4. In the same fashion as Step 2 of Section 3.4.2.1, buffers at 1m intervals⁸ were then individually drawn around the OSM highway_Line feature class at 1m intervals for distances of 1 – 20m.
5. The buffers generated were then used to clip the bounded NBRN roadways. The initial length of roadways contained within the clipped NBRN dataset were then compared to the length of the roadways contained within each of the 1m interval clipped files to determine a percentage of overlap between the datasets at a certain distance using the relationship:

$$C_{Routes} = \frac{\sum NBRN_{Clipped\ Length}}{\sum NBRN_{Total\ Length}} \times 100$$

6. The individual clipped datasets were then reviewed using Microsoft EXCEL;
 - a. The entities were sorted into the classifications of Major Roads, Urban, and Rural based on the labelling conventions used by GeoNB;
 - b. The overall total of entity lengths as well as the actual entities captured and their different classifications were summated, and
 - c. The values were then graphically depicted to determine the approximate offset distance between the two types of data at the individual buffer intervals (Figure 3.15)

⁸ The Multiple Ring Buffer tool also encountered difficulties when trying to apply it to the raw highway_Line network dataset in this OSM data context

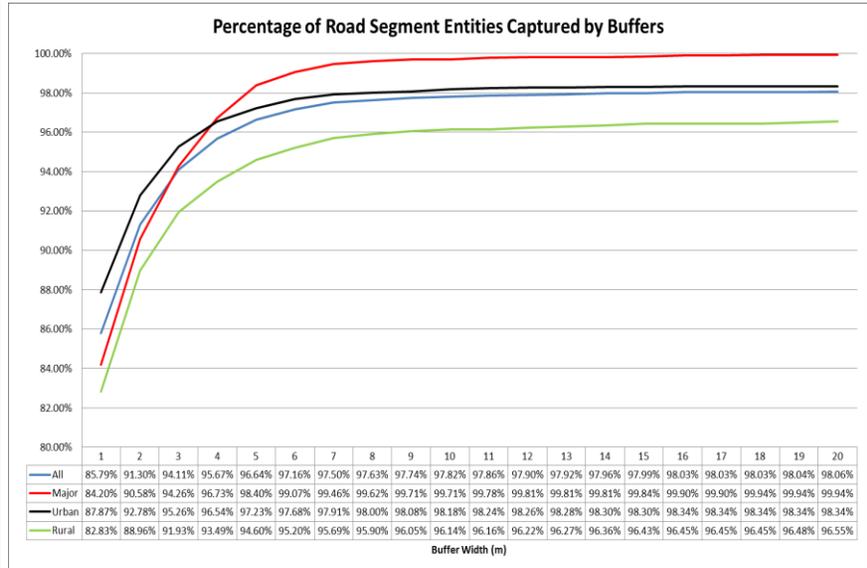
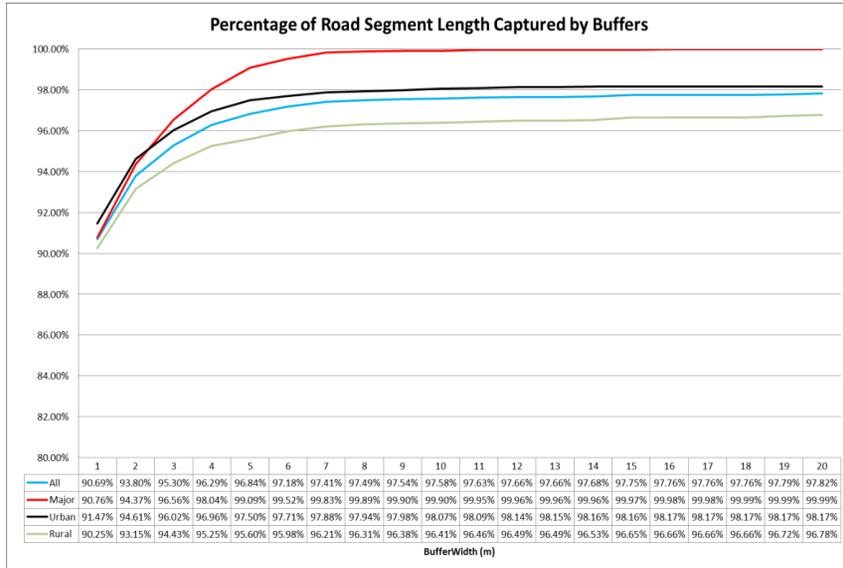


Figure 3.15 – Spatial variance of Overpass data to that of the NBRN analysis results

3.4.2.3 Temporal and Version Number Spatial Distribution Analysis of OSM Data

The data set aside in Step 3 of Section 3.4.2.2 were used to review the spatial distribution of the Temporal and Version Number attributes contained within the .osm data. For this, the process detailed in Figure 3.19 (enlarged at Annex B, Appendix 2) was employed. The standard OSM dataset does not include elevation values; therefore the “z-value” was used to illustrate various other attributes to review the spatial distribution of the data. The detailed process in Figure 3.19 remains the same for both cases; however, when reviewing the version number attribute, the timestamp used as the z-value is replaced by the version number attribute within the .osm data.

Rendering of Temporal Differences Found in the OSM Dataset

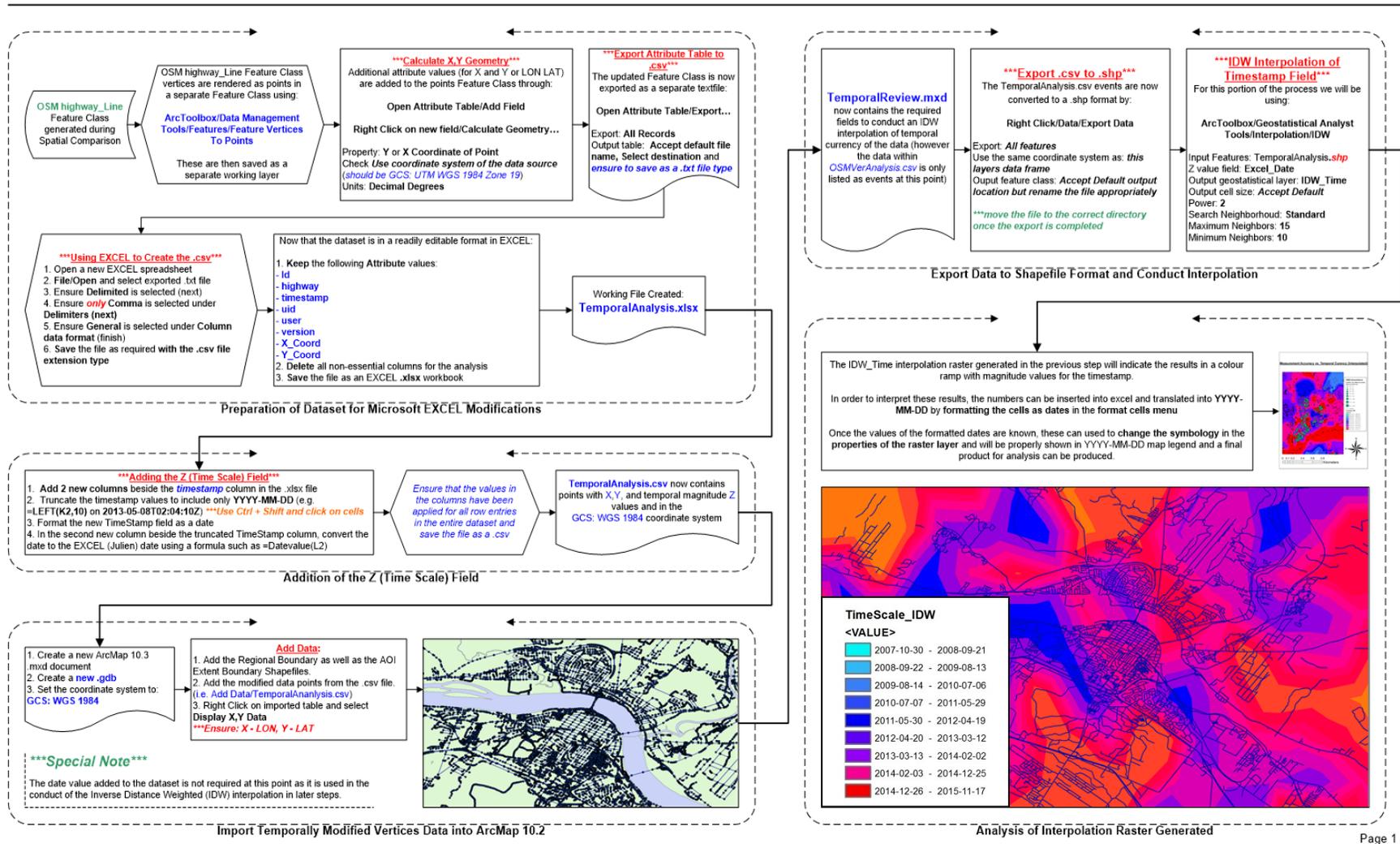


Figure 3.16 – Process Workflow for Temporal Review of Contributions to the OSM dataset

In order to spatially review the temporal attributes of the OSM contributions, the following approach was employed:

1. The highway_Line OSM data for the linear features within the AOI were rendered as points utilizing the feature vertices to points tool in the ArcMap toolbox;
2. The attribute table of the feature class created was then populated with the X and Y coordinates of the points and exported to a .csv format for review in Microsoft EXCEL.
3. The unnecessary attribute fields were removed and the timestamps were formatted into integer values which were used in an Inverse Distance Weighted (IDW) spatial interpolation:
 - a. In the case of temporal review, the timestamps were trimmed to a daily format and then converted to their equivalent [Microsoft EXCEL date](#) in order to provide an integer value which could be used as a z-value for the IDW interpolation.
 - b. For the spatial distribution analysis of the OSM Version Number the Version Number was unformatted and used directly.
4. The next step was to import the edited data back into ArcMap. As this data was initially extracted from a dataset utilizing the WGS 84 coordinate system, no further re-projection steps were required.
5. At this point the IDW Interpolation was performed using the Timestamp and Version Number to yield the outputs seen in Figure 3.17 (enlarged at Annex B, Appendices 3 and 4)

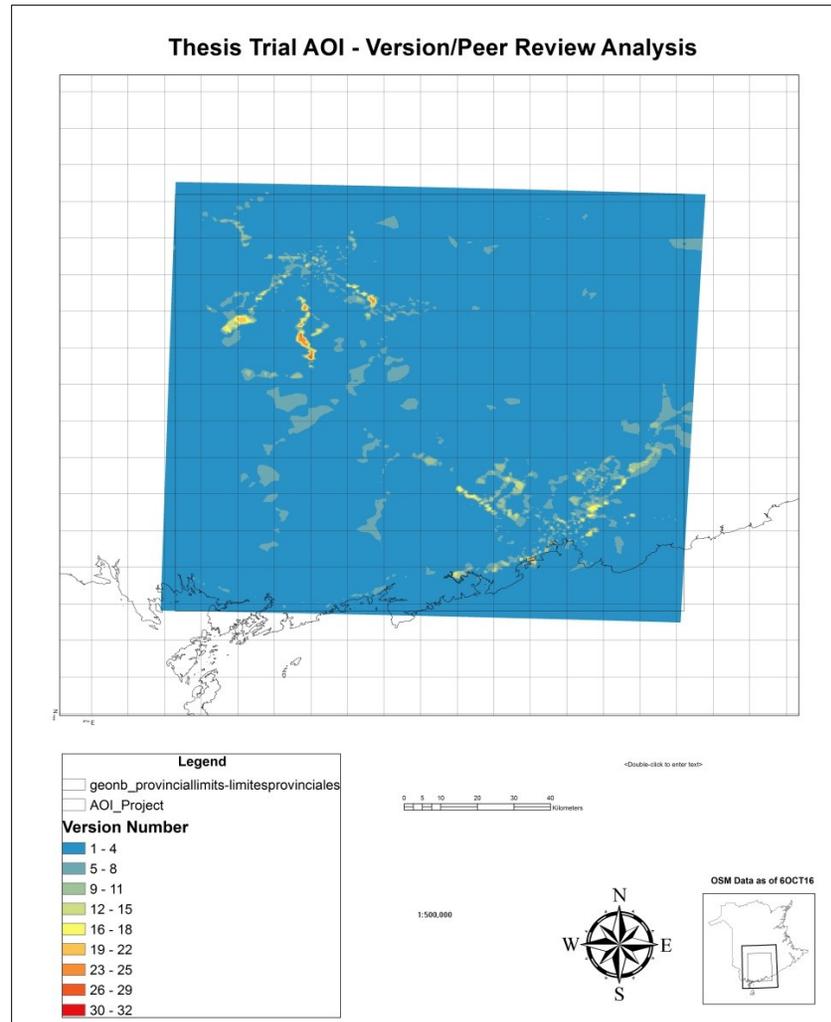
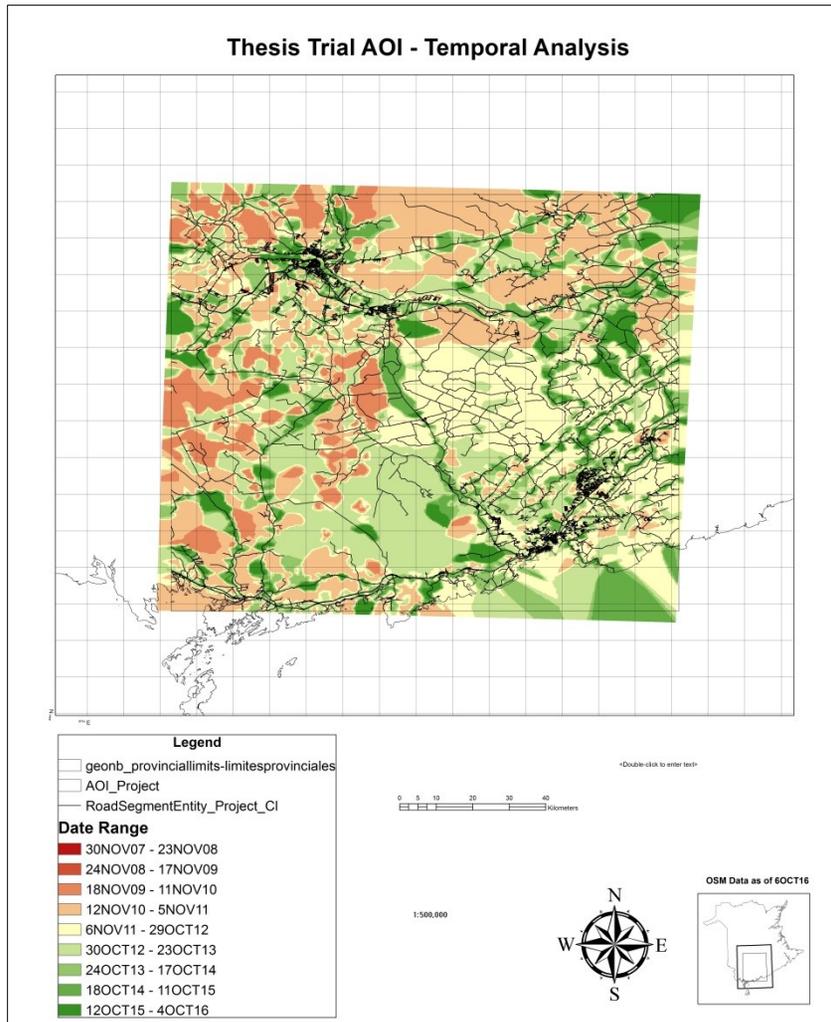


Figure 3.17 – IDW interpolation results for Timestamps and Version Numbers within the thesis test scenario, New Brunswick AOI

The outputs of this process could now be used to determine what areas should be prioritized for further data collection and which areas of the AOI contained data which has been reviewed and updated the most making it more likely to be a credible contribution source.

3.5 Thematic Analysis

3.5.1 A Simple Example of Thematic Comparison

In looking to examine the thematic accuracy of the .osm data in contrast to a trusted dataset, several shortcomings were encountered. As an example of this, a simple comparison of buildings is conducted (Figure 3.21).



Figure 3.18 – Carleton University athletic complex, OSM data vs. CanVec+ data

In Figure 3.18, the OSM buildings are depicted as the light blue polygons and CanVec+ buildings depicted by the black outlines. Although there are obvious

differences in the level of spatial detail captured by the two datasets, the focus of this comparison is in regards to the attributional information related to them. First, the attribute table for the OSM buildings has 74 unique attribute fields (Table 3-6), whereas the CanVec+ attribute table (Table 3-7) has only 8 unique attribute fields.

Table 3-6 – OSM Attribute Fields for the buildings in Figure 3.21

OSM Attribute Fields For The Carleton Athletic Center Buildings					
OSM Attribute Field	AG/UG⁹	OSM Attribute Field	AG/UG	OSM Attribute Field	AG/UG
FID	AG	osmvisible	AG	osm_foot	UG
highway	UG	osmversion	AG	osm_highwa	UG
building	UG	osmchanges	AG	osm_landus	UG
natural	UG	osmtime	AG	osm_leisur	UG
waterway	UG	osmSupport	AG	osm_lit	UG
amenity	UG	SHAPE_Leng	AG	osm_lock	UG
landuse	UG	SHAPE_Area	AG	osm_lock_n	UG
place	UG	osm_access	UG	osm_name	UG
railway	UG	osm_addr_5	UG	osm_natura	UG
boundary	UG	osm_addr_6	UG	osm_note	UG
power	UG	osm_addr_7	UG	osm_operat	UG
leisure	UG	osm_addr_8	UG	osm_park_r	UG
man_made	UG	osm_addr_9	UG	osm_parkin	UG
shop	UG	osm_amenit	UG	osm_phone	UG
tourism	UG	osm_area	UG	osm_source	UG
route	UG	osm_barrie	UG	osm_sport	UG
historic	UG	osm_bicycl	UG	osm_surfac	UG
aeroway	UG	osm_boat	UG	osm_water	UG
aerialway	UG	osm_buildi	UG	osm_waterw	UG
barrier	UG	osm_buil_1	UG	osm_websit	UG
military	UG	osm_capaci	UG	osm_wheelc	UG
geological	UG	osm_capa_1	UG	osm_whee_1	UG
OSMID	AG	osm_courts	UG	osm_wikip	UG
osmuser	AG	osm_create	UG	Area	AG
osmuid	AG	osm_fee	UG		

⁹ AG – Attribute is Automatically Generated by the OSM contribution process
 UG – Attribute is User Generated based on the input tags supplied by the user

Table 3-7 – CanVEC+ Polygon Attributes

CanVec+ Attribute Table							
FID ¹⁰	code	valdate	id	accuracy	theme	orientation	function
4	2010012	1995	82883c9301904dc58a9167a8e1bb1090	26	BS	0	-1
9	2010012	1995	e53f5fe68ac04afeb975c1623b9065ed	26	BS	0	-1

A closer look at the automatically generated attributes within the OSM dataset for the buildings in Figure 3.21 shows that the attribute fields covered in Table 3-8 are properly filled out with their integer and string values as outlined in the OSM common attributes at <http://wiki.openstreetmap.org/wiki/Elements>.

Table 3-8 – Automatically Generated OSM Attributes

Examples of the Automatically Generated Attributes found in the OSM Attribute Table							
FID	OSMID	osmuser	osmuid	osmversion	osmchanges	osmtimesta	Area
13	68999459	Rps333	596047	6	23222174	2014-06-27 0:00	1801.81
16	130299326	Greg Franks	3238084	4	34539155	2015-10-09 0:00	5429.17
17	130299327	WestCoastMike	1793499	2	18677668	2013-11-02 0:00	4552.22
18	130329216	mcwetboy	223056	1	9326999	2011-09-17 0:00	612.721
20	130330555	Johnwhelan	186592	3	27542778	2014-12-17 0:00	8539.17
44	130299329	Greg Franks	3238084	5	34984346	2015-10-30 0:00	4618.36
63	290034613	Rps333	596047	1	23222174	2014-06-27 0:00	99.4176

In Sections 3.3 and 3.4, methods for analyzing the attributes in Table 3-8, with regards to this study’s interest, have already been developed and as such these types of attributes will not be considered further here. Instead, a review of the User Generated attributes in Table 3-9 is conducted.

¹⁰ Feature Identification (FID) – automatically generated attribute from the GIS data extraction process

From a thematic perspective, the number of possible feature attributes in the OSM data has the potential to provide a richer source of information; however, a closer examination of the OSM attribute table highlights several factors relating to the lack of standardization found in the XML tagging structure of OSM data. One major factor is caused by the standard methods used for converting the OSM data structure into an ingestible GIS shapefile format. If no filter is applied during the conversion process, the thematic data available becomes quite heterogeneous and difficult to analyze using standard means. The ESRI tools trialed in this research did allow for the selection of specific attribute fields to be converted during the data import; however, the attribute fields themselves were often missing several values for each of the features covered in the attribute table. Table 3-9 exemplifies the heterogeneity of the thematic data encountered when analysis of the user generated fields present in the Carleton University athletic complex OSM dataset were attempted.

Table 3-9 – User Generated OSM Attributes

Examples of the User Generated Attributes found in the OSM Attribute Table						
FID	building	osm_addr_6	osm_amenit	osm_buildi	osm_leisur	osm_name
13	yes	Gymnasium		yes	sports_centre	Gymnasium
16	yes	Athletics		yes	sports_centre	Athletics
17	yes	Fieldhouse		yes	sports_centre	Fieldhouse
18	yes		kindergarten	yes		Colonel By Child Care Centre
20	yes			yes		Carleton Ice House
44	university	Alumni Hall		university		Alumni Hall
63	commercial			commercial		Walkway

It is at this point the lack of tagging standardization within the OSM project becomes most evident. Table 3-9 provides multiple examples of incomplete/incorrect information

and gaps which were found for the same features in Table 3-8. For example the attribute field “building” containing “yes” when used to describe the Gymnasium provides little to no thematic value. For the attribute `osm_name`, the OSM dataset did provide a richer description of the feature, but only when this field was properly completed by the contributor, which was not the case for the rest of the structures in the region. Upon visual review of Figure 3.18 it is also evident that the CanVec+ data is also quite out of date. For example, buildings such as the Field House, Alumni Hall/The Raven’s Nest, or the Ice House (commissioned in 2005) don’t exist in the CanVec+ dataset, so the data collection for structures in the CanVec+ dataset must pre-date 2005. At first glance the `osm_name` field in Table 3-9 would serve as a thematic advantage over the CanVec+ data reviewed (Table 3-7) which contained a very complicated string ID as its descriptor. Although the CanVec+ data may be out of date and only have contained 8 attribute fields, these fields were complete for all of the features in the dataset, and a review of the data documentation clearly shows that a standardized naming convention is applied through the unique ID for which a more detailed description is held in another information repository. It is therefore from these fundamentally different object identification processes, and the temporal discrepancies found between the two datasets, that clearly defining a thematic commonality for the basis of comparison becomes a qualitative matter, based on the data consumers’ requirements.

3.5.2 Thematic Analysis for the New Brunswick AOI

For an understanding of the thematic accuracy of the thesis test scenario OSM dataset, a very basic form of analysis was performed. This consisted of manually cross referencing the locations of key disaster management buildings such as those deemed as

the “Necessities of Life” (Zheng and Zheng, 2014) between Google Maps© and the OSM data for the AOI. Similarly to the CanVec+ dataset, the thematic attributes of the structures in the AOI were not directly available through GeoNB; instead, only a field listed as ID or Property Identification (PID)¹¹ was provided. Therefore cross-referencing between the OSM and Google Maps© datasets was performed. This process often encountered missing, incomplete, or incorrectly named attributes, similar to those in Table 3-9, making the generation of any usable metrics or statistics impossible. As discussed though, this process also required an interpretive approach rather than one which could be automated and was henceforth ceased as a viable option.

This form of analysis was intentionally left out of the final process workflows developed due to the fact that a trusted and reliable dataset which contained rich thematic attributes would need to already exist for a comparison to the OSM dataset to be carried out. As one of the primary foci of this research was to determine whether or not the OSM dataset could be used in an emergency when no other recent dataset existed, further thematic testing of the OSM data was not continued. If it is decided that further thematic analysis is to be conducted, it is recommended that the scope of the analysis be based on a spatial analysis process similar to that developed by Zheng and Zheng (2014) in which the centroids of known buildings constituting the “Necessities of Life” be compared to the locations of the centroids of those same buildings found in the OSM dataset. The conceptual model for this process is depicted in Figure 3.22.

¹¹ This deficiency could have been addressed with the 5 JUN 17 release of the province of New Brunswick’s Property Assessment dataset but the research phase of this thesis had already concluded

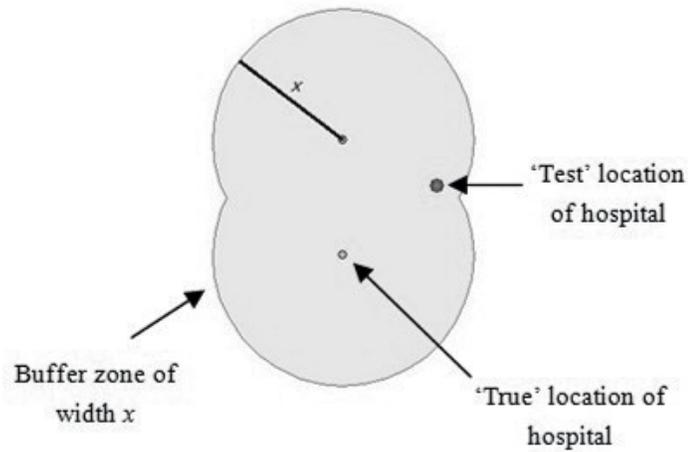


Figure 3.19 – Buffer Analysis of building (hospital) centroids between OSM (Test) data and Trusted (True) data (Zheng and Zheng, 2014)

This simple process would be relatively similar to the process workflow developed in this research already by substituting the building_Line centroid (point) features for the linear highway_Line features and buffering out to determine the spatial difference between the trusted known location of structures and that of the OSM buildings. As before, this process would require some knowledge of the true location of these structures, however, the volume of structures to be tested would be decreased significantly and the additional attribute information found in the OSM data could then be used to augment the existing knowledge. As a final note to this proposed analysis process, the ArcMap toolbox utilizes Planimetric Algorithms to calculate polygon (i.e. building) centroids. As the OSM data is delivered in a Geographic Coordinate System, it would need to be re-projected to a Projected Coordinate System before ArcMap could be used to carry out the spatial analysis.

3.6 Rapid OSM dataset Review

This section describes a rapid review process which could be used to determine whether or not OSM data use for a region should be considered at the onset of an aid

operation. Ideally, at the onset of any operation, a determination of data requirements would be made based on the operational geographic and temporal extent, which can drastically affect what data and information should be gathered. This can range from the procurement of proprietary Geo data and satellite imagery to standard collection by both land-based and air-borne reconnaissance elements. Since OSM data are a composite of all of these collection methods through its use of VGI and available open data sources, it is inevitable that errors based on linking the data collected together will lead to data consistency degradation within the dataset.

To combat these effects, the KeepRight.at initiative performs data consistency checks on the OSM dataset with an aim to identify common errors within the dataset such as missing/incorrect tags and incomplete/impossible geometry associated with features. Therefore, these data were evaluated as possible indicators of a relationship between volume of errors found and the number of contributions in the region. Unlike the Total Feature Count Quadtree generated by the VGI Analytics Framework (depicted in Figure 3.8), a quick review of the error count Quadtree generated in Figure 3.20 takes the total volume of errors found in the global dataset and splits it evenly into the different labelled regions. So, for example, region 20 has the same volume of errors as region 21 even though they vary greatly in size, which then provides a visual understanding of error density for the different regions depicted. For this research, the test scenario AOI studied falls within a region with a high to extreme density of errors and as such was also a consideration in determining the reliability of the data present.

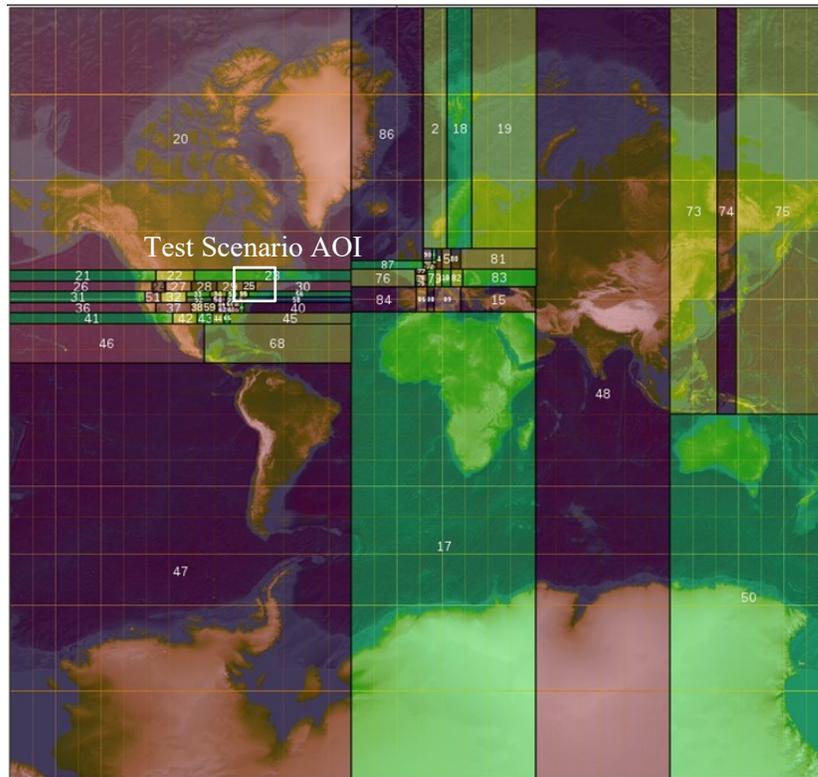


Figure 3.20 – KeepRight.at Error Count Breakdown Map

Figure 3.20 indicates that the error density in the AOI quadrant is quite high and therefore the wholeness of overall volume of entities, contained in the data for the AOI would also be heightened, however it should be remembered that errors included in the error count range from trivial (i.e. missing tags) to severe (i.e. presence of impossible geometry). When a review of the VGI Analytics Framework Quadtree representation in Figure 3.11 is made, it can clearly be seen that the contribution density in the AOI is also quite high and therefore counters the error density matter and the amount of risk associated with trusting the regional data is lowered. Abstractly, this relationship is outlined in Table 3-10.

Table 3-10 – Risk Relationship between Number of Errors and Volume of Contributions

Density of Errors in a Region	Density of Contributions in a Region	Risk Associated
High	High	Moderate/Low Increase
High	Moderate/Low	High Risk
Moderate/Low	High	Low Risk
Moderate/Low	Moderate/Low	Moderate/Low Increase

As a final note on rapid OSM dataset review, if a quick review of the error count breakdown map is all that is possible for determining whether or not to employ OSM data, the following should be considered:

1. Regions with relatively low error density would be an indicator that the contributions to the dataset are actually sparse and therefore the broadness of the regional OSM data would most likely be questionable.
2. Regions with a high density of errors would indicate large volumes of contributions and should be considered for inclusion or exclusive use, given the proper analysis.

3.7 Final notes on the development of the Idealized Process Workflow

During the development of the processes discussed in this chapter, several software tools were trialed in attempting to split the OSM Full Planet History file. This trial and error process was initially based around the constraint that it would be a process purely carried out in a MS-DOS/Windows data processing and handling environment. The different Open Source tools which were trialed are listed in Table 3.10. Initial efforts using the .osm.bz2 format of the Full Planet History were tested due to the fact that this format of the dataset contained the widest range of Open Source tools which could be used to process the XML format data in which it was structured.

Table 3-11 – MS-DOS/Windows based OSM processing tools trialed

Tool	Open Source Download Location	Notes:
JOSM - Java OpenStreetMap Editor (Edgewall Software, 2016)	https://josm.openstreetmap.de/	Extensible editor for OSM which runs on Java 8.
Osmconvert (Marqqs, 2017)	https://wiki.openstreetmap.org/wiki/Osmconvert	Convert and process OpenStreetMap files. Few functionalities, but processes much faster than Osmosis
Osmosis (Henderson, 2016)	https://wiki.openstreetmap.org/wiki/Osmosis	Command line Java application for processing OSM data
OSMembrane (Jarosch et al., 2017)	https://github.com/openstreetmap/OSMembrane	A GUI front-end for Osmosis

Each of the tools trialed had their merits; however in most cases the processing time to extract an AOI from the Full Planet History file in its compressed .osm.bz2 format was quite lengthy, ranging from about 8h to several days depending on the size and complexity of the AOI being extracted. The duration of this process was also exacerbated by the limited processing power of the computer being used, which was also an intentionally imposed constraint, arising from the concept of this process being employed by a forward deployed GST. In an attempt to reduce the processing time, trials were then conducted into decompressing the .osm.bz2 Full Planet History prior to extracting an AOI using various file decompression software packages such as WinZip™, the most commonly available for CAF purposes, and WinRAR©. These decompression software packages failed in most cases to decompress the data file as it would decompress beyond the scope of what the software package was designed for. A solution to this problem was finally found when the Open Source decompression software 7-Zip© was used. The resulting decompression of the .osm.bz2 file saw the initially downloaded file of approx. 86 GB grow into a decompressed volume of approx. 1.5 TB. Unfortunately due to the massive size of the decompressed dataset, all further efforts to extract data for targeted AOI failed to be accomplished as the processing

capabilities of the constrained hardware being utilized would encounter difficulties and stop mid-way through processing.

Since this review of the historical data is intended to be conducted on a routine basis prior to reacting to a natural disaster aid scenario, I decided that different platforms for analysis should be explored for their potential use on the more powerful processing machines available at MCE HQ. This decision would also allow for the consideration that non-Windows based tools could also be considered in the process development as they could be utilized on stand-alone workstations, which were not connected to the Defence Internal Network (DIN), rather than those under the information security constraints imposed on the equipment commonly used by a GST. Out of this consideration a LINUX/Ubuntu based approach was trialed. There have been many interesting tools developed within this Open Source environment which showed promise of providing manageable outputs. Although some difficulties were encountered in the set-up of the tools available, eventually the analysis tool developed by Rehl and Gröchenig (2016) was chosen and deemed suitable for the analysis purposes at hand.

The VGI Analytics Framework provides an integral capability contained in the ideal process workflow developed in that it allowed for rich data analysis of the contribution activities for time periods specifically surrounding a natural disaster. Although there was some difficulty encountered when trying to extract targeted AOI, this difficulty was countered by the decision to utilize regional .osh.pbf extracts which were readily available from Geofabrik.de. Unlike the .osm extracts also provided by Geofabrik.de, these .osh extracts contain their full attribute data which made them ideal for this process' analysis purpose. As described in Section 3.3, the outputs generated were delivered in a

.csv file format which was then imported back into a MS-DOS/Windows based environment to be analyzed using Microsoft EXCEL, since that is what is generally available to CAF personnel. The actual file size of the .csv outputs of the framework, are in fact quite small (i.e. < 1MB for most regions when a reasonable time period duration is specified) and could therefore be easily sent over classified or unclassified networks to forward GSTs, with limited bandwidth for the movement of data, should the raw analysis data be requested.

At this point the process workflow has been properly constrained to the conceivable limitations in the capabilities of forward deployed GSTs. Once the proper tools to conduct the analysis required for Stage 1 of the idealized process had been identified and tested for use, a review of the OSM dataset history surrounding natural disasters and the reaction of the OSM community could then be properly conducted. For Stages 2 and 3, a robust and repeatable process workflow had also been developed, and therefore final testing of the ideal process workflow outlined in Figure 2.9 was conducted using the thesis test scenario (section 3.4.2.2) as a final proof of concept.

Chapter 4: Results and Discussion

A review of the outputs for the historical analysis (Sections 3.3. and 4.1) of the Planet History (.osh.pbf) data and the spatial/temporal/version analysis and comparisons (Sections 3.4 and 4.2) of the current OSM (.osm) dataset are presented in this chapter. These include a review of contributor behaviours surrounding major earthquakes, the spatial accuracy of the OSM dataset when compared to a trusted standard, and the spatial distribution of the contributions contained within the OSM dataset with regards to their date of contribution (timestamp) and the current dataset's level of review/revision (version) for a targeted Area of Interest.

4.1 A Review of the Historical OSM Contributions Surrounding Natural Disasters

The outputs generated by the process developed in Section 3.3 allowed for a comprehensive review of the contributions to the OSM dataset, with a clear picture of how the mapping community reacted to a disaster situation such as those witnessed after the 2010 earthquake crisis in Haiti as depicted in Figure 4.1.

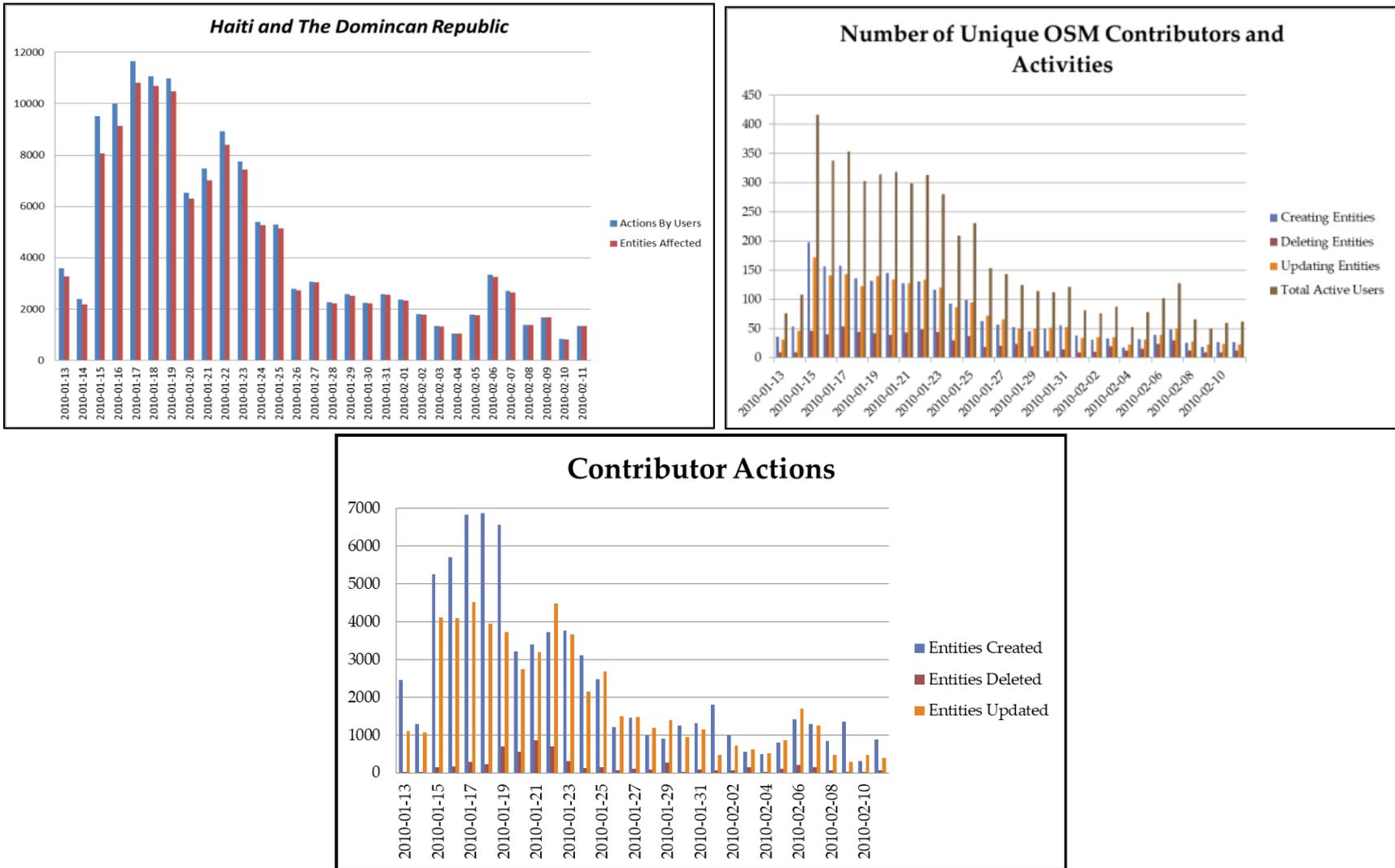


Figure 4.1 – OSM Contributions in the month following the 2010 Haiti earthquake

The surge behaviour witnessed could be attributed to many factors, and many are discussed in Neis et al. (2010), which reports that the larger VGI community was called into action by the HOT and several humanitarian gestures such as the donation of normally proprietary satellite imagery were donated to the cause of mapping the region. As depicted in the results at Annex A, this type of behaviour was also witnessed more recently in the country of Nepal; however, in both cases there was a delay for the spike in contribution activity to occur. This fact could have the most direct impact on understanding the overall research question of this thesis in that data for analysis may not yet exist when a natural disaster strikes a region. When a comprehensive review was conducted for all four of the study areas Figure 4.2, a clear bias is seen on the basis of the number of contributions associated with regular contributors and the source from which the data/information was generated.

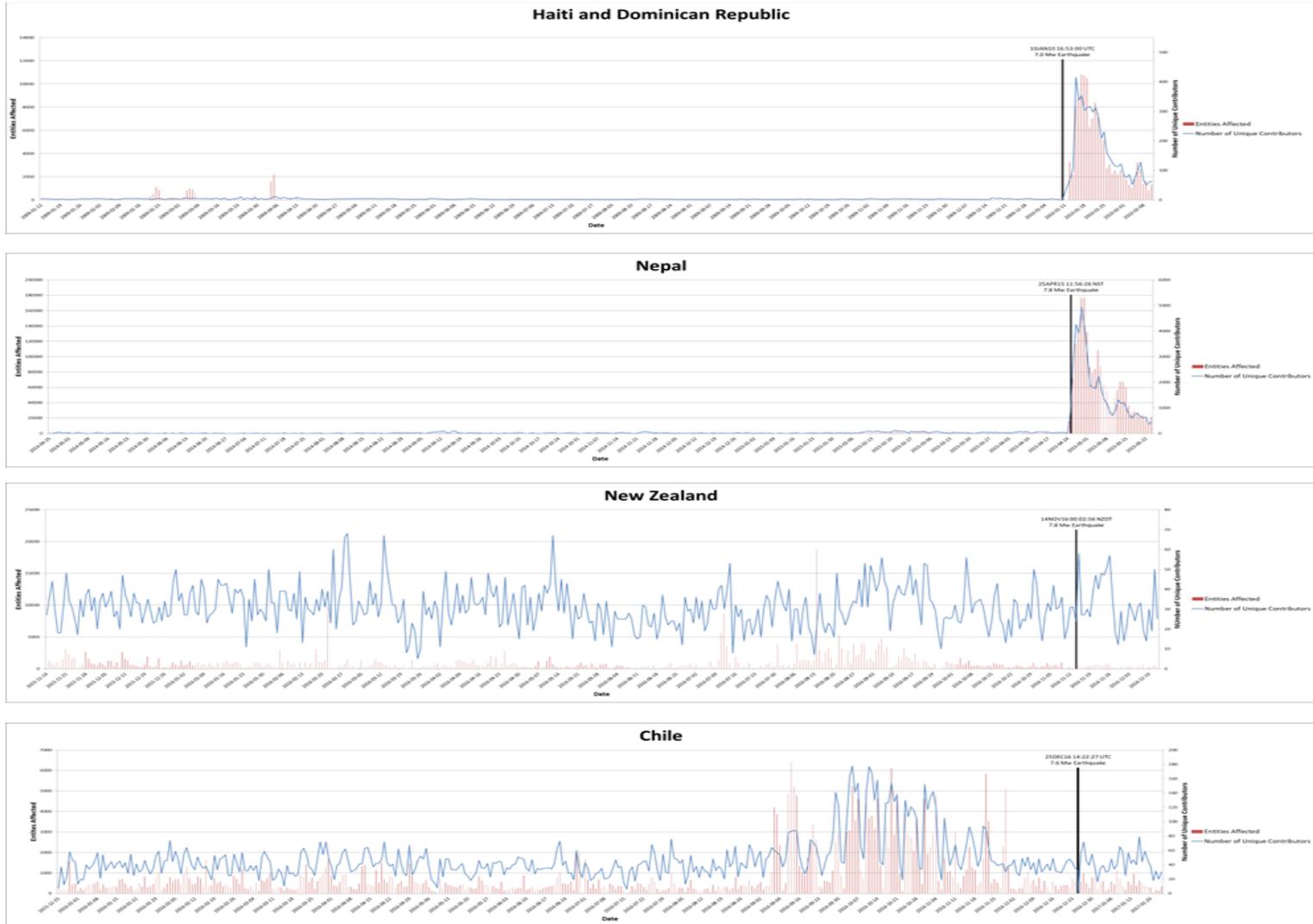


Figure 4.2 – Comparative review of the process results from Section 3.3(Annex A, Appendix 1)

The results clearly indicated that there is in fact a reactionary effect following a devastating natural disaster in cases such as those witnessed in Haiti and Nepal. In contrast to this, for the cases of Chile and New Zealand, the contributions to the dataset following the earthquakes remained relatively constant to the pre-event conditions as observed in the low level of activity seen in Annex A, Appendix 4.

These low activity rates following the earthquakes in Chile and New Zealand could be interpreted that due to the already robust open data sources available for the regions, which had already been bulk uploaded to the OSM dataset, was the direct cause for a lowered requirement for the regional data to be updated. It should also be noted that HOT was also not called into action and additional resources such as imagery were not provided as they would have been deemed unnecessary. If a macro perspective of the global OSM dataset is taken, it could be shown that there are indeed regional pockets where effects similar to the Index of Multiple Deprivation (Haklay,2008) can be observed. These effects could be viewed as part of the reasoning behind the varied contribution activities following each of the events but should not be viewed as the sole cause for trusting the reliability of a regions information and data.

Although there are clear indications and examples that the OSM dataset itself can be spatially trusted to be within an understood threshold (Ciepluch et al., 2010; Girres and Touya, 2010; Haklay, 2010; Mooney and Corcoran, 2012; Quinn, 2015; Zheng and Zheng, 2014; Zielstra and Zipf, 2010), the underlying concerns of its actual overall reliability remains a subject for consideration. One of the main concerns regarding the reliability of the dataset is the concept of “spoofing” by nefarious agencies. This process consists of contributions intentionally being made for the purposes of rendering a

region's data unusable or highly inaccurate. This type of activity has been witnessed historically throughout the OSM dataset for regions which are in turmoil such as the Ukraine. Although there can be no absolute measure of this type of activity, the basic tools available through the OSM project can be used to review and counter errors for the determination of a region's contribution reliability. In such cases where the concern is that spoofing may have occurred, an initial analysis of the dataset for the region could be conducted. This would involve a quick search using the *OpenStreetMap Analytics beta* platform or a more detailed review such as the process outlined in Section 3.3. The focus of this review would then be to observe any notable spikes in the contribution history, which could then be cross referenced with the OSM bulk data import catalogue at <https://wiki.openstreetmap.org/wiki/Import/Catalogue> to determine if the data contributions were generated from a reliable source. If a quick review of the results, for regions such as Chile and New Zealand, at Annex A is conducted, it is clear that there are several periods in which the number of contributors remained relatively constant; however, the overall volume of contributions spikes dramatically at different time periods leading to the provision of a timeframe for which a bulk upload of data may have occurred and the possibility of a spoofing input could be identified.

As a final consideration before conducting an analysis of the OSM data for a region, it should be understood that the OSMF is aware of these types of actions being carried out on the dataset. As discussed earlier, the OSMF not only acts as the repository for the OSM data itself, but also as its gatekeeper for the maintenance of the dataset's quality. It presently has tasked the Data Working Group (DWG) with detecting and stopping imports which do not comply with their pre-determined import guidelines as a form of

quality control. If an analysis is carried out and erroneous data are detected, the OSMF also maintains contacts which can be reached and are capable of reverting changes to the dataset. Although this does not enhance the instantaneous reliability of the dataset, it does allow for actions to be taken in order to rectify any issues detected through a recursive analysis cycle.

4.1.1 Understanding the Types of Contributions which have been made

Through the evolution of the OSM project from its initial inception to its current state, contributions and the methods in which contributors interact with the dataset have varied and improved widely. With this fact in mind, it is not hard to conceive that the attribute definitions initially set forth in the project are slowly becoming obsolete. Contribution frameworks such as the UN/WFP SDI-T have been conceptually applied to some regions within the dataset with great effect but have also caused increased heterogeneity to the overall atomic level operations (Table 3-3) conducted in the formation of the dataset over time. These variances in regional contribution efforts have led to large gaps between the regional data available (e.g. the OSM data for Haiti contains many more attribute fields than that of New Brunswick), and ultimately have caused the physical storage size of the dataset to expand disproportionately.

When an analysis of the dataset is conducted regarding the types of contributors, as outlined in Section 2.1.1 by Coleman et al. (2009), a clearer understanding that various contributors hold larger stakes in portions of the OSM dataset over others arises. This is most evidently seen in the varying degrees of contribution attribute labelling completion. In many cases, an Expert Authority who is trained in the use of GIS would most likely be responsible for the bulk uploads of open data into the OSM dataset whereas the casual

neophyte would compose the contributions which often lack key information such as the *source* attribute label. In Haklay et al. (2010) the process developed for testing the OSM dataset on targeted regions around London, U.K. suggested that roughly 13 active and unique contributors were required in order for the targeted areas to be considered spatially accurate and relatively complete. It should be considered though, that this study was also conducted during a time period when many open GIS data sources (globally) were just becoming available. Although the number of registered users of the OSM dataset has grown exponentially over the last few years, the 90:9:1 concept put forth by Neilson (2006) still holds true as evidenced in the data extracted for this thesis' research test scenario. In Figure 4.4 (Enlarged at Annex A, Appendix 4), a more comprehensive breakdown of contributor activities in the one month period following the major earthquakes studied is presented.

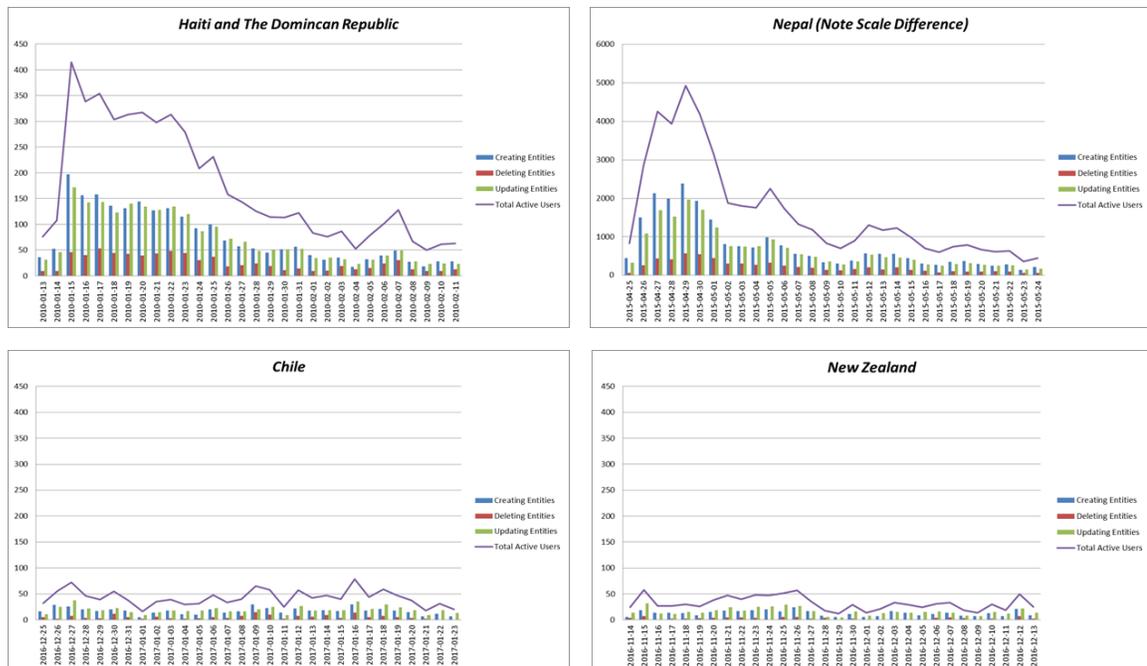


Figure 4.3 – OSM contributor activities for a one month period following a major earthquake

Figure 4.4 clearly depicts that for the cases of Haiti and Nepal, much of the contribution activity involves the creation and updating of entities in the region whereas in Chile and New Zealand, the activities seen are consistent with what would be considered regular maintenance of the dataset by contributors. Also interesting, the number of edits to the dataset remains relatively constant (approximately 60 per day) for Chile and New Zealand but spikes to numbers as high as 400 and 5000 per day for Haiti and Nepal respectively. As Haiti and Nepal did not have an open data infrastructure for their region for which bulk uploads of data could be made from, this behaviour indicates that the area was mapped by multiple independent contributors working together, either through mapping parties (Mooney and Corcoran, 2014) or interested amateurs with the humanitarian urge to help.

As a final note on the types and volume of contributions made to the dataset over time, consideration should always be given to the operational scope of use intended for the OSM data. In the case of using the most current dataset for mapping information a simple summative review of the *source*, *UserID*, and *timestamp* attributes of the dataset can provide vast indications as to the reliability of the data. For the *source* and *UserID* attributes, the application of a data filtering process would most likely provide good information on which users in the AOI are Super-Users (i.e. Expert Authorities) and where their contributions were commonly from.

4.1.2 Thesis Test Scenario Data Review

In order to emphasize how the presence of a reliable open data infrastructure could play a role in the contributor activity following a natural disaster, an analysis of the OSM data extracted for the AOI in New Brunswick was conducted. An examination of the

contribution activity surrounding the 1st of May 2008 was carried out using the VGI Analytics Framework with the outputs shown in Figure 4.5 (Enlarged at Annex A, Appendix 1).

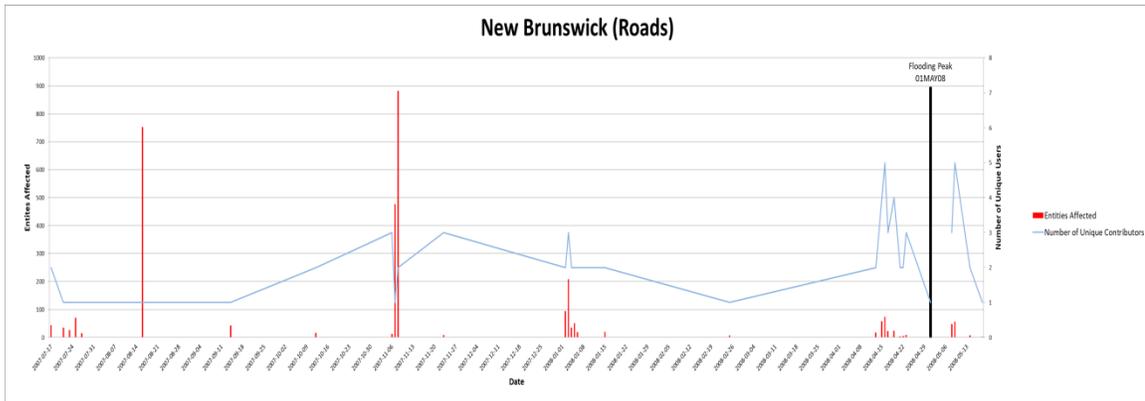


Figure 4.4 – OSM contribution activity for the province of New Brunswick surrounding the 2008 floods

The resulting outputs indicated a very low level of unique contributors were present in the region with spikes in activity which would be associated with bulk uploads by a single user. Due to the scope of this analysis method, a further understanding of the contributions currently held by the OSM dataset should be conducted. In this case the OSM bulk import catalogue indicated that there were in fact several bulk uploads of data from Geogratis.gc.ca in the late 2009 timeframe which explains the overall volume of contributors observed to be quite low and the volume of contributions in the region as spiking due to probable bulk uploads from these few contributors using other resources such as the uploading of multiple GPS tracks.

OpenStreetMap Analytics beta was employed to analyze the history of timestamps associated with road entities contained in the dataset for the province of New Brunswick as seen in Figure 4.5. The presence of very obvious spikes in the timestamps for the contribution activities, makes it clear that multiple bulk uploads have likely been conducted. In order to better understand this activity, analysis of the OSM data for the

AOI was then conducted where contributors to the regional data were ranked using the 12 providing greater than 1% of the road segment length contributions and the remaining 191 providing less than 1% each. The results of this analysis were then plotted as depicted in Figure 4.6, and broken down numerically in Table 4-1. With this breakdown of contribution volumes for the current dataset¹² (Figures 4.6 and 4.7) the effects of the 90:9:1 concept could be easily observed for the contribution percentages of the active contributors within the AOI and how the concept pertained to OSM contribution activities as having social media type behaviours.

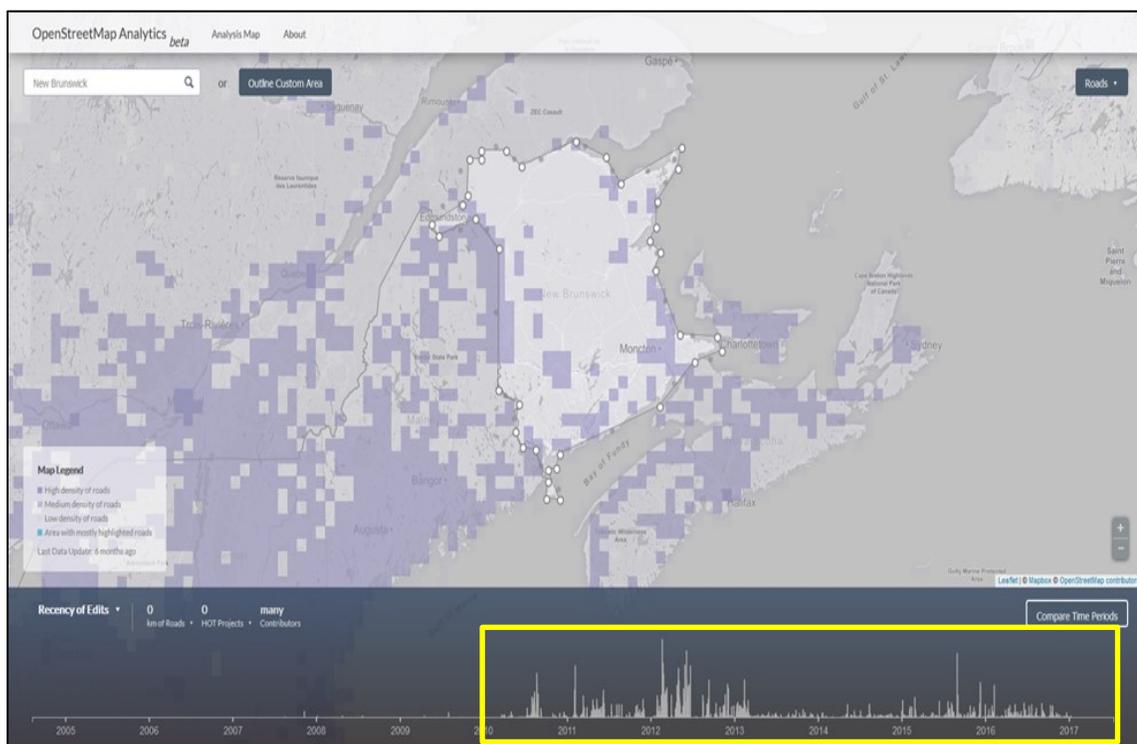


Figure 4.5 – Current Road Contribution timestamps for the province of New Brunswick

¹² The AOI data analyzed was from the 5OCT16 Thesis Test Scenario run through, extracted using the Section 3.4 XAPI process.

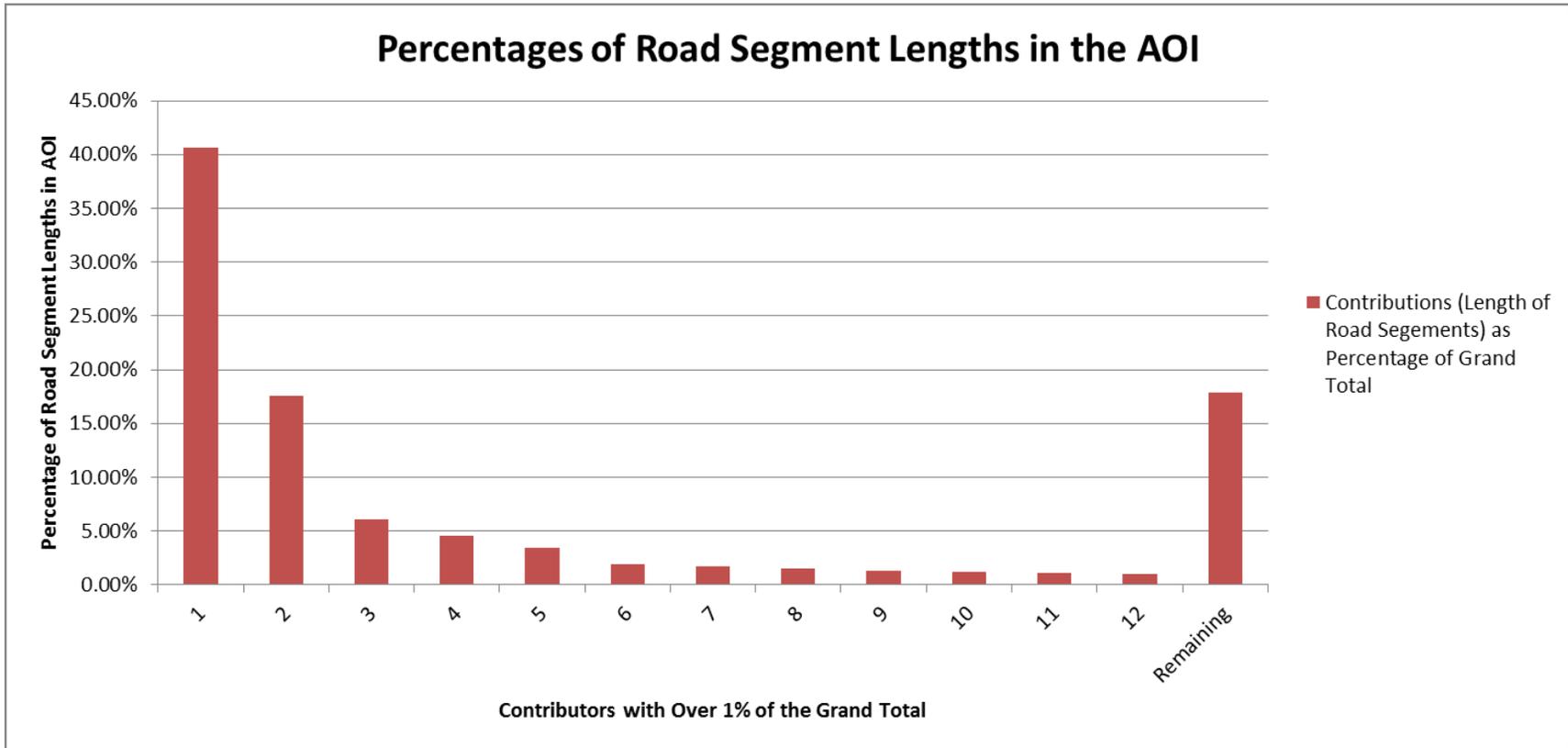


Figure 4.6 - AOI contributors and their volumes of contributions to the current OSM dataset

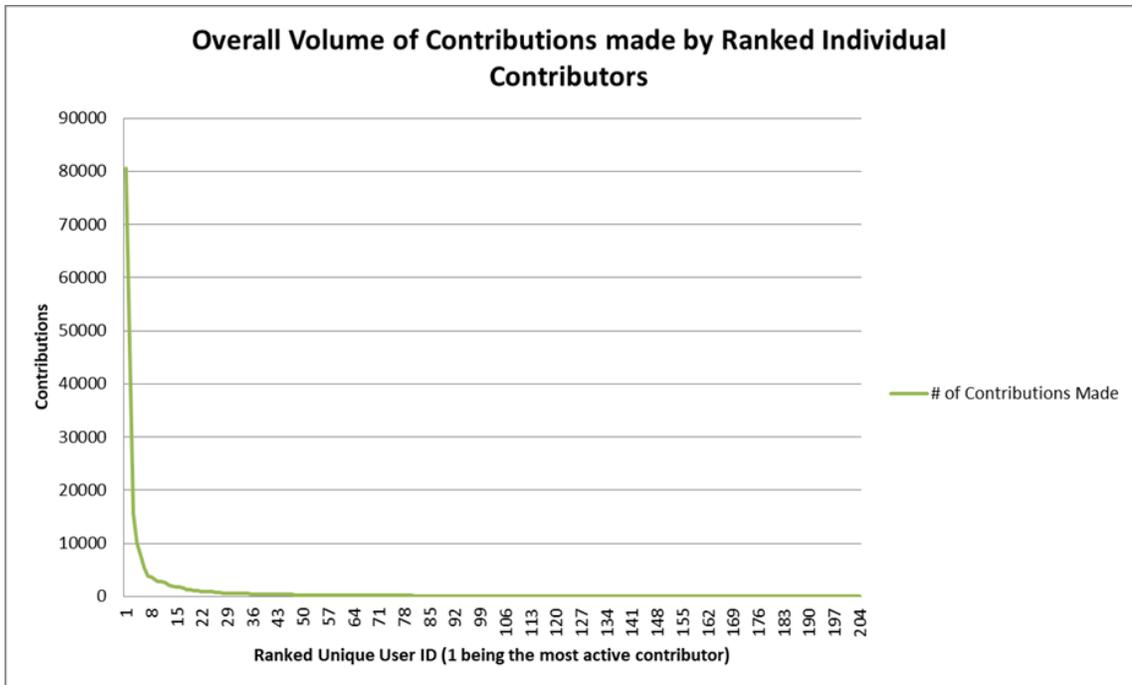


Figure 4.7 – Analysis of Unique Contributors (Ranked) to the OSM dataset for the Thesis Test Scenario Study Area and their Volume of Contributions Made

Table 4-1 – Breakdown of AOI Contributors and their Volumes of Contributions

Contributor Rank ¹³	Sum of Shape_Length (m) Attributes Contributed	Sum of Shape_Length attribute field as a Percentage of the Total found in the AOI highway_Line feature class
1	4,058,131.31	40.62%
2	1,760,171.63	17.62%
3	603,807.28	6.04%
4	458,803.10	4.59%
5	341,815.53	3.42%
6	191,221.22	1.91%
7	167,451.65	1.68%
8	154,750.02	1.55%
9	131,102.54	1.31%
10	120,481.77	1.21%
11	109,606.87	1.10%
12	103,553.58	1.04%
Remaining 191	1,789,795.36	17.91%

¹³ The Contributors were ranked arbitrarily at 1% due to the fact that contribution levels lower than this were mostly less than 1/10th of a percent and therefore the other 191 contributors in the AOI were lumped as a total sum ranking titled Remaining 191.

The actual data sources of the contributions themselves can also be extracted. Filtered and ranked current attributional data for the Thesis Test Scenario AOI are summarized by Figure 4.7 and Table 4-2. Figure 4.8 shows that the *sources* of the contributions, when actually included in the XML tagging of the data contributions, include large portions of data from Open Data Sources such as the GeoGratis.ca portal provided NRCan-CanVec datasets. Although this would not be considered a conclusive measure of the regions data reliability, it does indicate that both, multiple contributors were working from the same data source, and that in this case the data source was from a trusted professional entity. In Table 4-2, a breakdown of the percentages of contributions for each of the data sources is provided. As discussed in several works involving contribution activity (e.g. Kuutti 2006), even the highly active Expert Authorities in the region lowered their contribution activity when they felt that the entirety of the features represented in the dataset had been achieved.

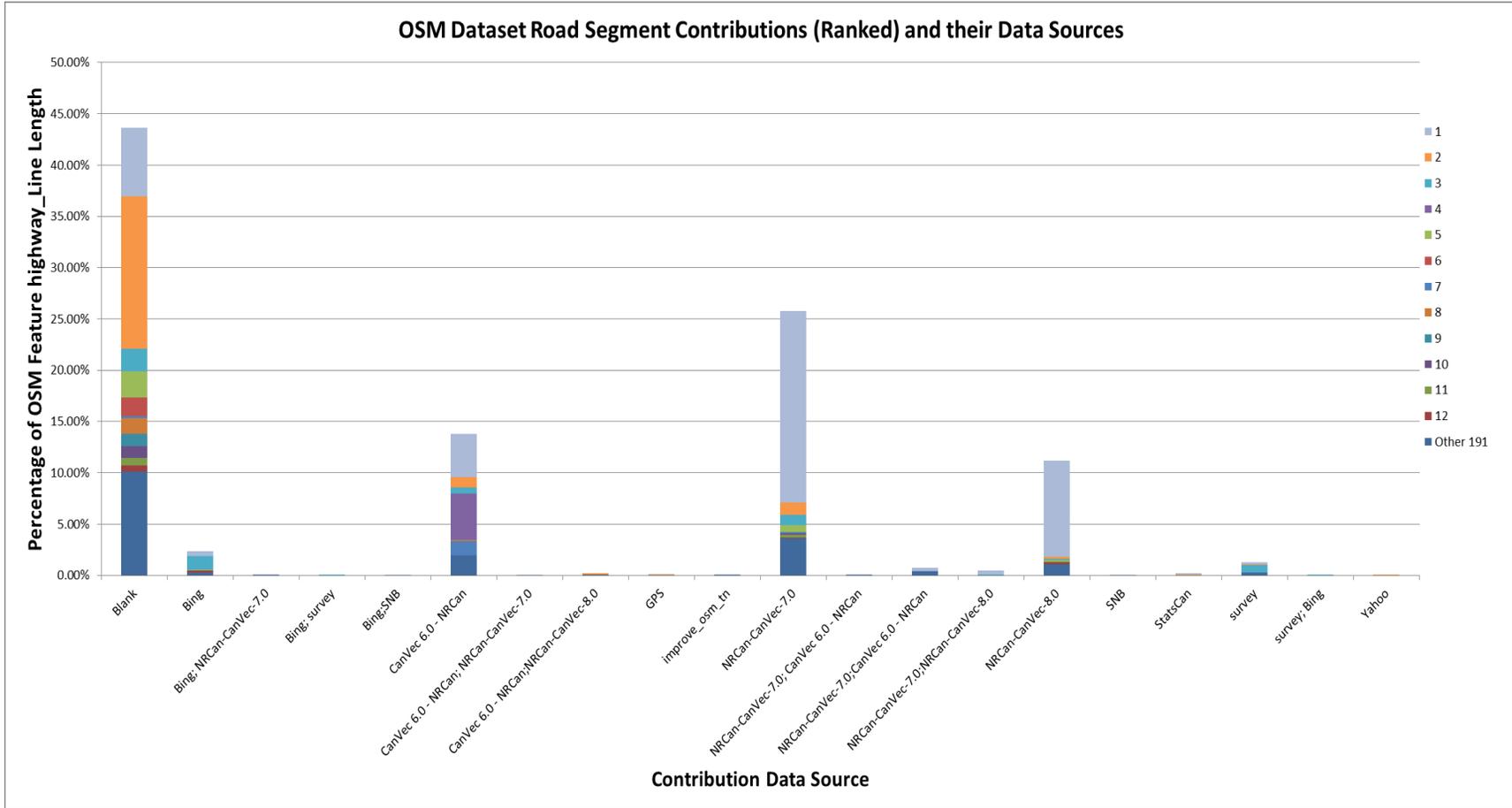


Figure 4.8 – Data sources in the AOI and the volumes of Contributions by Contributor

Table 4-2 – Detailed breakdown of the Contribution Data Sources for the Current AOI dataset

Contribution Data Source	Contributor Ranking												Other 191	Grand Total
	1	2	3	4	5	6	7	8	9	10	11	12		
Blank (No Source Provided)	6.7%	14.8%	2.2%	0.0%	2.5%	1.9%	0.1%	1.5%	1.3%	1.1%	0.8%	0.6%	10.1%	43.62%
Bing	0.4%	0.0%	1.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.2%	0.3%	2.32%
Bing; NRCan-CanVec-7.0	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.03%
Bing; survey	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.01%
Bing;SNB	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.08%
CanVec 6.0 - NRCan	4.3%	1.0%	0.6%	4.6%	0.1%	0.0%	1.3%	0.0%	0.0%	0.0%	0.0%	0.0%	1.9%	13.82%
CanVec 6.0 - NRCan; NRCan-CanVec-7.0	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.09%
CanVec 6.0 - NRCan;NRCan-CanVec-8.0	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.21%
GPS	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.13%
improve_osm_tn	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.03%
NRCan-CanVec-7.0	18.7%	1.2%	1.0%	0.0%	0.7%	0.0%	0.1%	0.0%	0.0%	0.1%	0.3%	0.0%	3.6%	25.75%
NRCan-CanVec-7.0; CanVec 6.0 - NRCan	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.01%
NRCan-CanVec-7.0;CanVec 6.0 - NRCan	0.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.4%	0.77%
NRCan-CanVec-7.0;NRCan-CanVec-8.0	0.4%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.46%
NRCan-CanVec-8.0	9.4%	0.3%	0.1%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.2%	1.1%	11.20%
SNB	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.00%
StatsCan	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.22%
survey	0.2%	0.1%	0.7%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.3%	1.26%
survey; Bing	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.01%
Yahoo	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.00%
Grand Total	40.62%	17.62%	6.04%	4.59%	3.42%	1.91%	1.68%	1.55%	1.31%	1.21%	1.10%	1.04%	17.91%	100.00%

Upon closer review of the results in Table 4-2, it was noted that the volume of contributions made by the top ranked contributor decreased in the *source* attribute field being left blank as more recent versions of the open datasets were released. This could be indicative of the level of comfort or process development which the contributor was developing with the OSM project over time, and would also most likely indicate a possible measure of reliability relating to the contributions made by that contributor.

As a final note on the contributors themselves, if one considers a sequential view of the results shown in Figures 4.4, 4.5, and 4.8 it is apparent that:

1. The results in Figure 4.4 pre-date the availability of Open Data from Geogratis.ca and GeoNB, hence the spikes in contributions to the OSM dataset for that period of time were probably due to interested amateurs conducting mapping activities in small pockets based on the low (2-3) number of active contributors.
2. The contribution patterns seen in Figure 4.6 were consistent with the launch of data.gc.ca/GeoGratis in March of 2011. The public roll-out of the CanVec series of datasets caused a spike in contribution volumes; however a lull in contributions occurred following the decision to convert the open data source into the CanVec+ format in 2013. This lowered level of contributions could also be attributed to the launch of the initial GeoNB NBRN dataset in December 2013, which could also have decreased the need to rely on the OSM project for any information not found in the often incomplete¹⁴ CanVec dataset.
3. Finally, the analysis presented in Figure 4.8 shows that approximately half of the OSM data for the region is attributed to data which is already several years older (i.e.

¹⁴ A clear example of this is shown in Figure 4.14

CanVec v.6.0,7.0, 8.0) than the data used in this analysis. Due to the fact that large portions of the contributions for the region came from only two unique contributors, this could be construed as indicating that they had abandoned any further work on the OSM project dataset for the region as they felt an appropriate level of completeness had been achieved.

Focussing on the actual contributions made, I tested whether, as discussed in Chapter 2, the version number of the entities within the dataset can be considered a marker of peer revision and process control. However, the dataset for the AOI does not support this idea. A common review of the version number of the contributions was conducted (Table 4-3), and clearly showed that over half of the contributions for the region are still only at the initial version or have had only one review/update (i.e. version 2). If these results are considered indicative of the level of peer review of OSM contributions for regions which have an available open data infrastructure, further analysis of the data should be conducted regarding the volume of contributions made from reliable or “Trusted” professional open data providers.

Such analysis was conducted for the AOI (see Figure 4.9 and Table 4-4) but it was again noted that over half of the contributions made for the region are devoid of the *source* attribute field being completed and thus counted as the Blank entries. On the other hand, approximately 1/3 of the contributions which have the *source* attribute field completed are actually sourced from trusted open data sources.

Table 4-3 – Volumes of AOI highway_Line Contributions and their Associated Version Numbers

OSM Version Number	Number of highway_Line contributions	Total Length (m)
1	9457	3,399,450.51
2	4618	2,536,209.08
3	1926	1,548,148.98
4	1005	821,966.31
5	494	436,388.62
6	323	327,334.06
7	200	180,287.06
8	97	117,792.21
9	83	124,226.98
10	58	119,097.65
11	28	54,908.72
12	24	48,849.28
13	23	40,292.55
14	16	33,820.94
15	8	16,721.74
16	15	28,291.25
17	8	16,263.00
18	9	11,181.05
19	5	17,403.17
20	4	15,177.25
21	5	32,090.57
22	6	7,792.93
23	2	1,990.69
24	2	297.40
25	2	17,007.84
26	1	92.49
27	1	21,210.05
28	1	437.54
31	2	973.32
33	2	8,935.69
54	1	6,052.95
Grand Total	18426	9,990,691.87

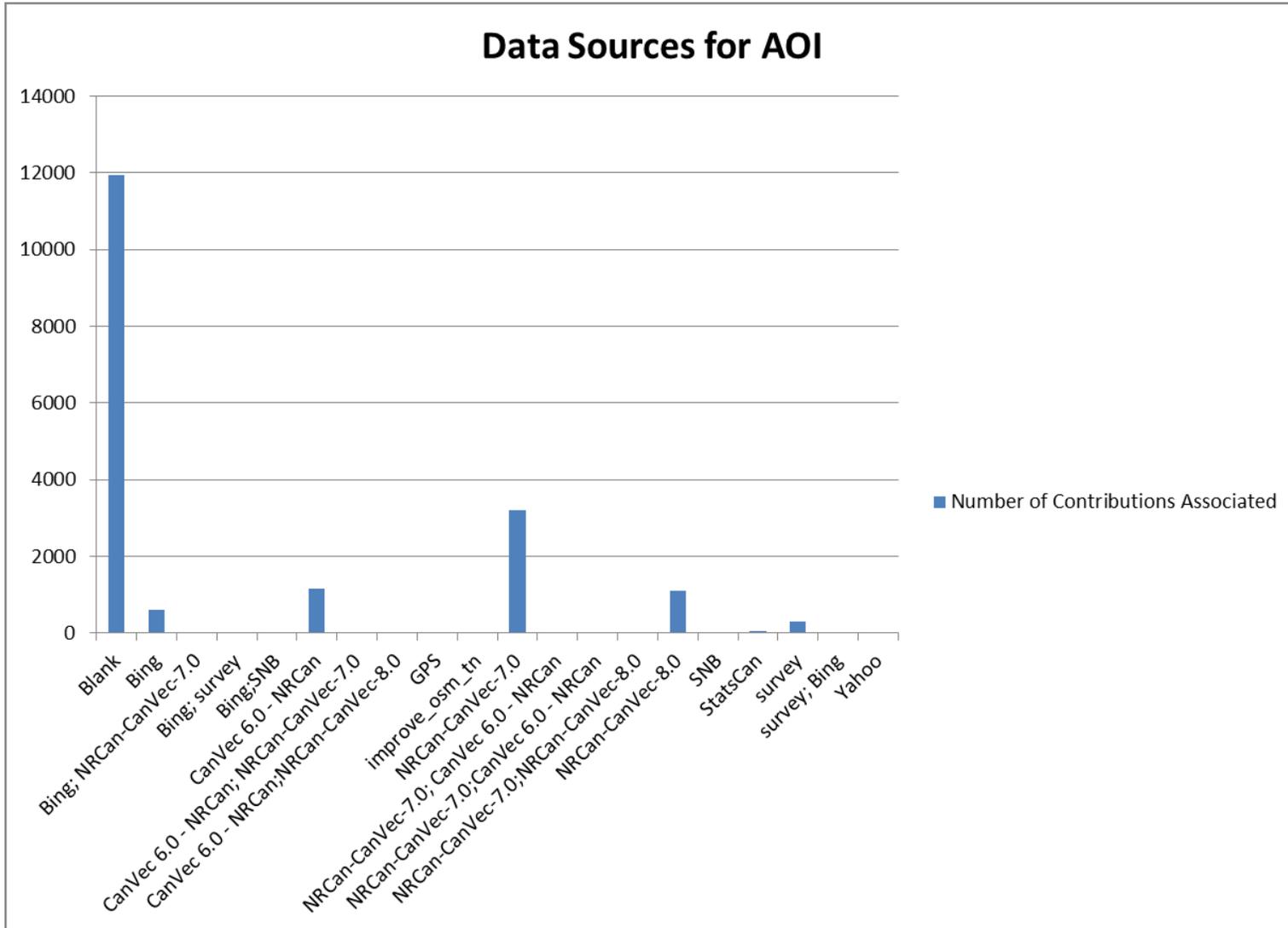


Figure 4.9 – highway_Line contributions and their related data sources for the AOI

Table 4-4 – Breakdown of Total highway_Line segments and their Contribution Data Source for the AOI

Source	Number of Contributions Associated	Total Length (m) of the OSM Contributions
Blank	11934	4357508.7600
Bing	595	232025.8673
Bing; NRCan-CanVec-7.0	4	2813.093022
Bing; survey	1	526.8318671
Bing;SNB	2	8134.757021
CanVec 6.0 - NRCan	1168	1380222.232
CanVec 6.0 - NRCan; NRCan-CanVec-7.0	7	8614.352439
CanVec 6.0 - NRCan;NRCan-CanVec-8.0	10	21211.27049
GPS	11	12702.62832
improve osm tn	18	3003.595
NRCan-CanVec-7.0	3199	2572759.394
NRCan-CanVec-7.0; CanVec 6.0 - NRCan	1	977.0977728
NRCan-CanVec-7.0;CanVec 6.0 - NRCan	22	76608.4363
NRCan-CanVec-7.0;NRCan-CanVec-8.0	6	45728.9121
NRCan-CanVec-8.0	1102	1118728.557
SNB	1	50.48140852
StatsCan	43	21981.06257
survey	300	126269.2914
survey; Bing	1	524.1010083
Yahoo	1	301.1523997
Grand Total	18426	9990691.874

Overall, this section shows that there are several factors which need to be considered before trusting the reliability of the OSM data for a geographic region under scrutiny.

The analysis has shown that:

1. Contributions to the dataset are often made by very few of the registered users to the OSM project and thus the impressive user growth stats for the project should not be seen as a true measure of the OSM mapping community’s true volume of actual contributors;
2. Peer revision of the contributions (i.e. version number as a consideration of Linus’ Law) should not be conclusively considered as a reliable measure for determining the reliability of the data for a region, and
3. The OSM dataset has grown in thematic and spatial completeness in relation to technological developments and the availability of open data infrastructures and is

directly shown by the frequency and regions which are bulk uploaded from trusted sources.

4.2 Spatial and Temporal Analysis and Use of OSM Data

The derived process for analyzing and rendering the OSM data attributes identified in Section 3.4 was developed through several trials before the final workflow was produced. Many tools and software platforms were trialed to determine the most robust and repeatable workflow possible given the constraints and limitations set forth in Section 3.1.

4.2.1 Spatial Analysis

The first portion of the process involves conducting a spatial accuracy analysis similar to those conducted by Ciepluch et al (2010), Girres and Touya (2010), Goodchild and Hunter (1997), Gröchenig et al. (2014), Haklay (2010), Quinn (2015), and Zheng and Zheng (2014); however, for this research, an approach was conducted wherein the buffer analysis method saw the buffers being placed around the OSM linear features versus the trusted professional dataset. Initially this analysis method was conducted using the entire NBRN dataset to test the .osm provincial extract data from Geofabrik.de. The results of this analysis indicated that there was minimal variance between the two datasets (i.e. $\pm 5m$ of spatial variance for roughly 97% of the roadways).

In testing of the thesis trial scenario the same spatial comparison approach was used. However, the AOI was now user defined and the OSM data was provided through the Overpass.XAPI. The main reasoning for these changes comes from the fact that the preprocessed .osm data from Geofabrik.de contains only spatial reference data and not the complete set of attributional values making further analysis (i.e. thematic cross-

referencing) impossible to achieve. The results of the test scenario were then plotted (Figure 4.10) and yielded similar outcomes where an overall average of 97% of the roads contained in the OSM dataset fell within $\pm 5m$ of the professionally generated and trusted NBRN dataset.

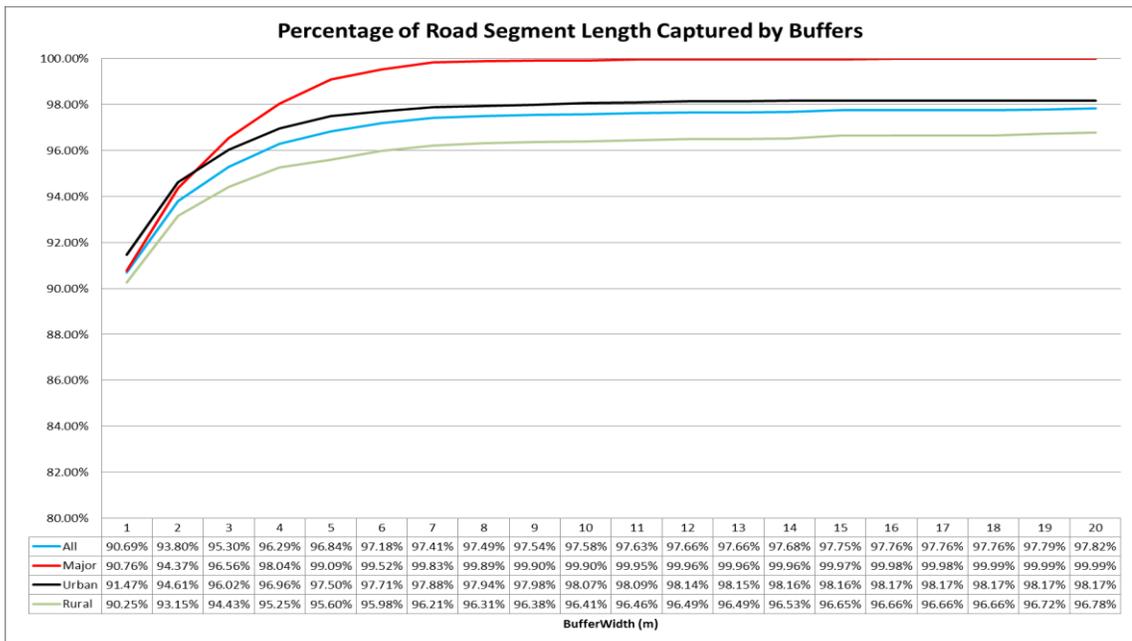


Figure 4.10 – Thesis Test Scenario Buffer Analysis Results

From this analysis it was also noted that over 90% of the OSM highway_Line feature class was within $\pm 1m$ of the NBRN features to begin with. As the buffer limits were extended, 95% of the two datasets were found to be within $\pm 4m$ of each other, at $\pm 7m$ a plateau effect begins to occur, with 99.83% of the Major roads, 97.88% of the Urban roads, and 96.21% of the Rural roads being captured, and beyond that no significant change was noted. These results are about what would be expected considering the discussion in Section 4.1.2 regarding the fact that the OSM dataset for the AOI contains multiple contributions based on a trusted professional (CanVec) dataset. With this form of contribution type being present in the OSM data for the AOI, it should be considered that these data are not pure VGI. This is due to it also being compiled using volunteer

labour to produce the dataset, not just volunteer primary data collection, and would then lend to the dataset's credibility based on the known source.

Due to the fact that the OSM project bases its highway_Line feature class (XML Tagging) around capturing all routes including bike paths, trails, etc. I saw this as a desirable trait for providing the most options for reaching a destination in the case of disaster aid provision. As mentioned in Section 3.4.2, a quick summation of the linear length of the features contained within the AOI datasets (NBRN and OSM) was conducted and compared. The total length of road segment entities contained in the AOI for the NBRN dataset is 7,535,733m and the total length of OSM highway_Line entities is 9,990,691m, a difference of 2,454,958m or 24.6% longer than the trusted dataset. When further analysis of the attributes is conducted, the total number of unique road segment entities within the AOI for the NBRN dataset is 19,703 and the number of unique entities contained within the OSM highway_Line feature class is 18,426, a difference of 7%. This brought to bear the question of *“how could the dataset with the longer linear length contain less unique entities?”* The reason for this discrepancy was found to be in the conversion process of the .osm data structure to the ArcGIS ingestible feature classes used in the analysis. As discussed in Section 4.1.2, improper tagging standards for the various highway_Line contributions were found to be the root of the problem. As an example of this, OSM would treat an improperly/unnamed Way contribution as being one long road segment, whereas that same road, which has a name/designation change part way along its length, would be represented as two unique entities in the NBRN dataset causing it to contain more unique entities for the same length of linear feature. Another possible aspect leading to this discrepancy is the fact

that all Ways in the OSM data fall into this feature class. Unnamed trails/paths/ etc. (e.g. the GPS track of a cyclist) would then be responsible for the increased linear length but be lumped together as a single entity and may not be identified as their own separate and unique entities.

4.2.2 Temporal Analysis and Rendering of OSM Timestamps

With the spatial accuracy of the OSM data for the AOI having been ascertained, it is then assumed that the decision to use purely OSM data has been made. The next step in the workflow is to create mapping products which could be useful in assigning data collection requirements. In order to accomplish this, analysis and temporal rendering of the OSM data for the region is conducted. For the Thesis Test Scenario, the results depicted in Figure 4.11 (Enlarged at Annex B, Appendix 3) were then rendered as the output product.

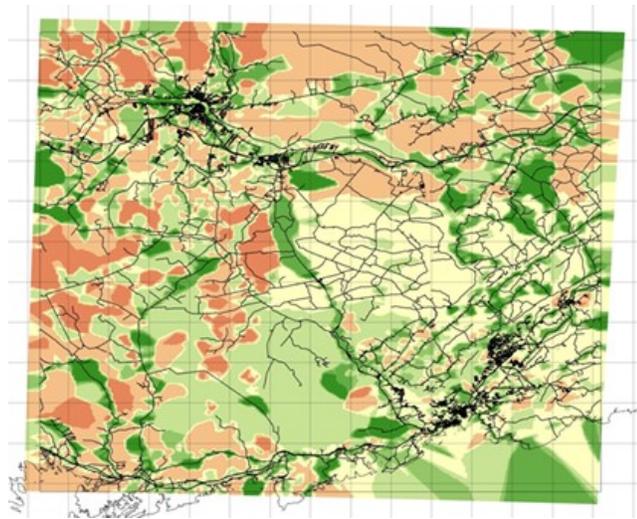


Figure 4.11- *IDW Interpolation of Timestamps for the Thesis Test Scenario AOI*

The resulting outputs seen in Figure 4.11 clearly indicate the areas, in red/orange, where the OSM data available was the oldest and hence data collection activities should be prioritized as being conducted there first. As stated earlier, this decision is made

under the context of an operational scenario in which no other recent data exist (e.g. the Haiti earthquake). With the presence of potential surge responses in contributions, this analysis would need to be conducted regularly (i.e. daily) during the follow on operational response, in order for this type of output product to remain current and useful. Also, this type of output product would only be useful as a starting point if no other credible intelligence regarding the disaster affected area could be collected ahead of time for use in data collection prioritization.

As a note on this part of the analysis and rendering process, the spatial scope of the AOI (i.e. if it is too large) will directly affect the ability to provide useable results in the interpolation and rendering process. When the ArcGIS based IDW interpolation is applied to a large area, with a vast number of data points, certain constraints such as the overall number of points to be included in the interpolation come into effect. This causes the output to become very block like and unusable when specific areas are magnified (Figure 4.12) which was observed in the output of the process workflow at Annex B Appendix 2.

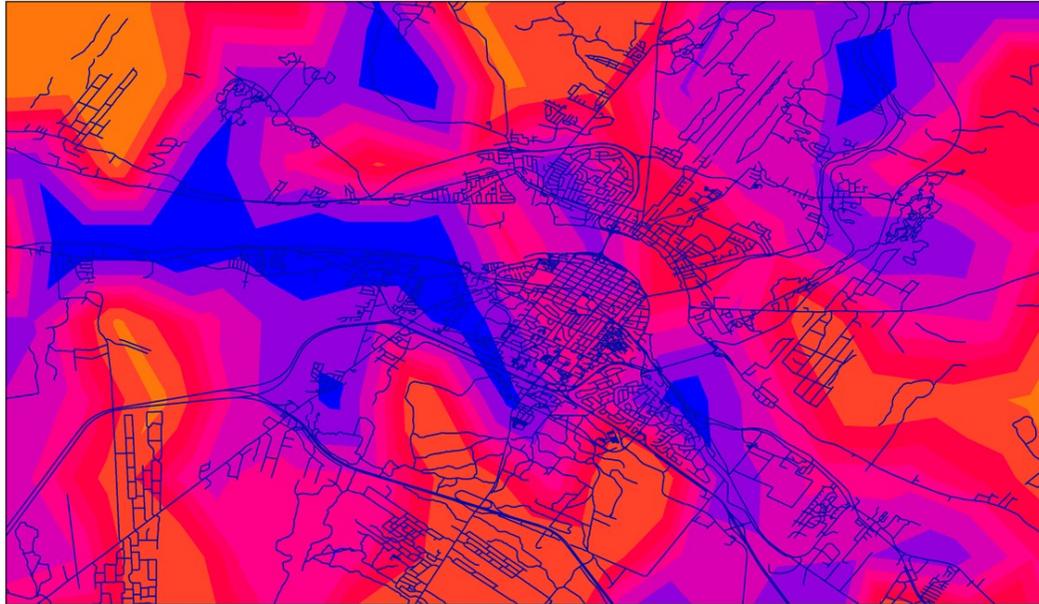


Figure 4.12 – Magnified view of Thesis Test Scenario Temporal Interpolation Output (Scale ~ 1:50,000)

The outputs seen in Figure 4.12 are still usable for the purposes of a large scale operation however if the AOI is shrunk, the results become more spatially precise in identifying exact locations for data collection needs. Figure 4.13 emphasizes this point by depicting a much smaller AOI, the Carleton University campus, in which this same workflow was applied during the process development.

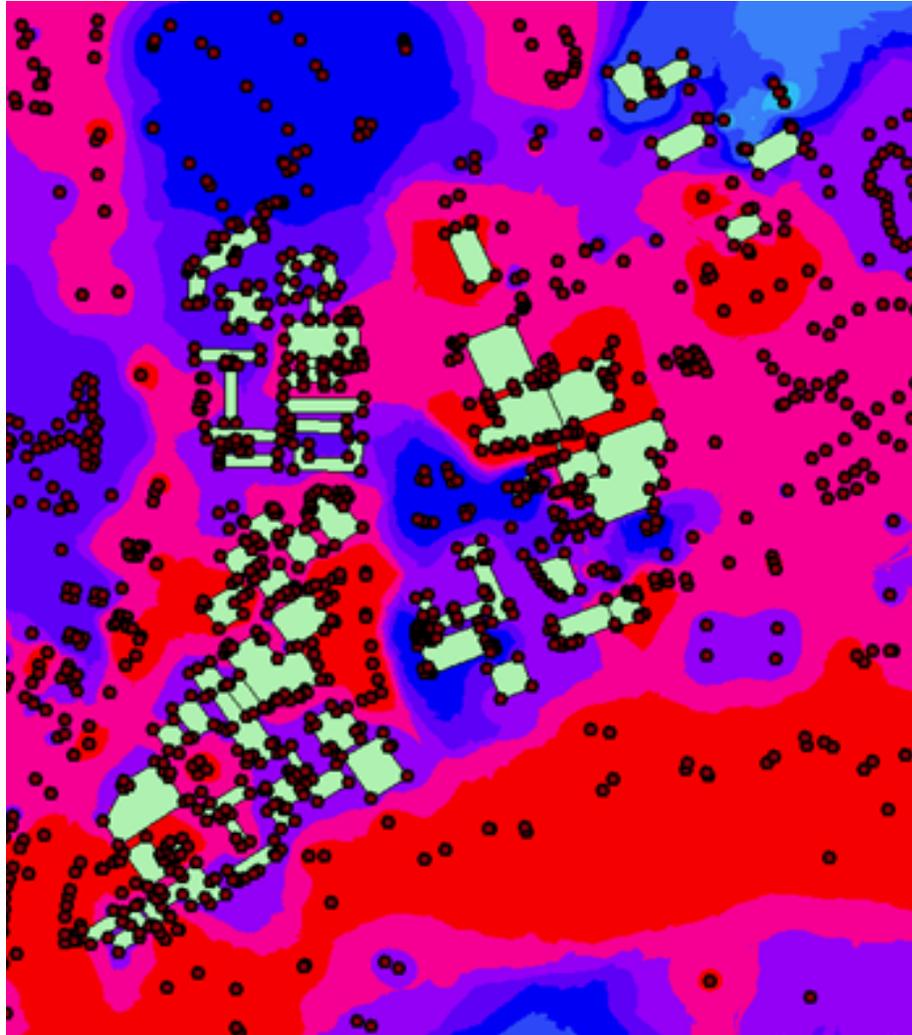


Figure 4.13 – Temporal Analysis of Carleton Campus OSM data (Scale ~ 1:10,000)

Figure 4.13 depicts the individual buildings (light green) with the relative temporal currency of the OSM data outlined spatially by an IDW interpolation of the area. When the temporal analysis process is applied to a smaller extent, the interpolation provides a more precise and manageable depiction of the specific areas which would require further data collection.

The analyses of the temporal aspects of the OSM dataset are useful for immediate matters; however, one must be mindful that this is a living dataset which is constantly evolving. As an enhancement to the reliability of the data though, the *timestamps*

analyzed in association with the contributions are an automatically generated attribute which are not subject to the interpretation of the contributor. As discussed in Section 4.1 and exemplified in Section 3.5.1, a solid understanding of the open data infrastructure (if any) available for a region will vastly drive the accuracy and reliability of the OSM data for the region's temporal attributes.

4.3 A Final Note on the Results Obtained

In this chapter, consideration of the results obtained throughout this research was conducted. The resulting outputs of the Thesis Test Scenario, once the process designed in Chapter 3 was applied, were scrutinized and several key discoveries were made with respect to additional considerations which must be made before employing OSM data as the sole source of data for an operation.

Chapter 5: Summary and Conclusions

5.1 Research Summary

This research has served to form a better understanding of the background regarding the OSM contribution history and how the behaviours of the VGI community can be viewed in terms of the type and accuracy of their contributions. The sub-components of the main research question provided guidance in the directions taken when conducting testing and analysis with regards to answering the greater question of:

“What are the quantifiable characteristics which can be used to determine the spatial and thematic accuracy as well as the reliability of the contributors of VGI in particular OpenStreetMap.org or OSM data?”

The following summary relates the results discussed in Chapter 4 with their respective importance to the research sub-questions posed in this thesis.

5.1.1 What limitations exist within OSM which must be further understood?

In Section 4.1.1, review of the historical contributions to the OSM dataset indicated that the OSM community reacts accordingly, in terms of mapping efforts following a natural disaster, for the volume of contributions to a region where no other open data source is available. If the HOT is called into action and humanitarian gestures such as the provision of free and recent imagery and data are provided, the OSM project has shown it can provide a collaborative platform to map a region well in a short period of time.

In the context of the operational process design created in this research, whether or not a region of interest has an open data infrastructure, should be considered a limitation. For regions with this infrastructure in place, heightened levels of accuracy and reliability can be assumed about the current OSM data than for regions without. From Gröchenig et

al. (2014b), it was indicated that mapping activity in the realm of VGI tended towards linear features (i.e. routes) always being contributed first and edited more frequently. The results I obtained (Annex A, Appendix 3) support this conclusion and as such I would recommend that in the case of operational OSM data use, only the route (i.e. highway_Line) information it contains should be sought.

With the absence of an enforced standard tagging schema for contributions and edits to the dataset, the results obtained in Section 4.1.2 indicate that OSM thematic analysis should not be a focus for operational purposes. The dataset is rich with unverified information and, thematically, I would recommend its use be limited to query or cross-verification only.

5.1.2 How do we quantify contributors and their contributions for use in the determination of a reliability index or measure?

The results of the Thesis Test Scenario discussed in Section 4.1.2 supported the 90:9:1 principle (Neilson, 2006). The results clearly indicated that very few contributors had actually performed the largest portion of effort in generating the data for the region. The results I obtained in this section (spatially represented at Annex B Appendix 4) support the conclusion that the version number of an entity/contribution contained within the AOI dataset cannot be used as a viable metric for determining the reliability of the source due to the fact that the majority of contributions are either the initial version or have had only one review/edit.

Due to the fact that the OSM project utilizes numerical user ids and protects the anonymity of its users through user names, a direct measure of reliability for each contributor is an untenable goal. Instead, this research sought to apply the 90:9:1 rule to the OSM data and review the top ranked contributors based on their volume of

contributions to the current dataset. It was found that a large number of the contributions made by these contributors came from already trusted data sources. From these findings I recommend further application of this type of analysis to other regions in which open data infrastructures are present, as the presence and use of data from an open data infrastructure could be seen as a possible metric for reviewing a contributors' level of reliability.

An analysis of the contribution sources themselves also provided a possible method for determining contribution reliability. Since large numbers of contributions were directly input from an already trusted source, I argue that this in itself could be used as a quantifiable measure for the reliability of the data present in the AOI. If a percentage of the data in a region had already come directly from a trusted source, it should be viewed as a partial measure of that region's data reliability.

5.1.3 What spatial and thematic variance thresholds must be established in order to determine the viability of applying the reliability index or measure before the OSM data is used?

Based on the findings of Section 4.2, a clear definition of these thresholds is necessarily an arbitrary matter. The question of whether or not having unreliable data is better than having no data at all, is the primary focus of this argument. In terms of spatial variance, the analysis showed a very strong correlation between the OSM dataset and that of the trusted dataset. However, the comparison process relied on trusted data being available for the comparison as well for applying the measures discussed in Section 5.1.2. The process I have developed in this research requires trusted data only for analysis and does not require its use in order to generate the final output products. I assert that given the results obtained, the operational necessity for data should be the primary driver in deciding whether or not to use OSM data for a region. As discussed further in Sections

4.3 and 5.1.1, if the proper considerations and constraints are placed on the usage of the data, then it should be considered as a viable and credible source.

5.1.4 Can OSM data be used to augment existing datasets once the spatial and thematic accuracy of the data as well as the reliability of its contributors has been ascertained as fitting through a form of threshold analysis?

The unique data structure used by OSM for the provision of flexibility when providing information makes it difficult to fully ingest into a traditional GIS format. During the course of this research it was noted on several occasions that the dataset did contain richer thematic and spatial characteristics than that of the comparison dataset, yet often lacked in its fullness of descriptive attributes for every unique entity it contained. As the OSM project's initial creator Steve Coast (2007) pointed out, the GIS community should let go of the concept of information completeness. I agree with this concept and maintain that the data contained in the OSM dataset could be used to augment existing data holdings given the proper understanding of the required datasets fitness for purpose. As contended in Section 5.1.3, I have come to believe it should be the information consumer's responsibility to not only constrain their data needs but be responsible for understanding the very nature of the OSM dataset. It is a living dataset which is evolving over time and, as with the situations encountered in Haiti and Nepal, can drastically grow in comprehensiveness in as little as a single week.

5.2 Conclusions and Future Research

The availability of open data sources continues to grow on a global scale. This growth, along with the combined efforts of the VGI community, will ensure that initiatives such as the OSM project will continue to be developed and thrive. The OSM project itself is a landmark effort which has shown some of the most remarkable achievements which can be attained through the collaborative efforts of interested

contributors. The project's inclusion of all levels of contributors has even driven governmental agencies to re-think the ways in which they gather data and information.

Through the use of the tools outlined in this research, it has been established that metrics for measuring the spatial accuracy and the reliability of contributions and contributors can conceptually be established. The results of the analysis conducted have shown several strengths and weakness inherent in the OSM data, however I feel that the employment of OSM data should remain a consideration by any agency deploying to a disaster affected region. Further testing of the processes developed across a larger scale could provide proper heuristics for regular application in an analysis process in order to streamline the trusted use of OSM data. The methods developed in this study have shown through comparative analysis that the data contained within the OSM dataset often parallels, and in some cases even provides additional value to that of trusted data sources.

Although the basis of this research and analysis was to develop an expedient process for actually employing OSM data once a certain threshold of spatial accuracy and temporal currency was understood, factors such as those discussed in Haklay (2010), were not studied. As indicated by Haklay (2008), factors such as those found in the IMD, wherein underlying indicators such as Income, Employment, Health Care, Education, etc. serve to affect whether or not a region should be viewed as having the capacity to maintain a solid contribution status within the OSM project. Further efforts should be made in this area to determine if there is in fact an effect on data quality and completeness based on income level or access to an open data infrastructure in a region as more open data sources emerge and are integrated into the OSM project.

REFERENCES

- Aedo, I., Díaz, P., Carroll, J. M., Convertino, G. and Rosson, M. B. (2010): End-user oriented strategies to facilitate multi-organizational adoption of emergency management information systems, *Information Processing & Management*, 46(1), 11–21, doi:10.1016/j.ipm.2009.07.002.
- Antoniou, V., Haklay, M. and Morley, J. (2010): A step towards the improvement of spatial data quality of Web 2.0 geo-applications: the case of OpenStreetMap, in the 18th GISRUK Conference, London, UK, pp. 197–201. [online] Available from: http://www.academia.edu/download/33774191/GISRUK2010_VyronAntoniouMukiHaklayJeremyMorley.pdf.
- Barron, C., Neis, P. and Zipf, A. (2014): A Comprehensive Framework for Intrinsic OpenStreetMap Quality Analysis, *Transactions in GIS*, 18(6), 877–895, doi:10.1111/tgis.12073.
- Buscher, M. and Wood, L. (2013): A new Manhattan project?: interoperability and ethics in emergency response systems of systems, [online] Available from: http://eprints.lancs.ac.uk/62390/1/269_ManhattanProject_Final.pdf.
- Canonical Ltd. (2017): Ubuntu, [online] Available from: <https://www.ubuntu.com/>.
- Ciepluch, B., Jacob, R., Mooney, P. and Winstanley, A. C. (2010): Comparison of the accuracy of OpenStreetMap for Ireland with Google Maps and Bing Maps, in *Proceedings of the Ninth International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences 20-23rd July 2010*, p. 337, University of Leicester. [online] Available from: <http://eprints.maynoothuniversity.ie/2476>.
- Coast, S. (2007): OpenStreetMap, NCGIA, Los Alamos National Laboratory, the Army Research Office and The Vespucci Initiative. [online] Available from: http://www.ncgia.ucsb.edu/projects/vgi/docs/present/Coast_openstreetmap-opendata.pdf.
- Cohn, J. P. (2008): Citizen science: Can volunteers do real research?, *BioScience*, 58(3), 192–197.
- Coleman, D. J., Georgiadou, Y., Labonte, J. and others (2009): Volunteered geographic information: The nature and motivation of producers, *International Journal of Spatial Data Infrastructures Research*, 4(1), 332–358.
- Corel Corporation (2017): WinZip for Windows, Mac and Mobile - Zip Files, Unzip Files, [online] Available from: <http://www.winzip.com/win/en/>.
- Edgewall Software (2016): JOSM - Java OpenStreetMap Editor, [online] Available from: <https://josm.openstreetmap.de/>.
- ESRI (2014): ArcGIS 10.3 for Desktop. Redlands, US: Environmental Systems Research Institute, [online] Available from: <http://www.esri.com/en/arcgis/products/arcgis-pro/overview>.

- FEMA (2017): FEMA.gov | Federal Emergency Management Agency, [online] Available from: <https://www.fema.gov/>.
- GeoGratis.gc.ca (2017): CanVec Information Index, [online] Available from: http://ftp.geogratias.gc.ca/pub/nrcan_rncan/vector/canvec/doc/info.html.
- Girres, J.-F. and Touya, G. (2010): Quality Assessment of the French OpenStreetMap Dataset, *Transactions in GIS*, 14(4), 435–459, doi:10.1111/j.1467-9671.2010.01203.x.
- Goodchild, M. F. (2007): Citizens as sensors: the world of volunteered geography, *GeoJournal*, 69(4), 211–221, doi:10.1007/s10708-007-9111-y.
- Goodchild, M. F. and Hunter, G. J. (1997): A simple positional accuracy measure for linear features, *International Journal of Geographical Information Science*, 11(3), 299–306.
- Goodchild, M. F. and Li, L. (2012): Assuring the quality of volunteered geographic information, *Spatial Statistics*, 1, 110–120, doi:10.1016/j.spasta.2012.03.002.
- Government, C. and L. (2007): Indices of Deprivation 2007, [online] Available from: <http://webarchive.nationalarchives.gov.uk/20100806161347/http://www.communities.gov.uk/communities/neighbourhoodrenewal/deprivation/deprivation07/>.
- Government of Canada, P. S. C. (2013): Canadian Disaster Database, [online] Available from: <http://cdd.publicsafety.gc.ca>.
- Gröchenig, S. (2012): Using Spatial and Temporal Editing Patterns for Evaluation of Open Street Map Data, Master's Thesis, Carinthia University of Applied Sciences, Villach, Austria. [online] Available from: <https://static1.squarespace.com/static/559921a3e4b02c1d7480f8f4/t/5859dccb9de4bb73f2fef867/1482284239471/Groechenig.pdf>.
- Gröchenig, S., Brunauer, R. and Rehrl, K. (2014a): Digging into the history of VGI data-sets: results from a worldwide study on OpenStreetMap mapping activity, *Journal of Location Based Services*, 8(3), 198–210, doi:10.1080/17489725.2014.978403.
- Gröchenig, S., Brunauer, R. and Rehrl, K. (2014b): Estimating Completeness of VGI Datasets by Analyzing Community Activity Over Time Periods, in *Connecting a Digital Europe Through Location and Place*, edited by J. Huerta, S. Schade, and C. Granell, pp. 3–18, Springer International Publishing, Cham.
- Haklay, M. (2008): Indices of Deprivation 2007, Po Ve Sham - Muki Haklay's personal blog [online] Available from: <https://povesham.wordpress.com/2008/01/17/indices-of-deprivation-2007/>.
- Haklay, M. (2010): How Good is Volunteered Geographical Information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets, *Environment and Planning B: Planning and Design*, 37(4), 682–703, doi:10.1068/b35097.
- Haklay, M. (2013): Citizen Science and Volunteered Geographic Information: Overview and Typology of Participation., in *Crowdsourcing Geographic Knowledge*, edited by D. Sui, S.

- Elwood, and M. Goodchild, Springer Netherlands, Dordrecht. [online] Available from: <http://link.springer.com/10.1007/978-94-007-4587-2>.
- Haklay, M. (2015): Citizen Science and Policy: A European Perspective, The Woodrow Wilson Center, Commons Lab, 10.
- Haklay, M. (Muki), Basiouka, S., Antoniou, V. and Ather, A. (2010): How Many Volunteers Does it Take to Map an Area Well? The Validity of Linus' Law to Volunteered Geographic Information, *The Cartographic Journal*, 47(4), 315–322, doi:10.1179/000870410X12911304958827.
- Henderson, B. (2016): Osmosis, [online] Available from: <https://wiki.openstreetmap.org/wiki/Osmosis>.
- HOT (2017): Humanitarian OpenStreetMap Team, [online] Available from: <https://www.hotosm.org/>.
- Jarosch, J., Snunu, S. and Wood, H. (2017): OSMembrane, Java, OpenStreetMap on GitHub. [online] Available from: <https://github.com/openstreetmap/osmembrane>.
- Kuutti, K. (2006): Activity Theory as a potential framework for human computer interaction research, in *Context and Consciousness: Activity Theory and Human-Computer Interaction*, pp. 17–44, The MIT Press: Cambridge, MA, USA.
- Lauriault, T. P. and Mooney, P. (2014): Crowdsourcing: a geographic approach to public engagement, Available at SSRN 2518233 [online] Available from: http://papers.ssrn.com/sol3/Papers.cfm?abstract_id=2518233.
- Linux Foundation (2016): Linux, Linux.com | News for the open source professional [online] Available from: <https://www.linux.com/home>.
- Marqqs (2017): Osmconvert, [online] Available from: <https://wiki.openstreetmap.org/wiki/Osmconvert>.
- Martin, J. (1983): *Managing the Data Base Environment*, 1st ed., Prentice Hall PTR Upper Saddle River, NJ, USA.
- Microsoft Corporation (2017): Microsoft – Official Home Page, [online] Available from: <https://www.microsoft.com/en-ca/>.
- Mooney, P. and Corcoran, P. (2012): Characteristics of Heavily Edited Objects in OpenStreetMap, *Future Internet*, 4(4), 285–305, doi:10.3390/fi4010285.
- Mooney, P. and Corcoran, P. (2014): Analysis of Interaction and Co-editing Patterns amongst OpenStreetMap Contributors, *Transactions in GIS*, 18(5), 633–659, doi:10.1111/tgis.12051.
- Neilson, J. (2016): Participation Inequality: The 90-9-1 Rule for Social Features, [online] Available from: <https://www.nngroup.com/articles/participation-inequality/>.

- Neis, P. and Zipf, A. (2012): Analyzing the Contributor Activity of a Volunteered Geographic Information Project — The Case of OpenStreetMap, *ISPRS International Journal of Geo-Information*, 1(3), 146–165, doi:10.3390/ijgi1020146.
- Neis, P., Singler, P. and Zipf, A. (2010): Collaborative mapping and emergency routing for disaster logistics—case studies from the haiti earthquake and the UN Portal for Afrika, in *Geospatial Crossroads @ GI Forum*, edited by A. Car et al., pp. 239–248, Geoinformatics Forum, Salzburg.
- Neis, P., Zielstra, D. and Zipf, A. (2013): Comparison of Volunteered Geographic Information Data Contributions and Community Development for Selected World Regions, *Future Internet*, 5(2), 282–300, doi:10.3390/fi5020282.
- OSMF (2017): OpenStreetMap Foundation, [online] Available from: https://wiki.osmfoundation.org/wiki/Main_Page.
- Pavlov, I. (2016): 7-Zip, [online] Available from: <http://www.7-zip.org/>.
- Quinn, S. (2015): Using small cities to understand the crowd behind OpenStreetMap, *GeoJournal*, doi:10.1007/s10708-015-9695-6.
- Ramm, F., Topf, J. and Chilton, S. (2011): *OpenStreetMap: Using and Enhancing the Free Map of the World*, UIT Cambridge, UK.
- Ramm, F., Topf, J., Karch, C., McCann, R., Beelmann, P. and Reichert, M. (2016): GEOFABRIK, Geofabrik [online] Available from: <https://www.geofabrik.de/>.
- Raymond, E. S. (1999): The Cathedral and the Bazaar, [online] Available from: <http://www.catb.org/~esr/writings/cathedral-bazaar/>.
- Rehrl, K. and Gröchenig, S. (2016): A Framework for Data-Centric Analysis of Mapping Activity in the Context of Volunteered Geographic Information, *ISPRS International Journal of Geo-Information*, 5(3), 37, doi:10.3390/ijgi5030037.
- Rehrl, K., Gröchenig, S., Hochmair, H., Leitinger, S., Steinmann, R. and Wagner, A. (2013): A Conceptual Model for Analyzing Contribution Patterns in the Context of VGI, in *Progress in Location-Based Services*, edited by J. M. Krisp, pp. 373–388, Springer Berlin Heidelberg, Berlin, Heidelberg.
- Roshal, A. (2017): WinRAR archiver, a powerful tool to process RAR and ZIP files, RARLAB [online] Available from: <http://www.rarlab.com/>.
- See, L., Schepaschenko, D., Lesiv, M., McCallum, I., Fritz, S., Comber, A., Perger, C., Schill, C., Zhao, Y., Maus, V., Siraj, M. A., Albrecht, F., Cipriani, A., Vakolyuk, M., Garcia, A., Rabia, A. H., Singha, K., Marcarini, A. A., Kattenborn, T., Hazarika, R., Schepaschenko, M., van der Velde, M., Kraxner, F. and Obersteiner, M. (2015): Building a hybrid land cover map with crowdsourcing and geographically weighted regression, *ISPRS Journal of Photogrammetry and Remote Sensing*, 103, 48–56, doi:10.1016/j.isprsjprs.2014.06.016.
- Silvertown, J. (2009): A new dawn for citizen science, *Trends in ecology & evolution*, 24(9), 467–471.

- Tobler, W. R. (1970): A Computer Movie Simulating Urban Growth in the Detroit Region, *Economic Geography*, 46, 234, doi:10.2307/143141.
- USGS (2017a): Earthquake Lists, Maps, and Statistics, [online] Available from: <https://earthquake.usgs.gov/earthquakes/browse/>.
- USGS (2017b): Landsat Processing Details: Landsat Missions, [online] Available from: <https://landsat.usgs.gov/landsat-processing-details>.
- Veregin, H. (1999): Data quality parameters, *Geographical information systems*, 1, 177–189.
- Zhang, C., Zhao, T. and Li, W. (2015): Towards an interoperable online volunteered geographic information system for disaster response, *Journal of Spatial Science*, 60(2), 257–275, doi:10.1080/14498596.2015.972996.
- Zhao, Y., Gong, P., Yu, L., Hu, L., Li, X., Li, C., Zhang, H., Zheng, Y., Wang, J., Zhao, Y., Cheng, Q., Liu, C., Liu, S. and Wang, X. (2014): Towards a common validation sample set for global land-cover mapping, *International Journal of Remote Sensing*, 35(13), 4795–4814, doi:10.1080/01431161.2014.930202.
- Zheng, S. and Zheng, J. (2014): Assessing the Completeness and Positional Accuracy of OpenStreetMap in China, in *Thematic Cartography for the Society*, edited by T. Bandrova, M. Konecny, and S. Zlatanova, pp. 171–189, Springer International Publishing, Cham. [online] Available from: http://link.springer.com/10.1007/978-3-319-08180-9_14.
- Zielstra, D. and Zipf, A. (2010): A comparative study of proprietary geodata and volunteered geographic information for Germany, in *13th AGILE international conference on geographic information science*, vol. 2010. [online] Available from: http://agile2010.dsi.uminho.pt/pen/shortpapers_pdf/142_doc.pdf (Accessed 1 December 2016).
- Zulfiqar, N. (2008): A study of the quality of OpenStreetMap.org maps: a comparison of OSM data and Ordnance Survey Data, MEng Thesis, UCL, London, UK.

..

KEY DEFINITIONS

Crowdsourcing – is the practice of obtaining information or input into a task or project by enlisting the services of a large number of people, either paid or unpaid, typically via the Internet

Index of Multiple Deprivation (IMD) – is a UK government qualitative study of deprived areas in English local councils. The first study (released in 2007) covers seven aspects of deprivation. The statistics described by the Neighbourhood Renewal Unit are:

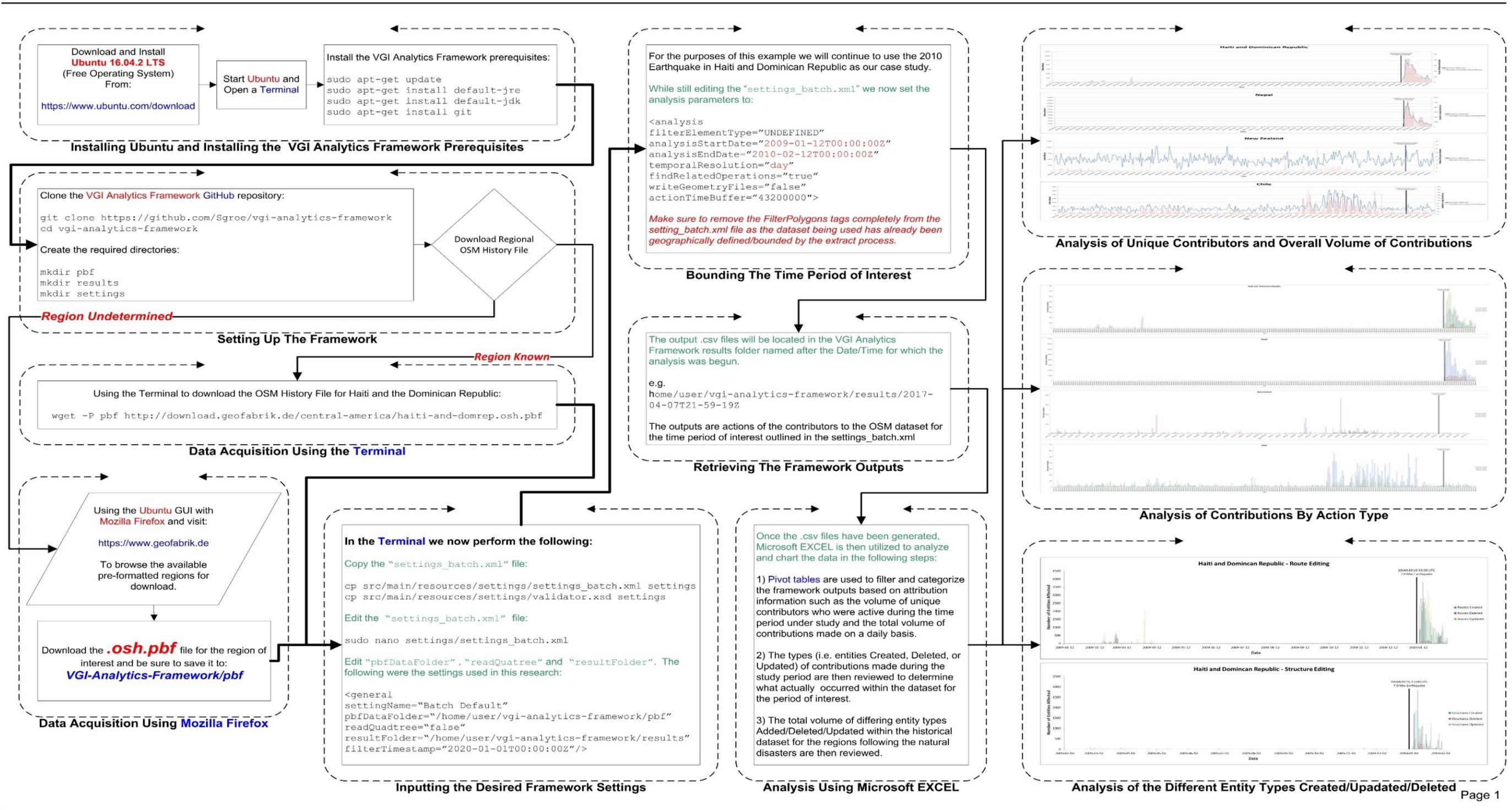
- Income
- Employment
- Health deprivation and Disability
- Education Skills and Training
- Barriers to Housing and Services
- Crime
- Living Environment.

Necessities of Life – Are key buildings and infrastructure which would be deemed as vital in an ongoing disaster management response. Outside of standard emergency services (Police, Fire, EMS), these would include places such as schools, shopping malls, sports complexes, etc. where large groups of people can safely gather to seek refuge as well as critical services such as water management, bridges, power grid, communications, etc.

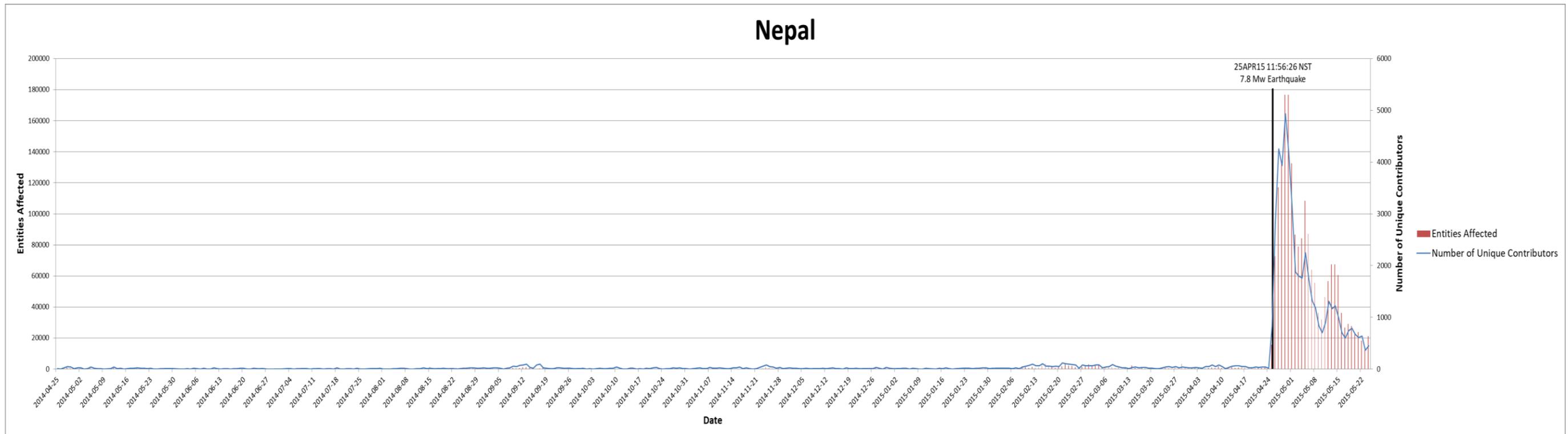
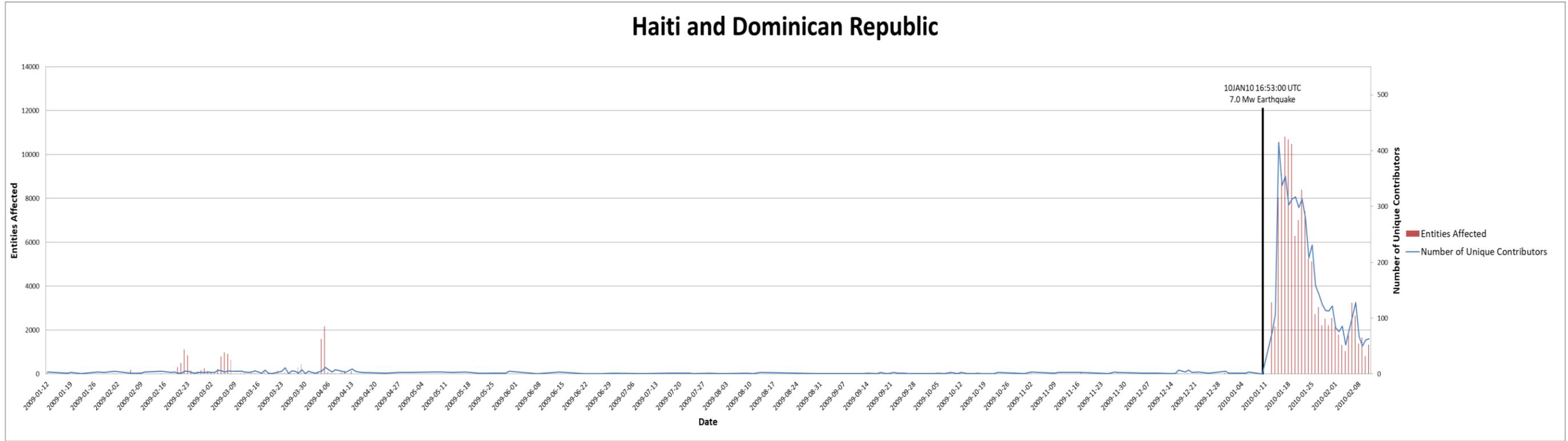
Volunteered Geographic Information (VGI) - is the harnessing of tools to create, assemble, and disseminate geographic data provided voluntarily by individuals (Goodchild, 2007)

ANNEX A – Workflow for Historical Review of the OSM Dataset

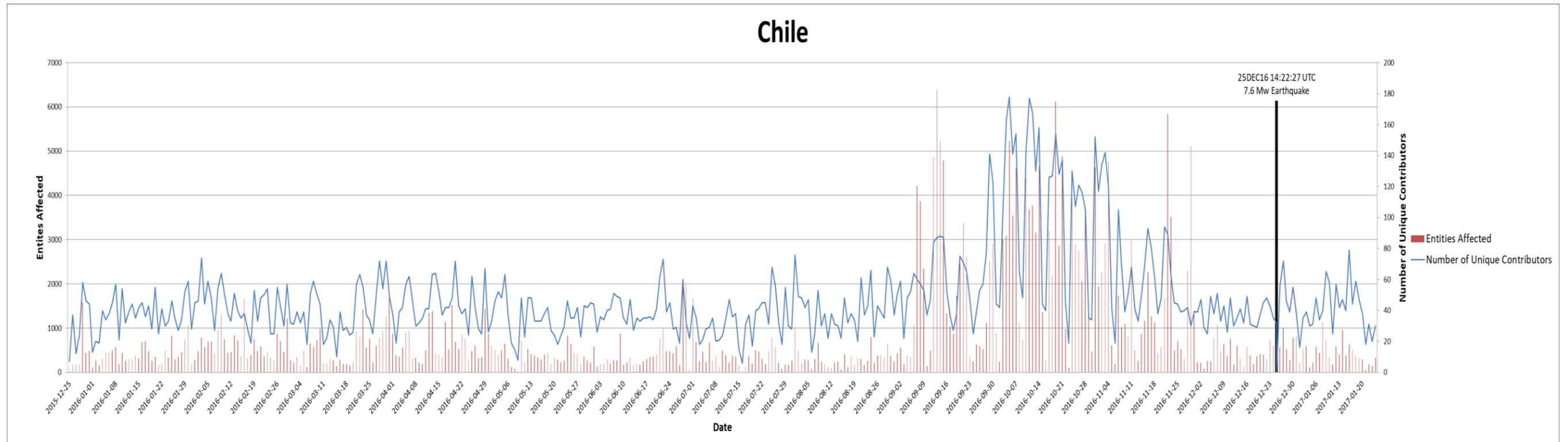
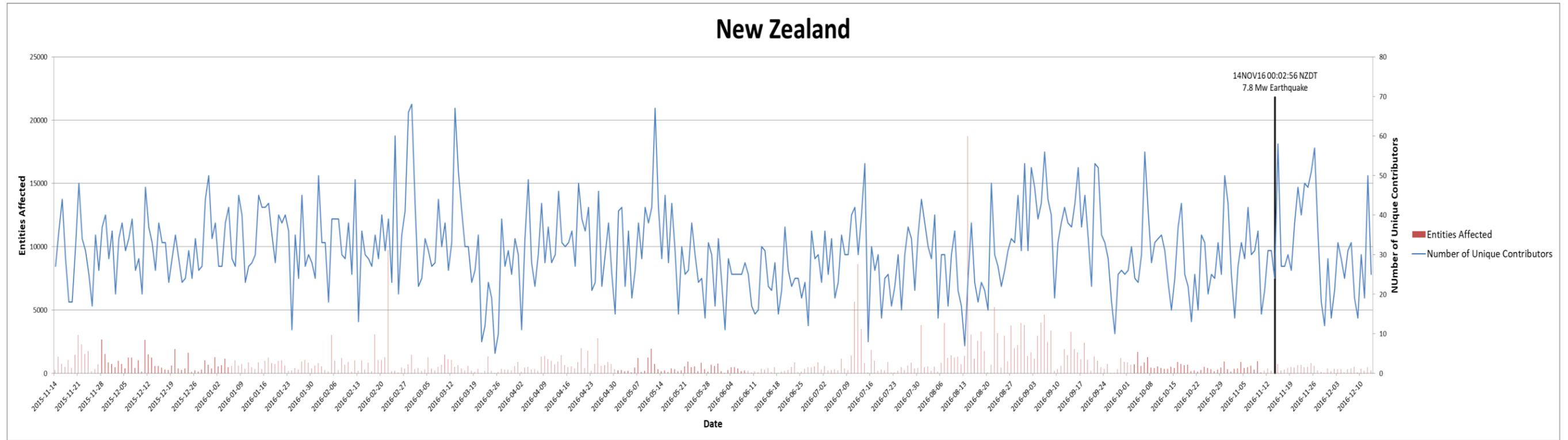
VGI Analytics Framework use for Temporal Review of OSM Contributions



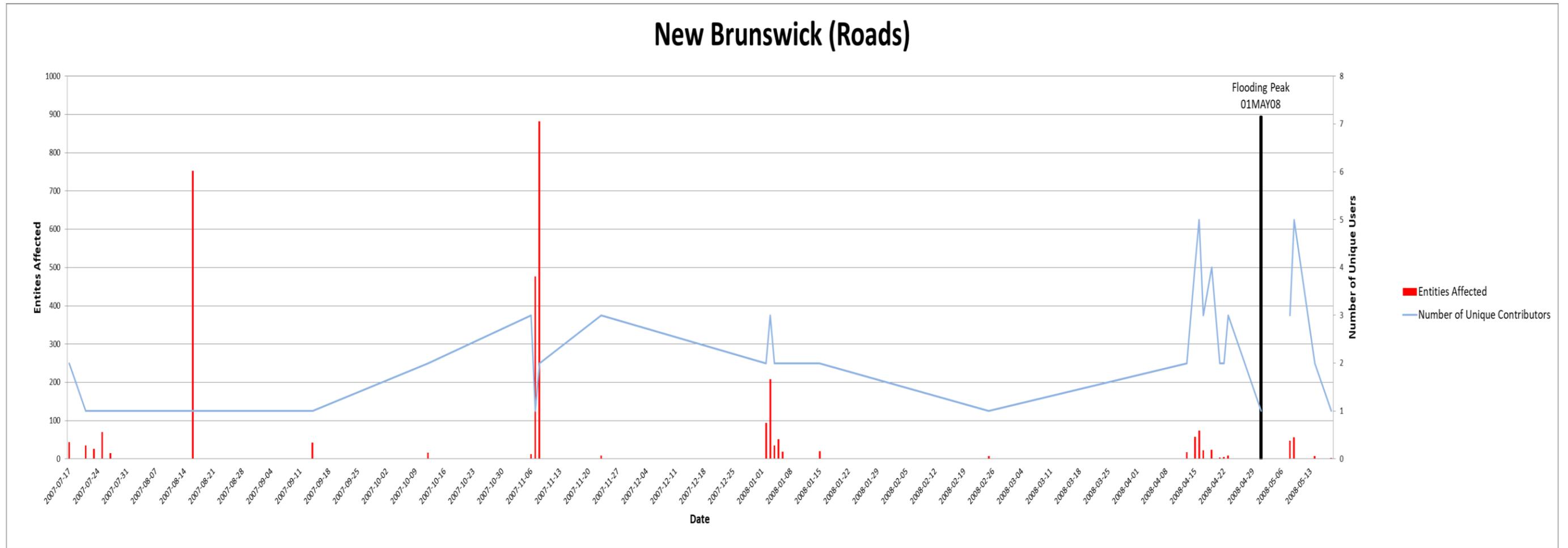
APPENDIX 1 – Number of Active Contributors and Total Volume of Contributions



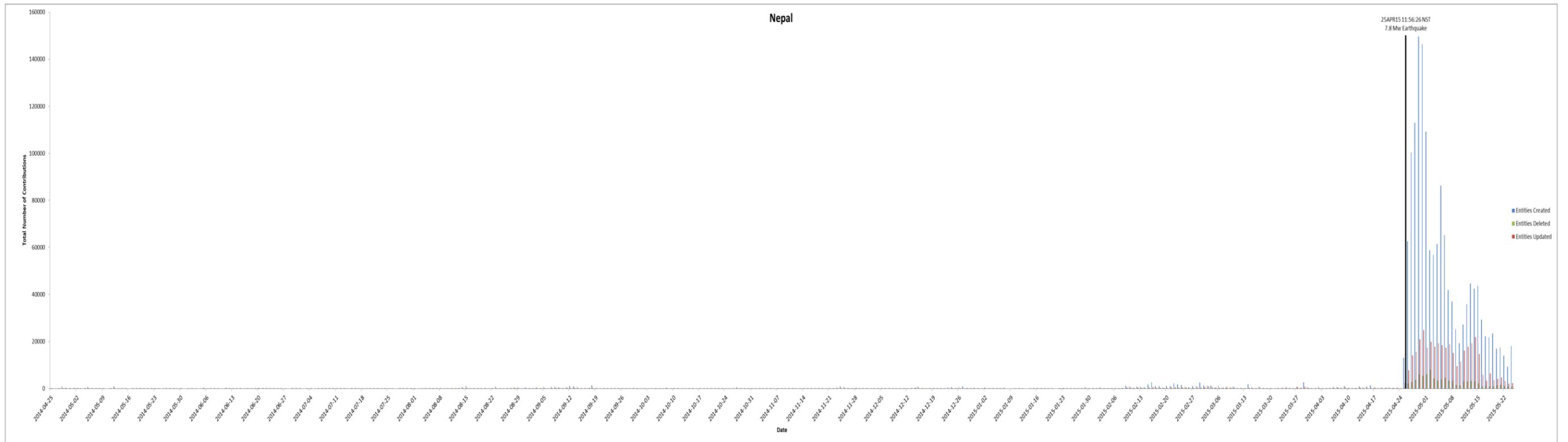
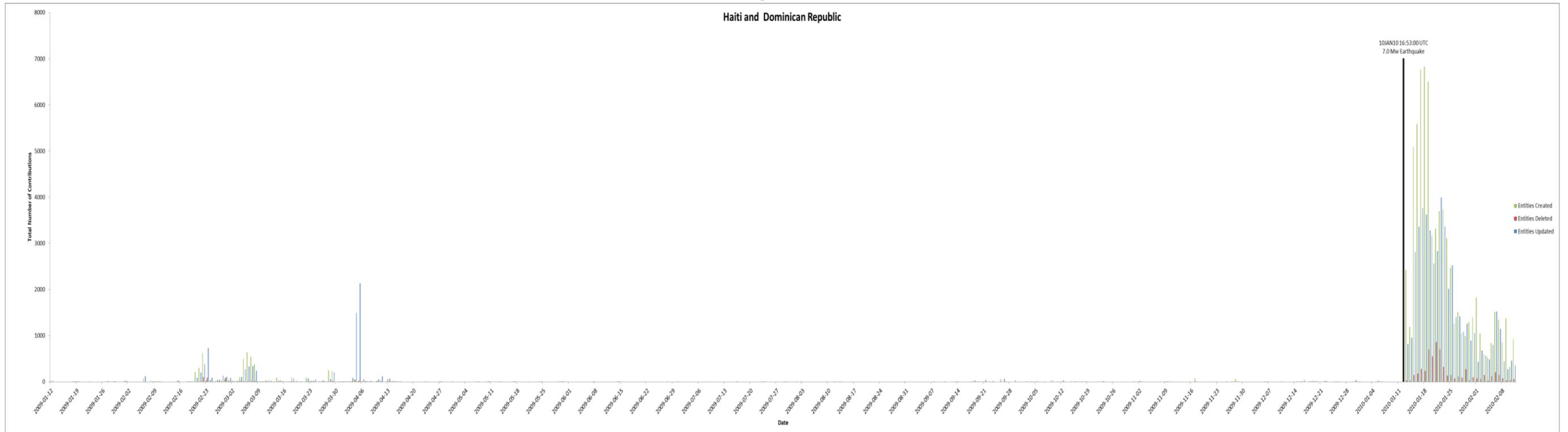
APPENDIX 1 – Number of Active Contributors and Total Volume of Contributions



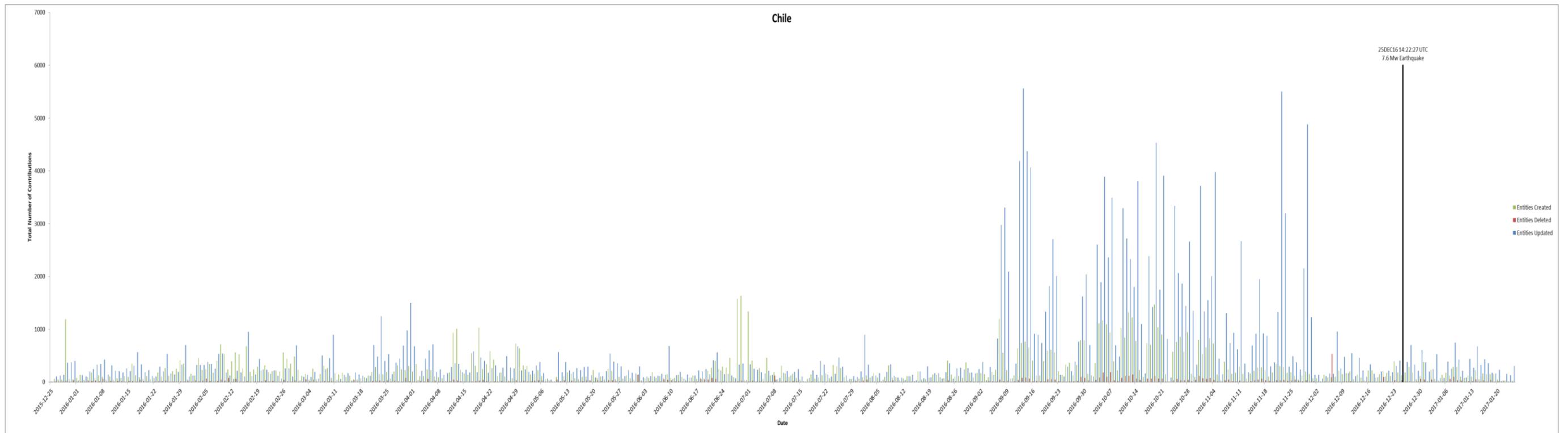
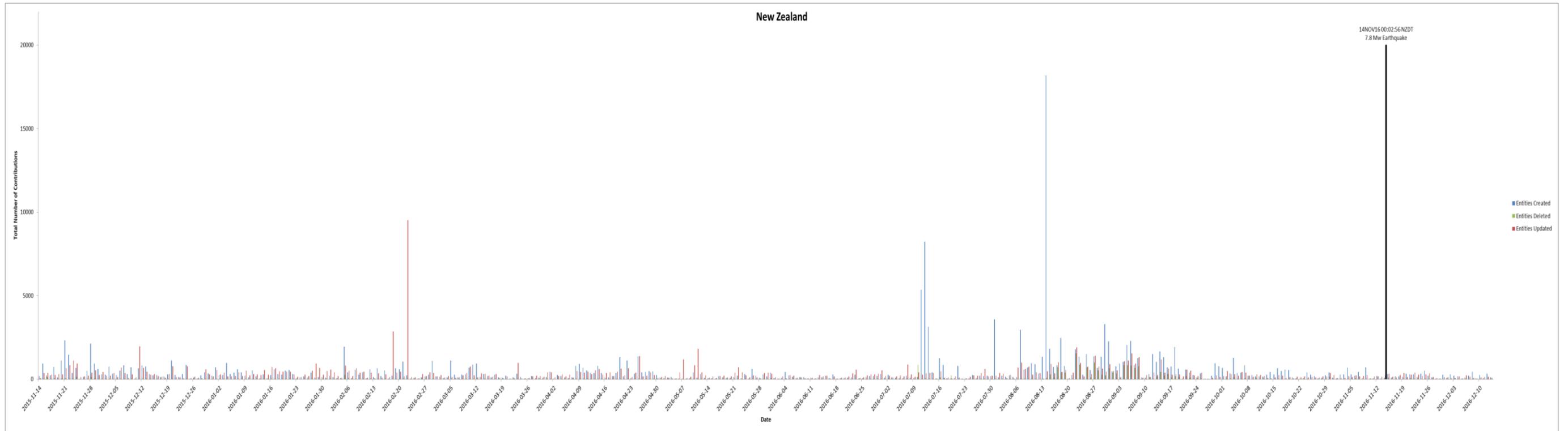
APPENDIX 1 – Number of Active Contributors and Total Volume of Contributions



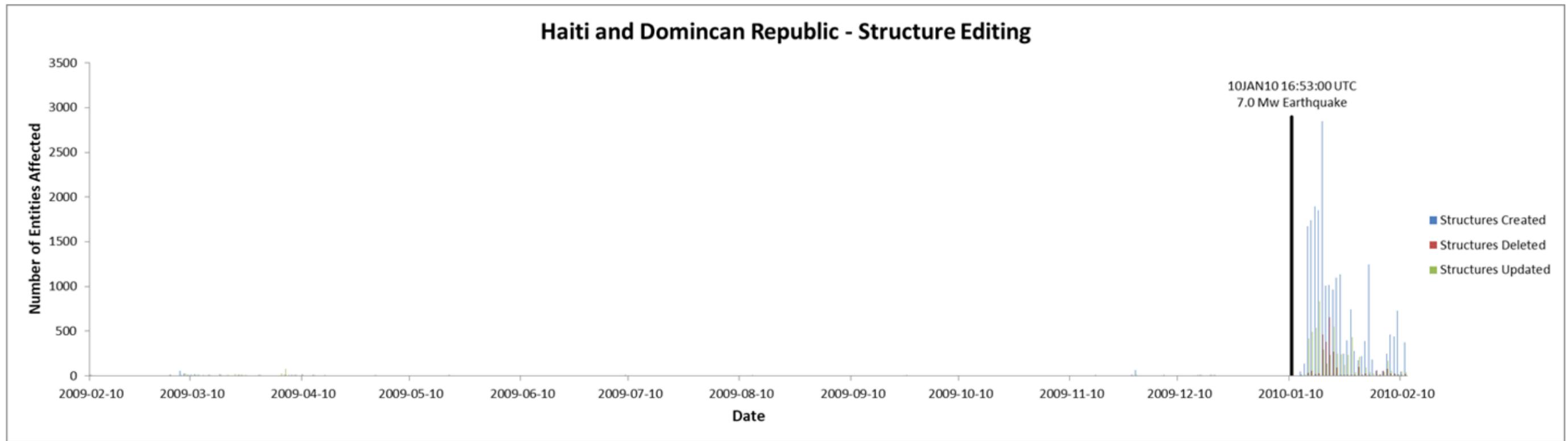
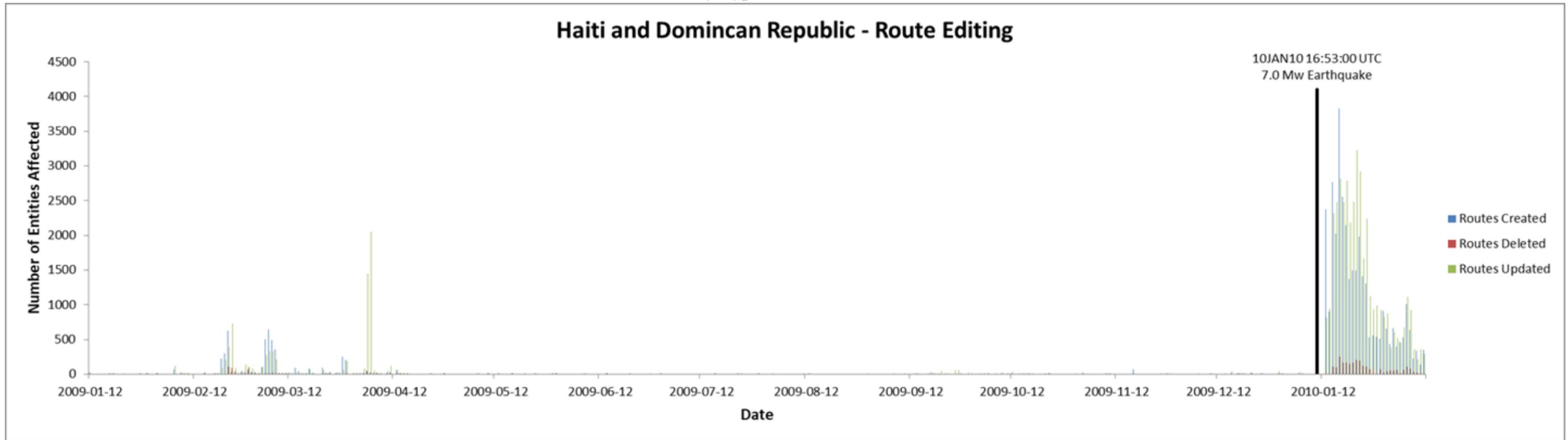
APPENDIX 2 – Actions Being Carried Out on the OSM Dataset

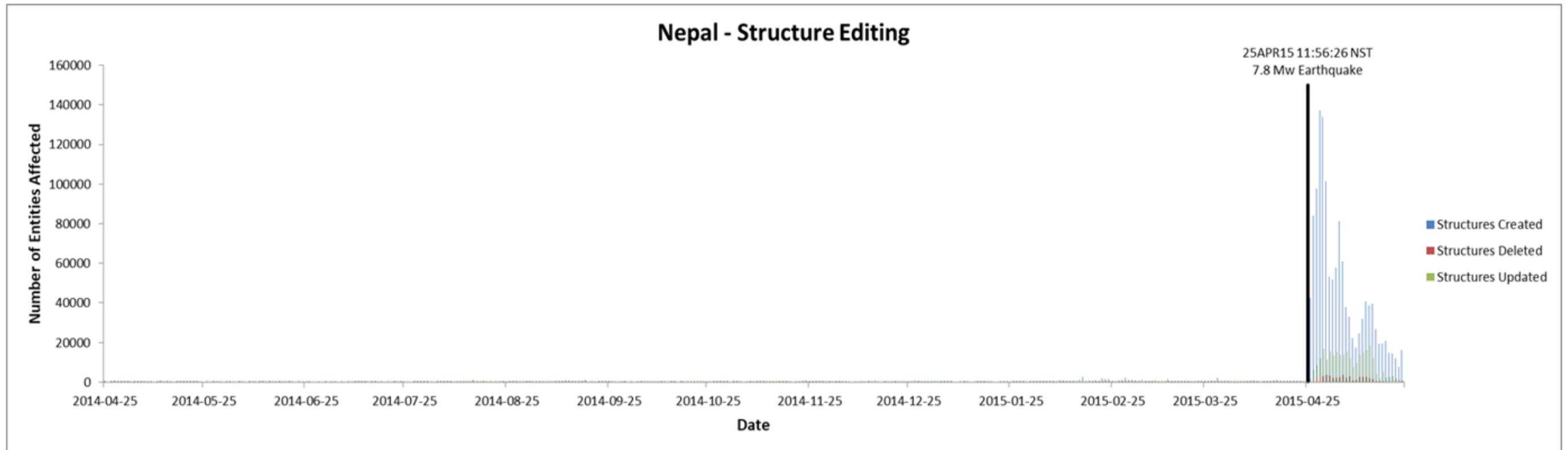
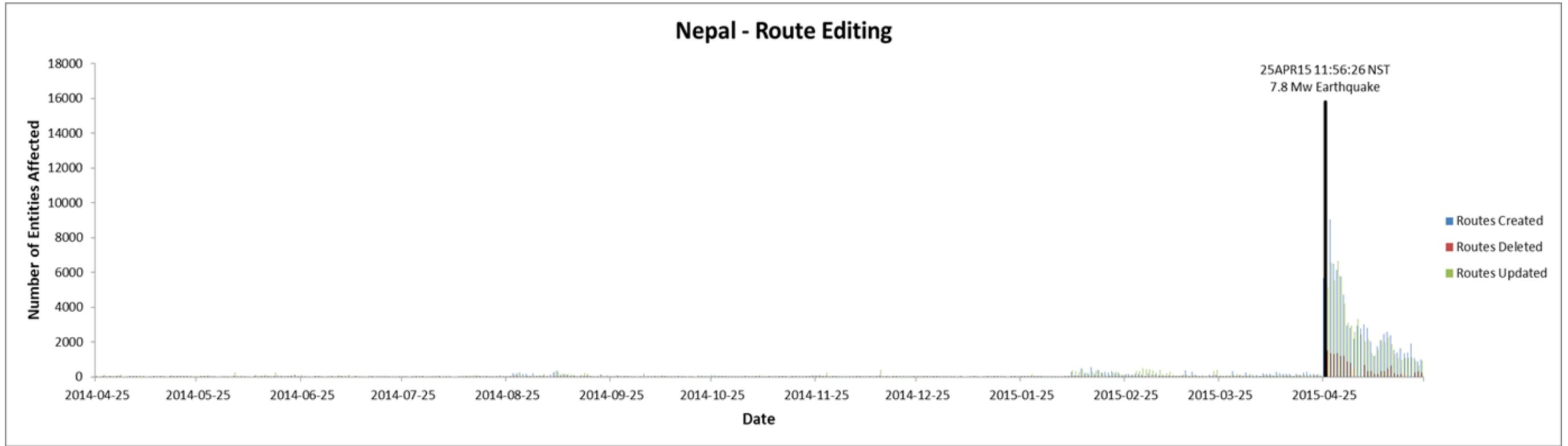


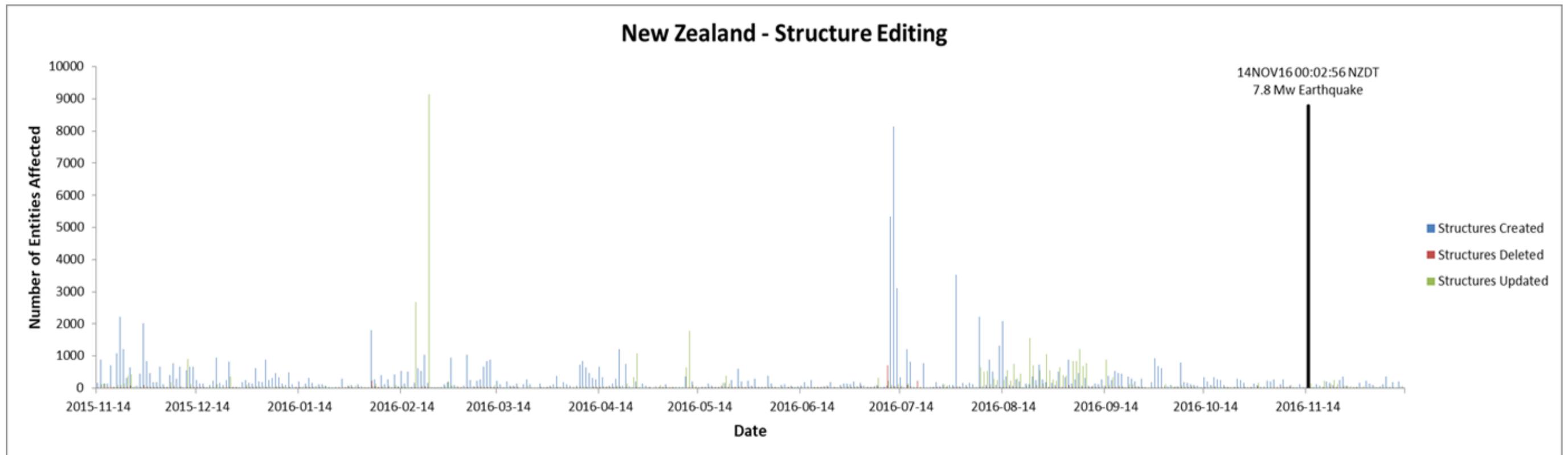
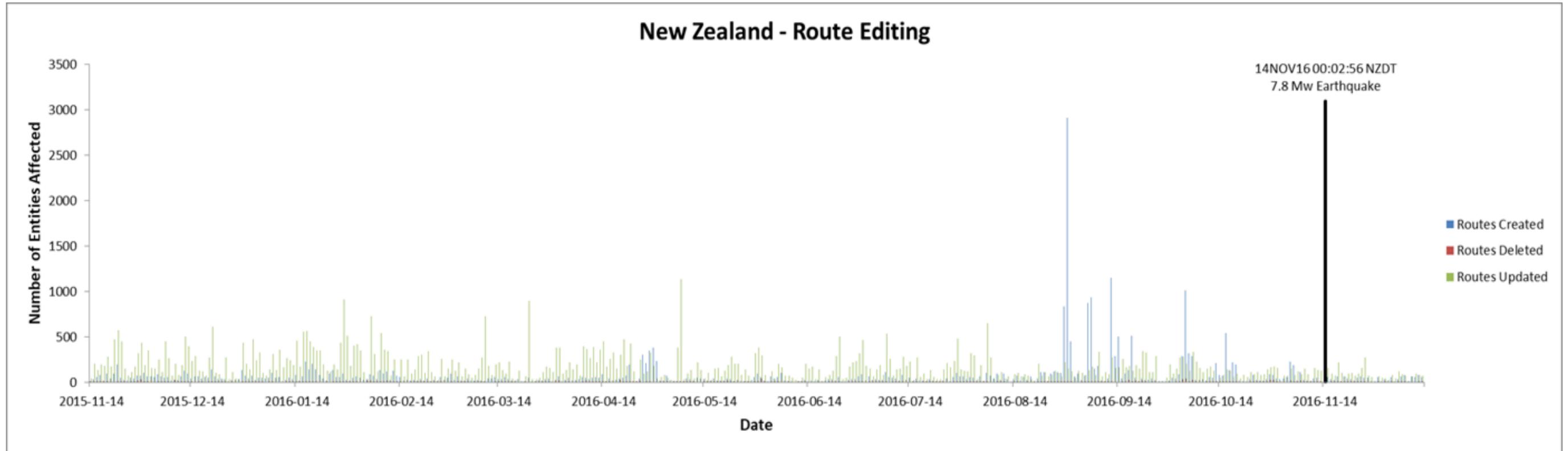
APPENDIX 2 – Actions Being Carried Out on the OSM Dataset

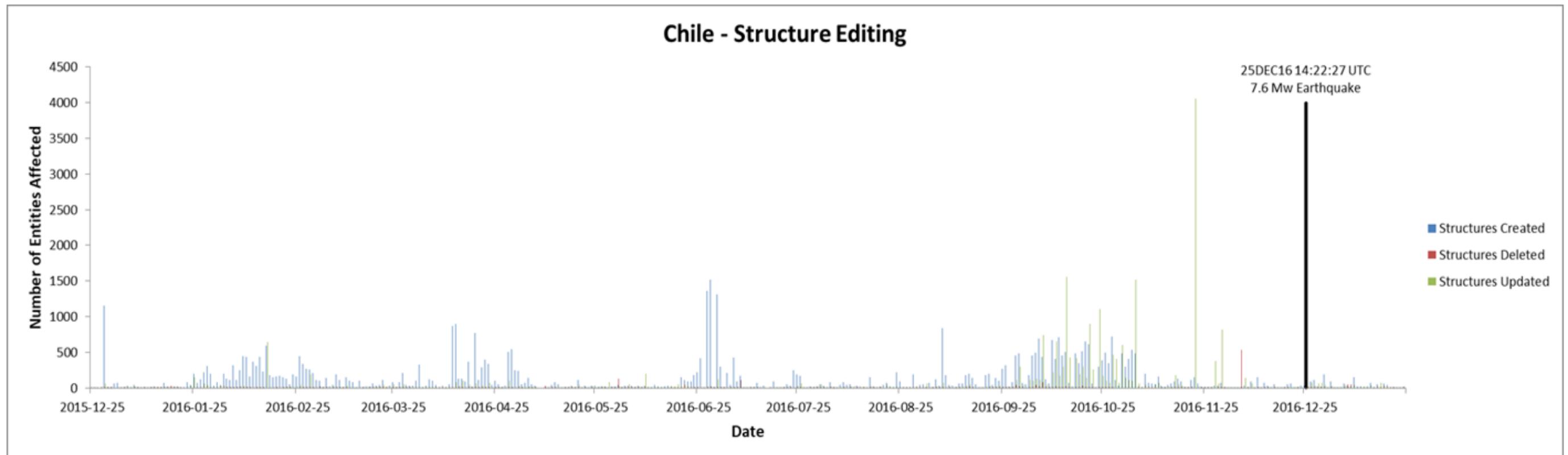
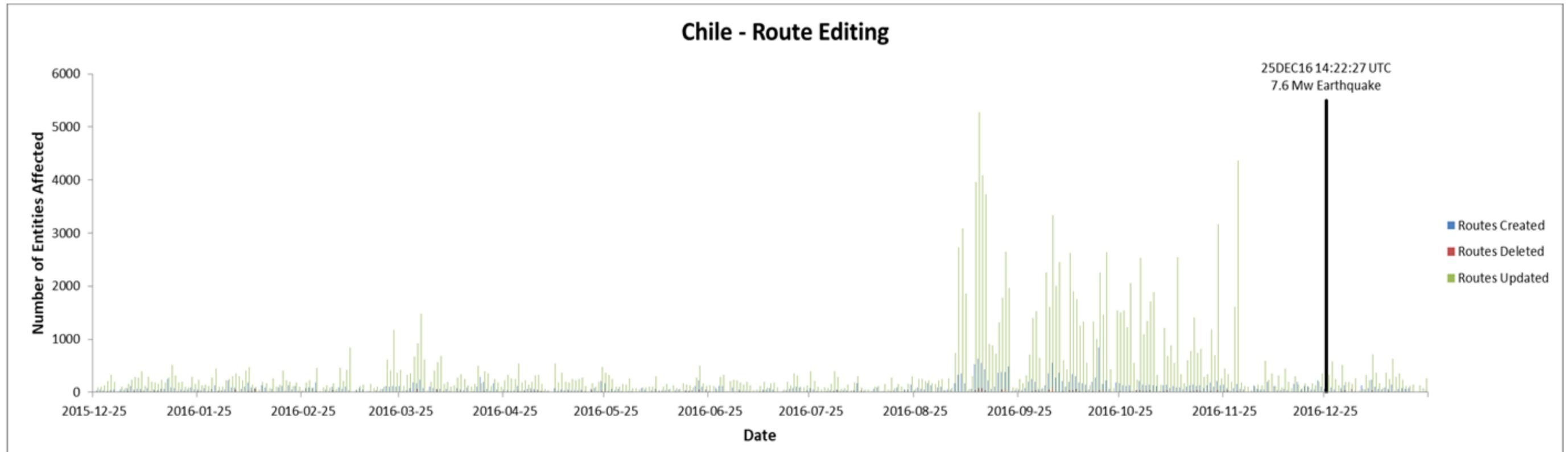


APPENDIX 3 – Entity Types Contributed to the OSM Dataset

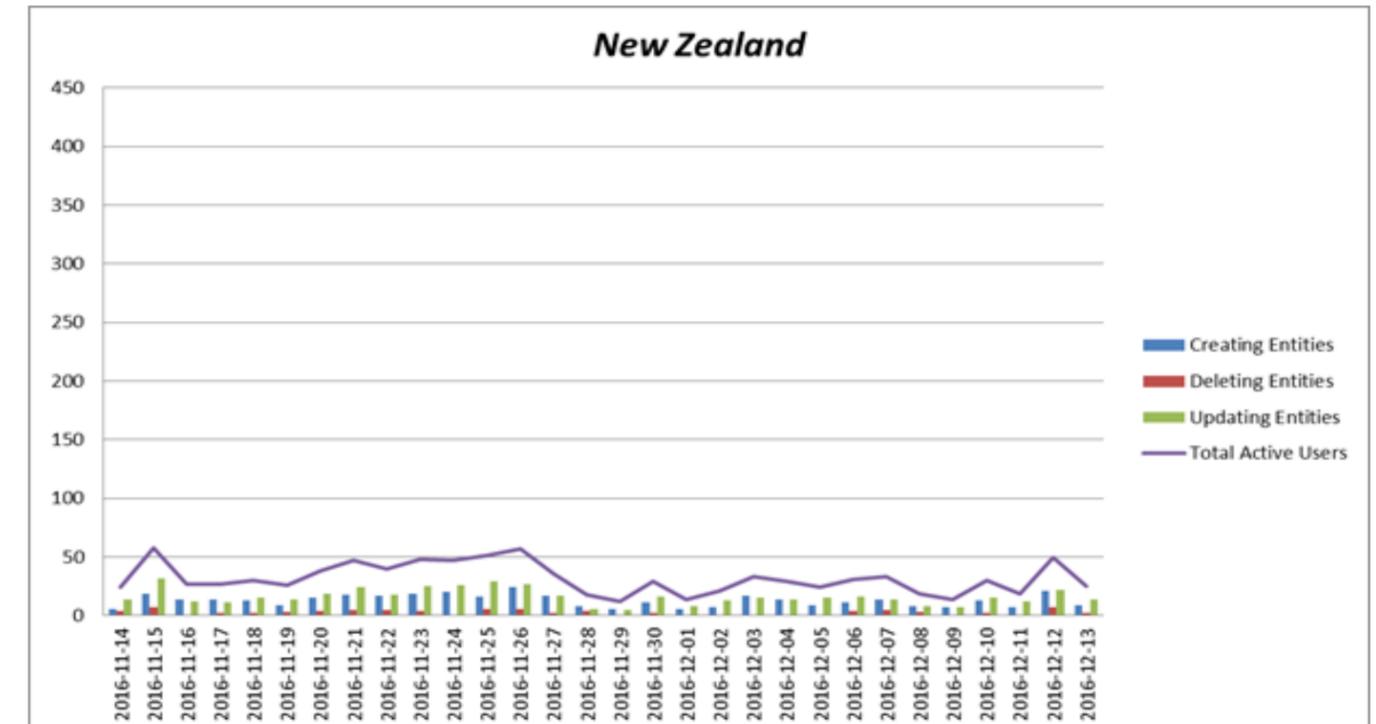
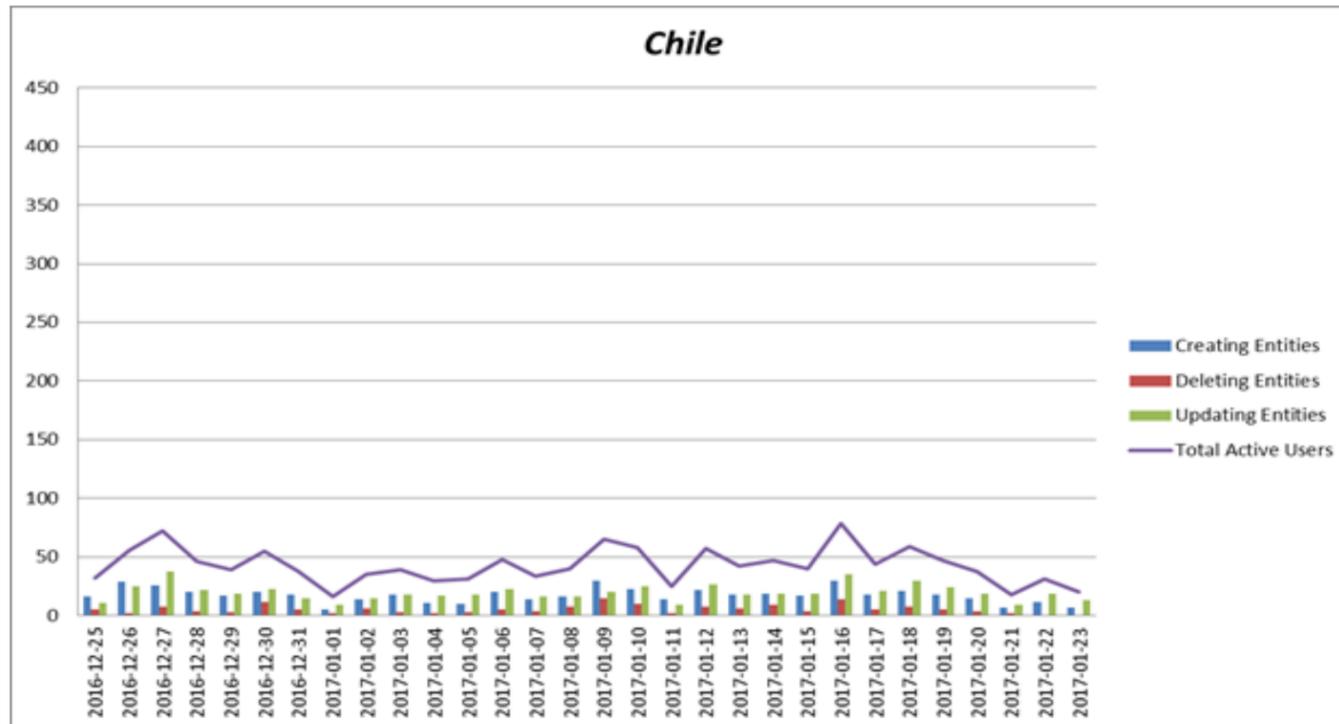
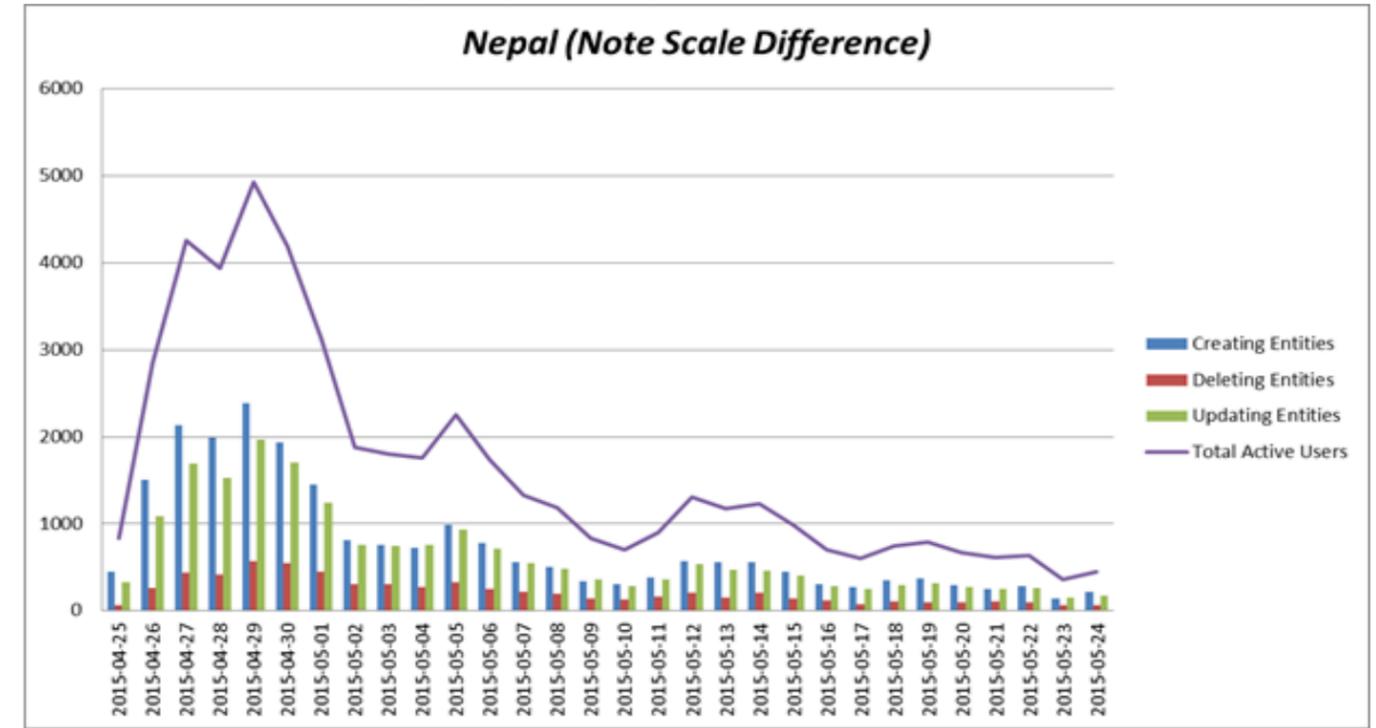
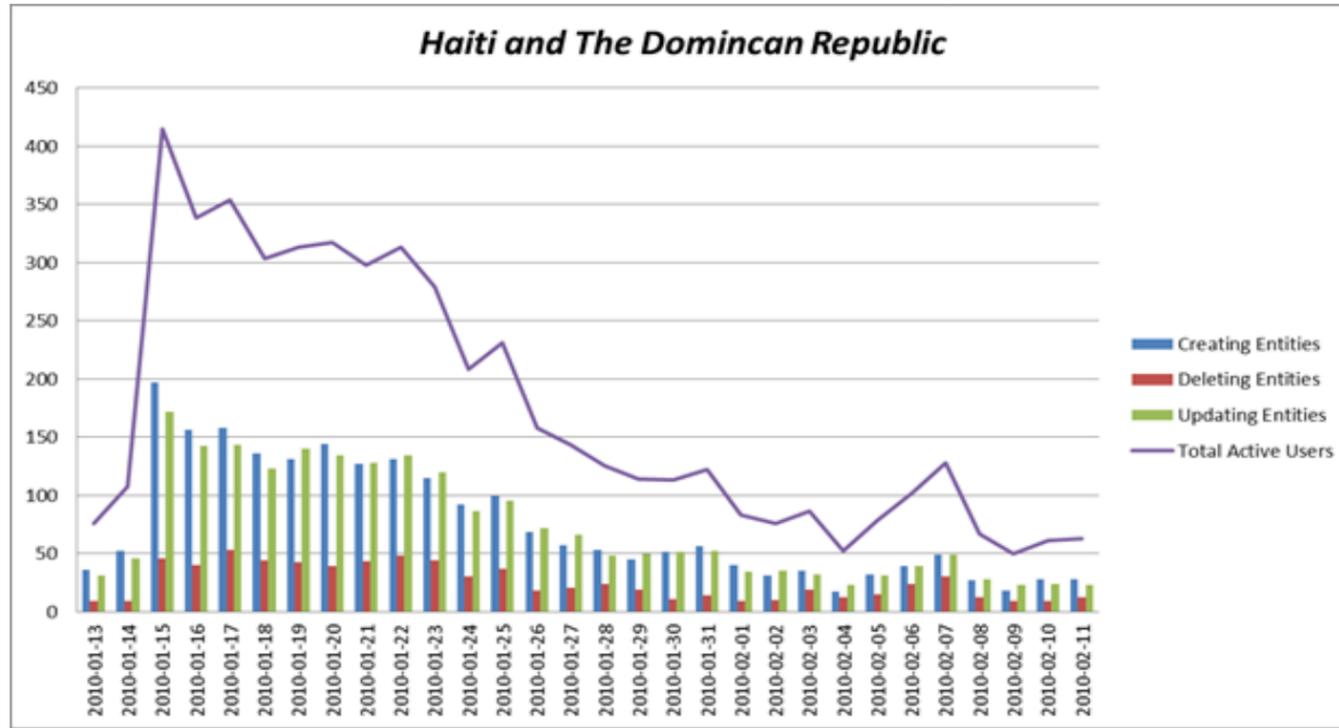








APPENDIX 4 – Unique Contributor Activities for a One Month Period Following a Major Earthquake



ANNEX B – Fictional Scenario Generated to Test Thesis Workflow

UNCLASSIFIED

3350-1
Fictional Scenario – ANNEX M TO Op EASTERN LYRE
Sept 16

Ref: A.	6 Canadian Division OPLAN 001 (Operation EASTERN LYRE)
B.	B-GL361-01/5/FP-001 Geomatics Support
C.	Geographic Coordinate System, World Geodetic System 1984 (GCS_WGS_1984)
D.	6 Canadian Division OPLAN 001 Annex M: Geomatics Operating Instructions
E.	www.OpenStreetMap.org

Distribution List
Internal: CO/DCO/RSM/OC Geo Sp/OC GI&S/Ops O
External: CFINTGFP G3/CFINTGFP G5/6 CMBG G3/6 CMBG G5/ 6 CMBG Ch Geo

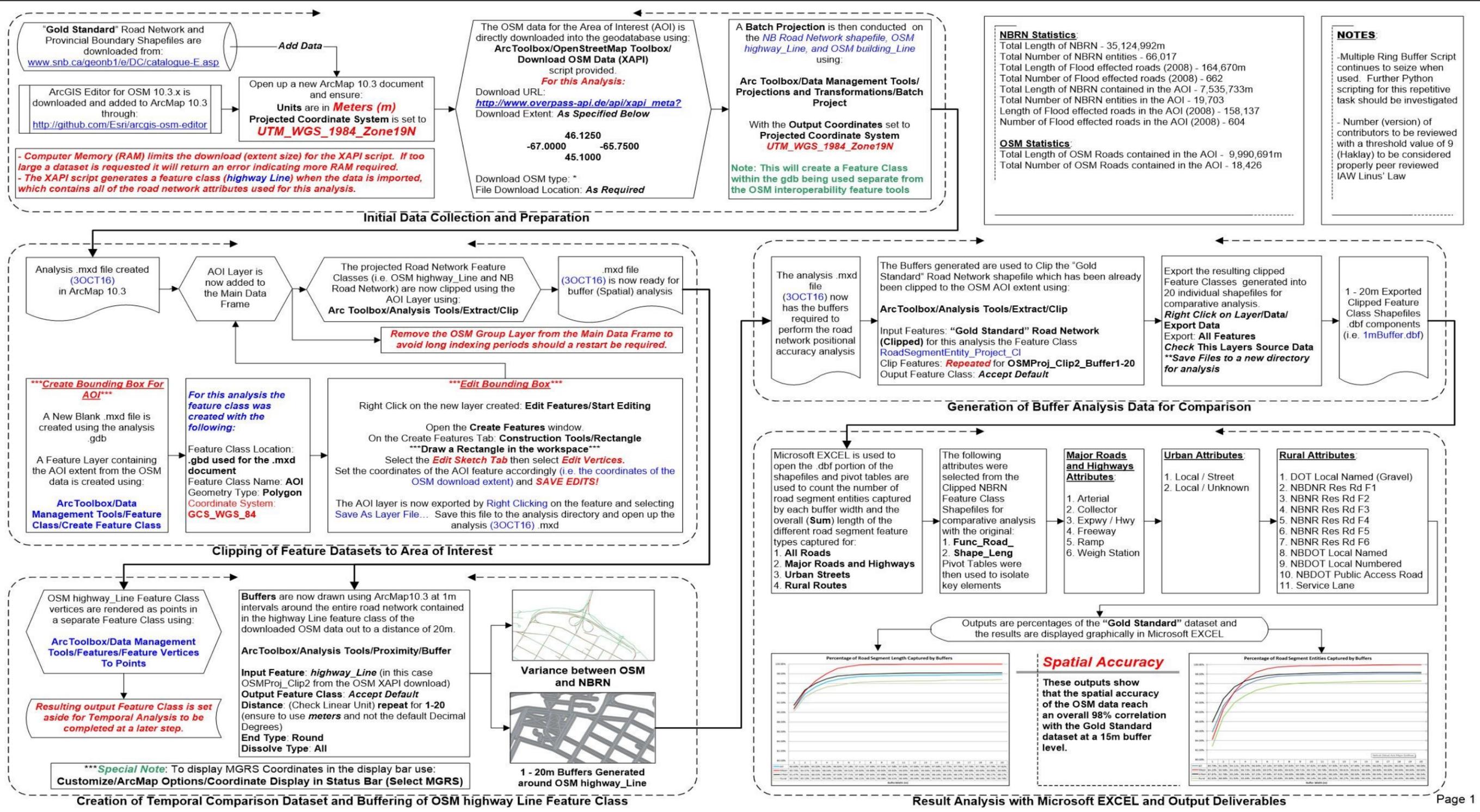
SITUATION		MISSION							
<p>General: This annex is to serve as guidance for the Geomatics Support Team (GST) assigned to Op EASTERN LYRE. As the GST, you will be the direct support provider for 6 CMBG and liaison between 6 CMBG ESCC and MCE for the purposes of providing sound planning and conduct of geomatics support to the Op. Your support and guidance will be primarily conducted at the operational level (Brigade) with a view to ensure that additional tactical level (Sub-Unit) products can be generated and provided to the supporting assets working in the Area of Interest (AOI). The role of Geomatics Support will be to analyze and interpret any available data on the flood affected area and determine Courses of Action (COA) for the efficient deployment of resources into the area. Standard 1:50,000 (Tactical) and 1:250,000 (Operational) products will be the focus deliverables as well as detailed assessment of the reliability (Spatial and Thematic) of the products provided.</p> <p>Phase 1 - Planning: For this operation, OpenStreetMap (OSM) data will be the primary source for our geospatial data requirements. Due to the requirement for the most recently available data to be utilized in this Domestic Operation, the traditional "Gold Standard" dataset provided by Geo New Brunswick will be used only for spatial and thematic verification. The on-site GST will be co-located with the 6 CMBG Engineering Support Coordination Center (ESCC) and will be responsible for producing small quantities of products for 6 CMBG use and the Canadian Forces Map Depot (CFMD) will be prepared to print and ship larger quantities of products as required with the understood lag (approx. 72h) in production and shipping times.</p> <p>Phase 2 - Deployment: Ch Geo and GST will deploy as part of the 6CMBG ESCC. Movement to be coordinated by ESCC.</p> <p>Phase 3 - Employment: Geo elements will distribute standard maps and retain a capability of creating bespoke products on demand from Comd 6 CMBG / 6 CMBG G3 / ESCC Ops guidance and direction.</p> <p>Phase 4 - Redeployment: Ch Geo and GST will re-deploy with 6 CMBG ESCC at Operation conclusion. Movement to be coordinated by ESCC.</p>		<p>6 CMBG Geo Tp will provide the GI&S to the HQ and brigade units during Op EASTERN LYRE. IOT ensure common operating picture in the geomatic domain.</p> <p>AREA OF INTEREST</p> 							
EXECUTION		COORDINATING INSTRUCTIONS							
<p>Concept for Geomatics Support: HQ 6 CMBG GST will deploy on Op EASTERN LYRE to provide comprehensive and timely geomatics support to 6 CMBG in response to domestic operations being conducted to aid in flooding relief efforts.</p> <p>Scheme of Manoeuvre: Phase 1 - 1:250,000 analysis products of the AOI to 6 CMBG G3 / ESCC Ops based on the following: a. Collection of the most current dataset possible from www.OpenStreetMap.org (OSM); b. Analysis of Temporal Variance in the AOI c. Determination of Spatial Variance between "Gold Standard" AOI dataset (http://www.snb.ca/geonb1) and OSM data. d. Determination of Thematic Variance between "Gold Standard" AOI dataset (http://www.snb.ca/geonb1) and OSM data.</p> <p>Phase 2 - Movement and set-up of Deployable Geomatics Support System (DGSS) to forward HQ</p> <p>Phase 3 - Responsive provision of geomatics products: (small scale paper and electronic) based on the dataset chosen for Op EASTERN LYRE</p> <p>Phase 4 - Tear down and movement of DGSS back to HQ 6 CMBG</p> <p>Support Requirements: 1. The HQ 6 CMBG GST will deploy with sufficient deployment & reserve stocks (basic load (BL) and maintenance load—(ML)) for 14 days usage of BL + ML requirements. 2. Resupply of products will be IAW SOP resupply demands submitted to MCE with a lag of delivery of products to be approx. 72H. 3. National support, quick response, new product requirements, and possible requirements for MCE base-plant support will be resolved by CO MCE.</p>	<p>4. Survey elements will not be employed during this Operation. Existing data sources will be the primary source of information with any additional information provided by Bde ESCC to be reviewed by ESCC Ops O prior to inclusion and release of new product editions.</p> <p>5. Map Supply: Small quantities of products to be generated by on site GST with any large (greater than 100 copies) requests to be submitted by Ch Geo through MCE Ops with production and delivery through the Canadian Forces Map Depot (CFMD).</p> <p>6. TERA and TVIs: The geomatics common operating picture (COP) will consist of data collected prior to deployment with additional information to be provided by Bde ESCC. The Land Command Support System (LCSS) will form the primary Bde COP with additional electronic and hard copy products provided for daily Command Update Briefs (CUBs). No on-line support exterior to the Bde HQ will be available for the duration of this operation.</p> <p>7. Digital Geospatial Information (DGI) Access and Management: DGSS and information/data brought into theatre will remain separate from LCSS network with access restricted to GST members and deployed Ch Geo.</p> <p>8. Advice and Liaison: OPCON of GST to Bde ESCC will be primary routing for any additional information requirements / requests with direct line technical support through MCE / Geo Sp Sqn as required.</p> <p>Groupings: See Appendix 5.</p> <p>Staff and HQ Responsibilities: Any 6 CMBG elements requiring Geo support will submit their requirements to the TF Ch Geo; requirements sent by radio shall be routed through the ESCC C/S E9B. These elements are AuthDirLn for consultation and review of Geo products and requirements.</p>	<p>1. 6 CMBG GST: Provide support to all 6 CMBG elements.</p> <p>2. 6 CMBG Ch Geo: Prioritize requirements and coordinate Geo Sp. Authority for Supplemental Geospatial Information (SGI) List</p> <p>3. 6 CMBG G3: Final Authority for priority of work.</p> <p>4. 6 CER CO: Mission specific toponym allocation authority.</p> <p>5. CO MCE (or CJOCC JEng if you switch to CJOCC): Final authority for Designated Geospatial Information (DGI) List</p> <p>SERVICE AND SUPPORT</p> <p>1. Supply and Storage Facilities. Locations. See Appendix 2.</p> <p>2. Transportation. 6 CER will provide 1 x MLVW and 1 x MILCOT for transportation of GST equipment and personnel. Movement of geomatics products to and from theatre will be coordinated by deployed Ch Geo through 6 CMBG ESCC Ops and 6 CMBG G5 Ops.</p> <p>3. FIN. All costs associated with the production of geomatics products will be through CC: 1234BC Fund: B202 GL: 1000 IO:BCB000M1234</p> <p>4. Sustainment. See Appendix 1.</p> <p>5. Stock Levels/Inventory Control. As per the execution paragraph, the GST will deploy with SOP stocks for Operations. Additional topo, hydro, and aero products can be requested through MCE Ops. a. Consumption. Requests for more than 100 copies of any given hard copy product will require a request from the deployed Ch Geo for MCE support via the CFMD. b. Replenishment. Deployed Ch Geo and GST will be self-sufficient for the duration of the Operation barring any unforeseen circumstances which will require authority of CO 6 CER action. c. Restocking. Primary distribution / consumables routing for this Operation will be through the 6 CMBG ESCC</p>	<p>For this Domestic Operation, MCE is the lead authority for data validation. Data sources to be utilized vary in collection standards and as such all data will be spatially compared prior to product generation.</p> <p>1. Release Authority: The products generated for this operation will be based on open source information / data and as such all products are to be UNCLASSIFIED and releasable to the general public.</p> <p>2. Introduction of New Products: All products will be reviewed by ESCC Ops / Ch Geo prior to general release of new editions.</p> <p>3. Datum/ Grid: For this Operation, a Geographic Coordinate System (GCS) will be employed from the World Geodetic System (WGS 84). The Universal Transverse Mercator (UTM) map coordinates for the AOI are as seen below:</p> <table border="1" data-bbox="1218 909 1470 974"> <tr> <td>-67.000</td> <td>46.125</td> <td>-65.750</td> </tr> <tr> <td></td> <td>45.100</td> <td></td> </tr> </table> <p>With Parameters: GCS_WGS_1984 Angular Unit: Degree (0.0174532925199433) Prime Meridian: Greenwich (0.0) Datum: D_WGS_1984</p> <p>4. Current Stocks: The AOI "Gold Standard" dataset is approx. 6 months old and as such will be used only as a validation when generating new products unless OSM data proves undesirable. Current reliability monitoring of OSM data in the AOI indicates that spatial accuracy is approximately 98% at +/- 10m from the existing "Gold Standard" dataset.</p> <p>5. Noted Geographic Errors: The OSM data to used is generated in the standard WGS 84 GCS format; however, the "Gold Standard" dataset utilizes a North American Datum 1983 Projected Coordinate System. This difference will be minuscule when spatial validation is conducted; however, the output products are in no way to be used for precision targeting purposes. Any noted errors in completed products are to be passed to ESCC Ops/Ch Geo.</p> <p>6. Funding of Geomatics Products: Output products will be directly added to operational budget following operation close-out.</p> <p>Command Relationships: 1. Comd 6 CMBG 2. CO 6 CER 3. CO MCE (National Coordination of Geo Sp) 4. 6 CMBG ESCC Ops O 5. Deployed Ch Geo - In-Theater Geo Advisor TF Comd 6. GST Comd</p> <p>Primary location of GST will be colocation with 6 CMBG HQ ESCC with alternate locations to be established during as per operational requirements.</p> <p>Communications: Deployed GST will have direct access to 6 CMBG LCSS assets for the purposes of collecting and disseminating Geomatics information and support. No direct link to internet sources except those provided by the LCSS Tactical Information Management System (TIMS) will be available after the point of deployment until communications in the effected AOI are restored.</p> <p>Link to National Capabilities: For any additional Air or Hydrographic Information, contacts can be found in Appendix 4.</p> <p>Reports: GEOSITREP to be provided daily to MCE / 6 Can Div ESCC daily at 2000H Z. See Appendix 3 for format.</p>	-67.000	46.125	-65.750		45.100	
-67.000	46.125	-65.750							
	45.100								
		<p>APPENDICES (NOTIONAL)</p> <p>Appendix 1. Geomatics Requirement List and Planning Tables (Locally reproduced).</p> <p>Appendix 2. Distribution Plan—Storage Location Graphics/Maps.</p> <p>Appendix 3. Reporting and GEOSITREP Format.</p> <p>Appendix 4. Geomatics Points of Contact.</p> <p>Appendix 5. Geo Sp Squadron TFM/TOR&E Requirements.</p> <p>Appendix 6. Designated Products (Map Index Coverages).</p>							

UNCLASSIFIED

APPENDIX 1 – Thesis Trial Process Workflow

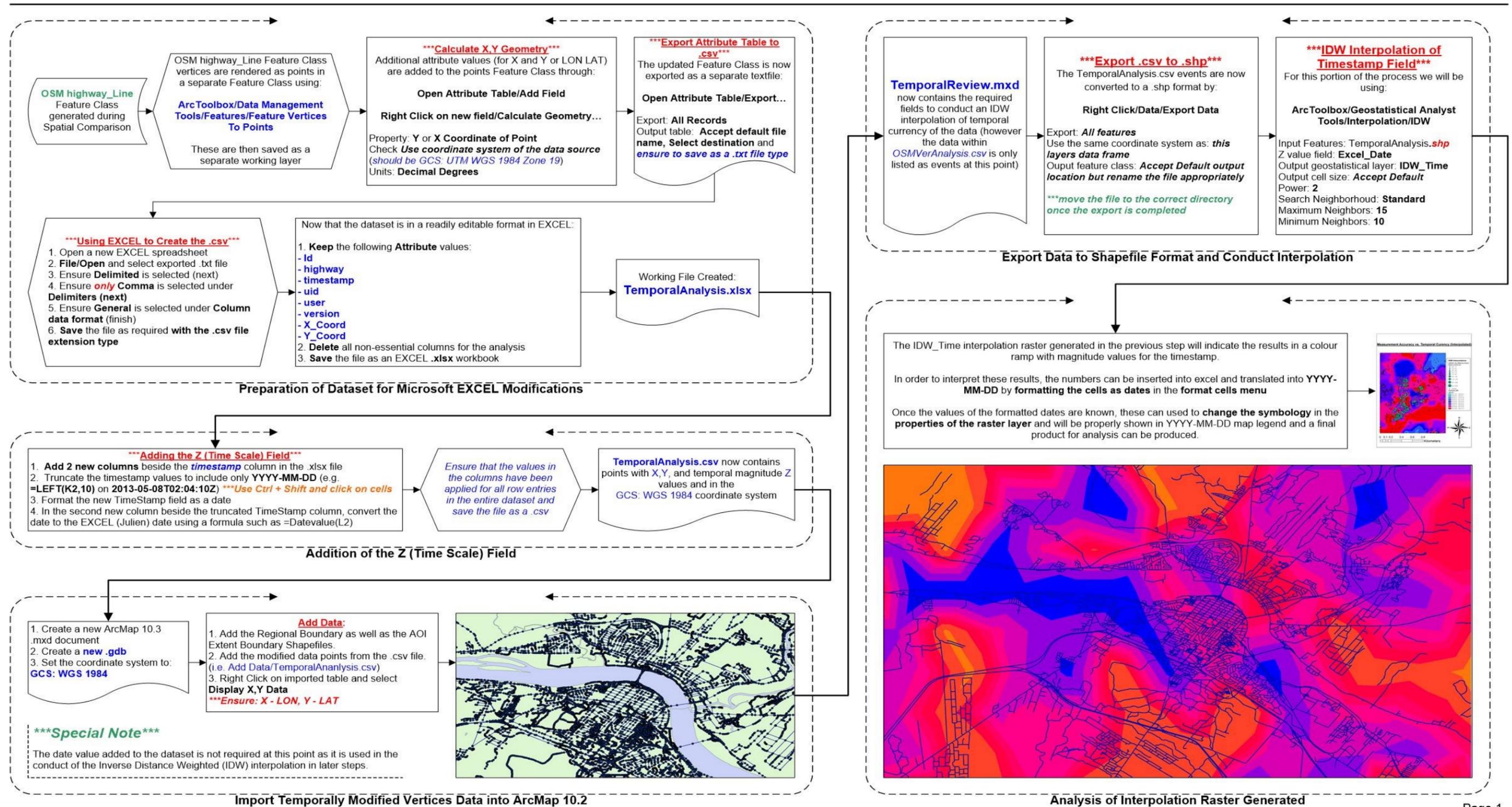
Thesis Process (XAPI) Trial Run Based on Fictional Operation EASTERN LYRE

October 18, 2016



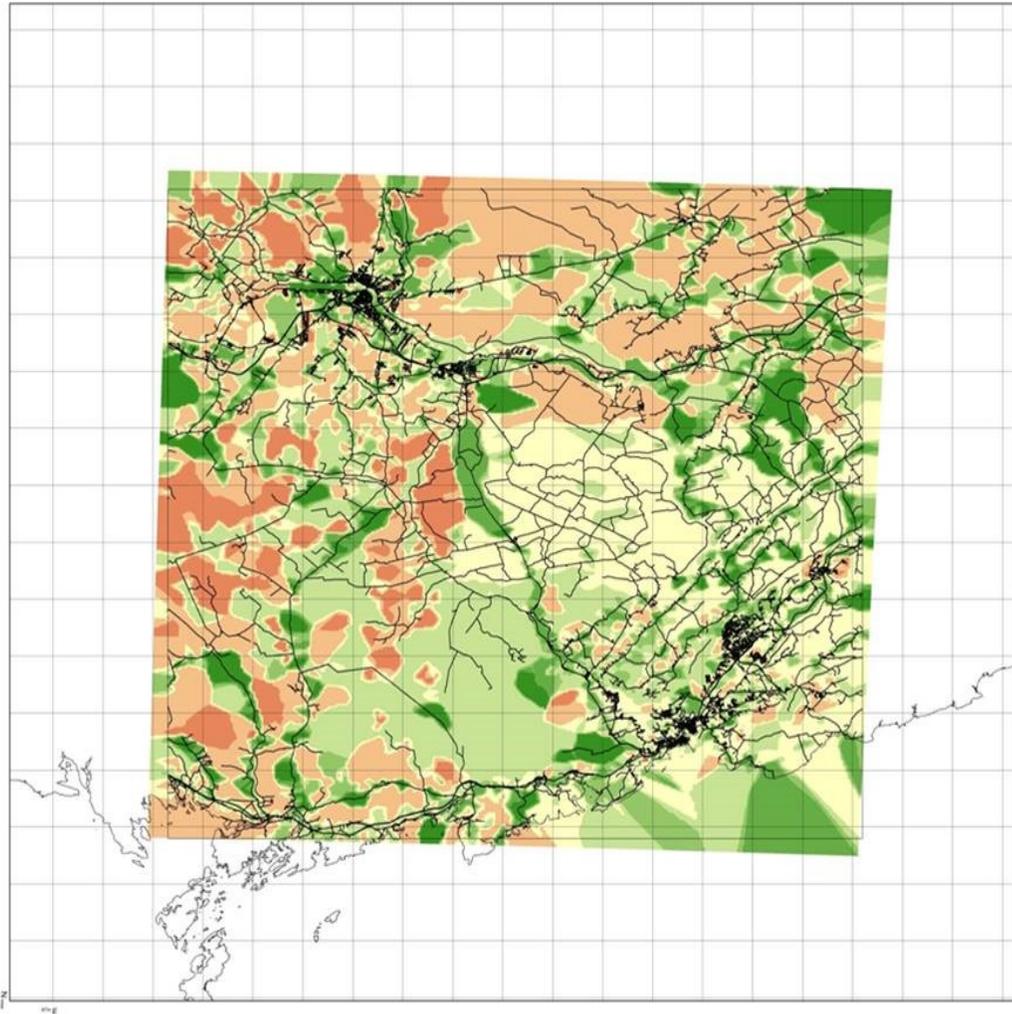
APPENDIX 2 – Workflow for Rendering of Temporal Differences Rendering of Temporal Differences Found in the OSM Dataset

June 6, 2017



APPENDIX 3 – IDW Temporal Analysis

Thesis Trial AOI - Temporal Analysis

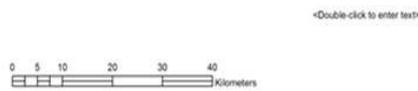


Legend

- geonb_provinciallimits-limitesprovinciales
- AOI_Project
- RoadSegmentEntity_Project_CI

Date Range

- 30NOV07 - 23NOV08
- 24NOV08 - 17NOV09
- 18NOV09 - 11NOV10
- 12NOV10 - 5NOV11
- 6NOV11 - 29OCT12
- 30OCT12 - 23OCT13
- 24OCT13 - 17OCT14
- 18OCT14 - 11OCT15
- 12OCT15 - 4OCT16

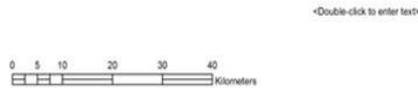
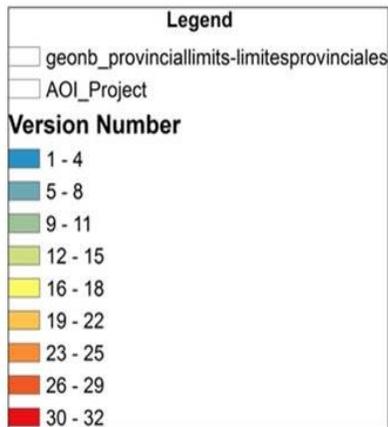
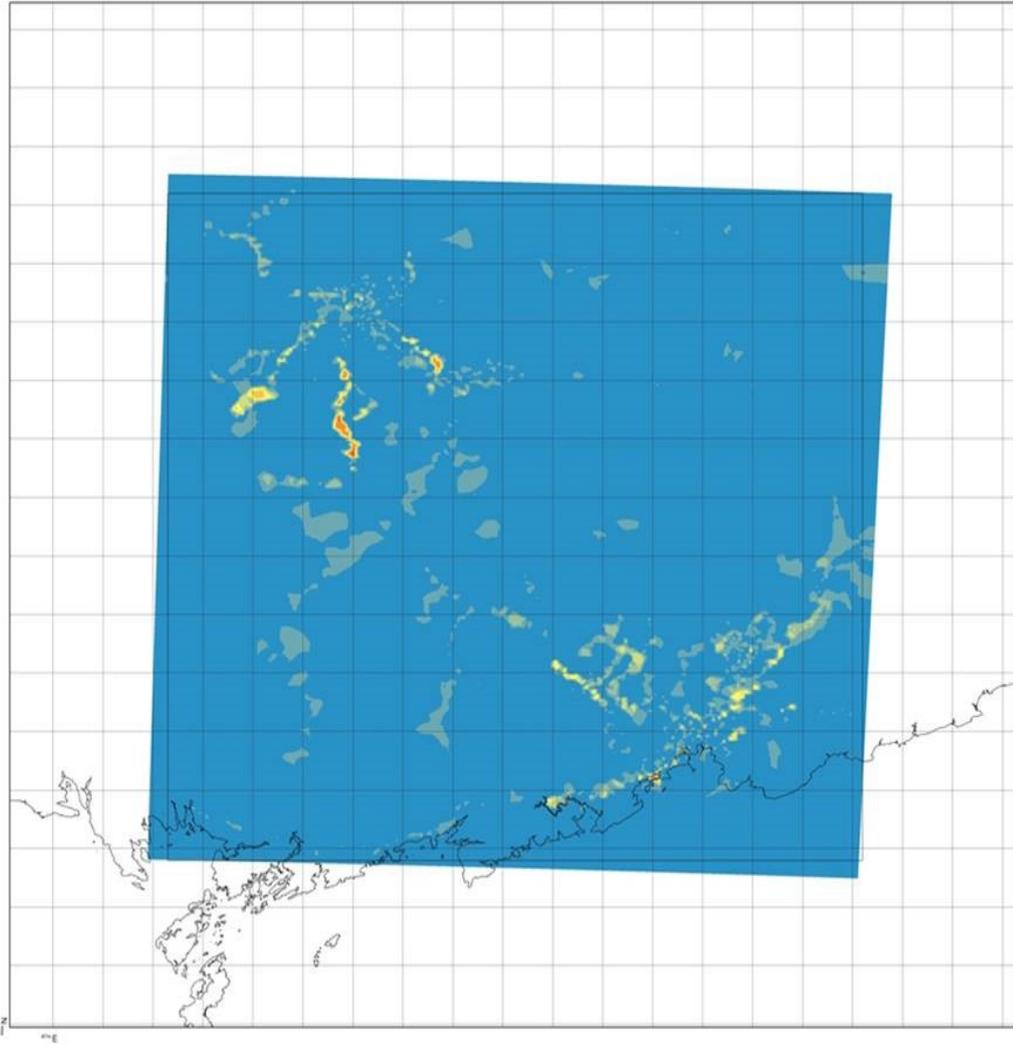


1:500,000

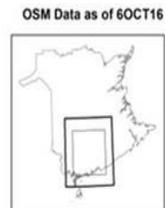


APPENDIX 4 – IDW Version Analysis

Thesis Trial AOI - Version/Peer Review Analysis

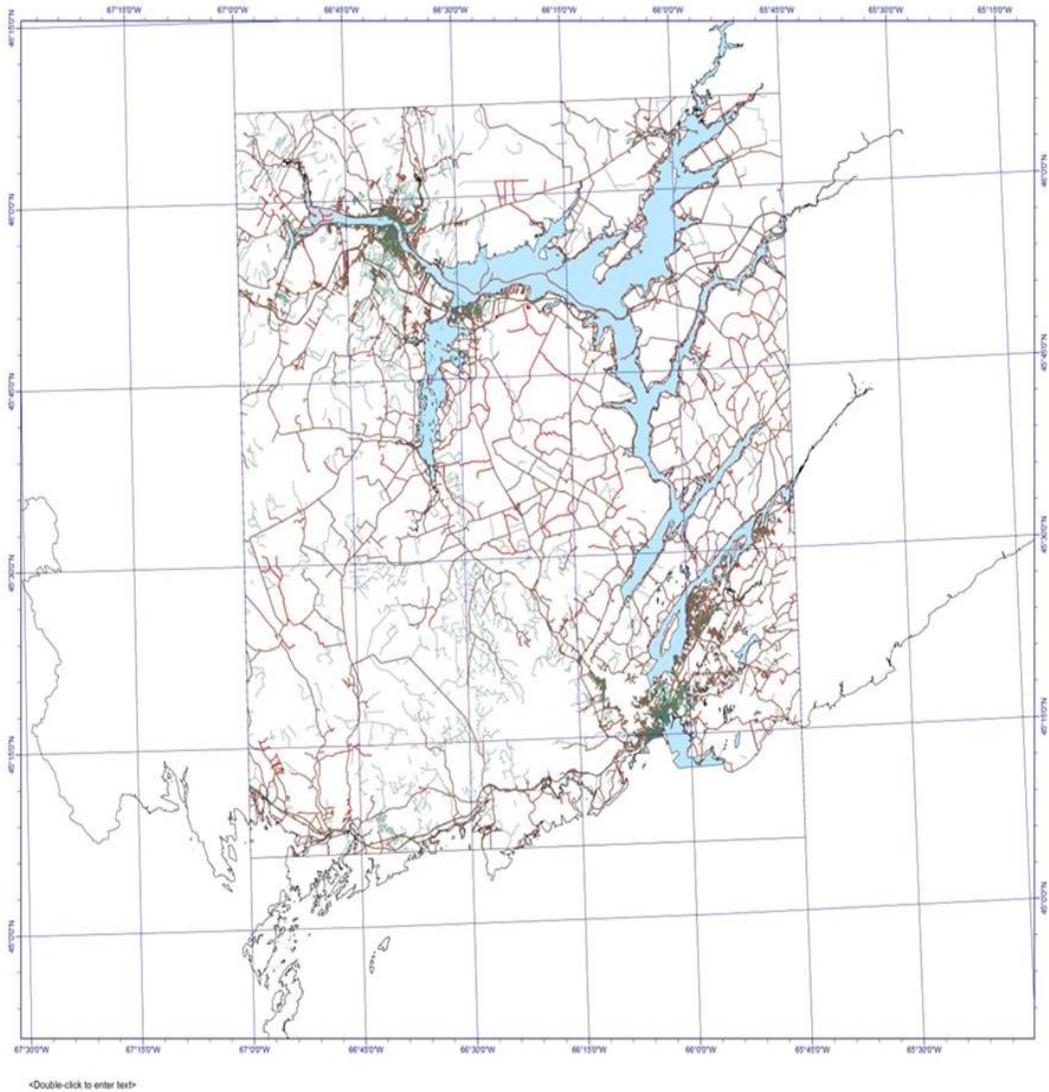


1:500,000



APPENDIX 5 – AOI used for Thesis Trial Scenario

Thesis Trial AOI



<Double-click to enter text>

Legend

-  geonb_provinciallimits-limitesprovinciales
-  AOI_Project
-  AOI
-  OSMProj_Clip2
-  RoadSegmentEntity_Project_CI
-  SDEDO_Flood_2008

