A Planning based Evaluation of Spatial Data Quality of OpenStreetMap Building Footprints in Canada

by

Nicholas Copes

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Master of Arts

in

Planning

Waterloo, Ontario, Canada, 2019

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

OpenStreetMap (OSM) is an editable world map where users can create and retrieve data. Building footprints are an OSM dataset that is of particular interest, as this data has many useful applications for planners and academic professionals. Measuring the spatial data quality of OSM building footprints remains a challenge as there are numerous quality measures that can be used and existing studies have focused on other OSM datasets or rather a single quality measure. The study performed in this thesis developed a set of ArcGIS models to test numerous spatial data quality measures for OSM building footprints in a sample of mid-sized Canadian municipalities and gain a comprehensive understanding of spatial data quality. The models performed tests by comparing to municipal datasets as well as determining other quality measures without a reference dataset. The results of this study found that the overall spatial data quality of OSM building footprints varies across mid-sized municipalities in Canada. There is no link between a municipality's location or perceived importance and the level of spatial data quality. The study also found that commercial areas have a higher level of completeness than residential areas. While the models worked well to test numerous spatial data quality measures for building footprints and can be used by others on other building footprint datasets, there exist some limitations. Certain tests that identify potential building footprint errors need to be checked to see if they are indeed errors. Also, the models were not able to measure any aspect of shape metrics. Suggestions for further studies include measuring shape metrics of building footprints from OSM as well as encouraging and subsequently monitoring OSM contributions in a particular area.

Acknowledgements

I would like to acknowledge my supervisor, Dr. Robert Feick for his guidance and support throughout the entire thesis process. I wish to thank my committee member, Dr. Peter Deadman and reader, Dr. Jane Law for their feedback on my work. I would also like to acknowledge, James McCarthy from the Mapping, Analysis and Design Lab at the University of Waterloo for providing the dated OSM layers used in the completeness over time analysis. I thank my parents and my sister for their support throughout this process, listening to me talk about things they knew little about and providing useful suggestions from a non-expert viewpoint.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

BC: British Columbia

BF: Building Footprint

CGDI: Canadian Geospatial Data Infrastructure

CSD: Census Subdivision (as defined by Statistics Canada)

ESRI: Environmental Systems Research Institute (maker of ArcGIS)

FFU: Fitness for Use

GI: Geographic Information

GIS: Geographic Information System(s)

GPS: Global Positioning System(s)

HOT: Humanitarian OpenStreetMap Team

ISO: The International Organization for Standardization

LULC: Land Use and Land Cover

OSM: OpenStreetMap

SDI: Spatial Data Infrastructure

VGI: Volunteered Geographic Information

Chapter 1: Introduction

1.1: Overview

Spatial data quality is a complex issue which many people have difficulty understanding. This is because the measurement of spatial data quality can be done in a variety of ways and there is still some lack of standardization when it comes to defining measures of spatial data quality. Of the measures of spatial data quality, Devillers et al. (2007) note that fitness for use (FFU) is an important concept. FFU is concerned with users of geospatial data being able to understand how the dataset fits the intended use. It concerns the match between the data's characteristics and the user's requirements for a given task. The creators must ensure that the intended users understand the data and how to interpret quality. The fact that each dataset is different also leads to issues in terms of standardizing the way spatial data quality measures are performed (Devillers et al., 2007).

Volunteered Geographic Information (VGI) is free spatial data that is created by "volunteers" such as amateur geographers and citizens providing spatially referenced data. In particular, VGI has its own unique set of challenges in terms of assessing spatial data quality. As anyone can create VGI, there is no assurance of quality or knowledge of the creator's expertise. Kalantari et al. (2014) also note that VGI often does not contain data describing where it came from or how accurate it is, making it difficult for users to assess FFU.

OpenStreetMap (OSM) is a popular form of VGI and has many contributors. Created in 2004, it is an editable world map which gains data through crowdsourcing. OSM allows anyone to create and edit features such as buildings, roads and points of interest anywhere in the world. OSM has its own unique challenges in terms of measuring quality. Haklay (2010) notes that no assumptions can be made about the quality or background knowledge of contributors. This thesis focuses on measuring the quality of datasets from OSM, in particular, building footprints. Building footprints refer to polygons that show the outline of a building in its perceived real geographic location. They are important data that can be used for many types of location analyses including determination of density or catchment areas. There are a variety of measures that can be used to assess spatial data quality of OSM datasets. The International Organization for Standardization (ISO) in their Standard 19157, notes many aspects of data quality including positional accuracy, completeness, commissions (excess features) and thematic accuracy (level of attributes). While these aspects of data quality can be applied to many forms of spatial data, in this study they will be optimized for OSM evaluation. Authors such as Fan et al. (2014) have developed tests to measure such aspects of quality for data from OSM; However, their methods can be difficult to understand for the average person or planner as they involve complex formulas and specialized software. A lack of studies have been performed on building footprints. Many studies such as those by Boeing (2017) focus instead on the quality of roads from OSM. Additionally, other studies focus on trying to improve the quality of building features by modifying them through generalization (Pászto et al., 2015) or squaring (Lokhat and Touya, 2016) instead of developing new ways to measure their quality.

As the OSM user base is rapidly expanding, it is important to ensure that the data available meets a certain level of quality. OSM data are being used more frequently and have great social value. Barrington-Leigh and Millard-Ball (2017) note that OSM is particularly beneficial in low-income countries where government data is often unavailable. In many cases, it is also more accurate and complete than government data. Thus, it is important to ensure that this data is of quality since it is heavily relied upon. For planners, OSM data also brings great benefit. Boeing (2017) notes that OSM road data can be used for planning analysis and recommendations for roads improvements, bike lanes and more. Building footprints allow planners to have an inventory of buildings in their municipality which is often non-existent or outdated in government databases. Planners can use OSM datasets in place of non-existent municipal ones or simply to compare datasets and ensure accuracy for local citizen use.

The approach presented in this thesis for evaluating the quality of building footprints from OpenStreetMap (OSM) aims to combine existing methods and new ones into an easy to understand and easy to implement solution. This will use a combination of methods including an ArcGIS model which can be used to measure the quality of building footprints in comparison to a reference dataset by performing various tests. An additional model with different tests is used to evaluate the quality of OSM building footprints where a reference dataset is not available. This model is helpful because reference datasets are not always available, but certain spatial data quality measures can still be performed. By dividing the tests into two models, the second model can be used in areas without a reference dataset, but also in conjunction with the first model, when a reference dataset is available, thus giving a more complete quality assessment. The tests that the models perform calculate many of the spatial data quality components identified by the International Organization for Standardization (ISO) and academic literature including completeness, positional accuracy and thematic accuracy.

1.2: Research goals and objectives

The first objective is to develop a simple to use set of models to evaluate the spatial data quality of building footprints from OpenStreetMap. This approach is designed to be replicable and will be used to evaluate building footprint datasets across various mid-sized Canadian municipalities. The models are designed to be used by planners and other GIS users.

The approach developed in this thesis aims to have a comprehensive evaluation of building footprint quality. Also, it simplifies the existing approaches used by others by allowing the tests to be performed using ArcGIS models. By developing an approach that uses common shapefiles and can perform all quality tests in the widely-used ArcGIS program, it can be replicated by others. It is important to understand the quality of building footprint data in order to ensure accurate analysis. The approach aims to help the understanding of building footprint quality and to introduce methods that allow for a better comprehension of spatial data quality in general. Through the approach, users will be able to see with their own data the level of quality including errors, offsets and commissions, which they can then choose to investigate or improve. The objective of this method is to design a simple model that can be reused to measure OSM building footprint quality against a reference dataset.

All of the model tests have been designed to be aggregated in a workflow; however, they can also be performed separately. Tests were chosen that were relevant specifically to building footprints and that aim to measure quality and difference between datasets. This is different from other studies that try to “improve the quality” by using tools, rather than develop new ways to measure quality.

In addition to developing a model to compare the quality of OSM datasets to reference datasets, another model is used to evaluate OSM building footprints when a reference dataset is not available. The goal of this model is to help people understand whether or not the OSM dataset is reliable enough to use when there are limited alternatives. This is, in essence, a measure of the fitness for use of these building footprint datasets when a reference dataset is not available.

The second objective is to understand how datasets from OpenStreetMap can be useful to planners. This is done through an investigation of the literature and by making links between existing uses of OSM data and what a planner could do with such data. Furthermore, identifying the link between assessing the quality of the data and ensuring the necessary fitness for use of the data for planning purposes.

The third objective is to determine the level of quality of building footprints from OpenStreetMap for a small sample of mid-sized cities in Canada and investigate reasons for variations. An investigation will be done for ten Canadian cities to note the level of completeness, accuracy and overall quality in different areas of the city and determine the level of contributions over time. Making links between a city's location, perceived importance, local factors and the level of quality and completeness will be discussed.

Chapter 2: Literature Review

This chapter provides the current state of research in the area of spatial data quality. Section 2.1 describes VGI as a source of spatial data. An overview of VGI and examples are given. OpenStreetMap is then described as an example of VGI. Section 2.2 introduces the concept of Spatial data quality. The International Organization for Standardization's measures of spatial data quality are described, along with methods for assessing spatial data quality of authoritative GI, VGI and OSM. Contributor motivations and evaluations are discussed in section 2.3. In section 2.4, the importance of VGI and OSM for planning is highlighted. This section discusses the need for geospatial data at the municipal level as well as the value of VGI for planning the quality requirements of this data for planners' and citizens' use. This section concludes by stating how VGI and OSM are important sources of data for planners and the need to be able to understand and evaluate spatial data quality to ensure accurate analysis.

2.1: Volunteered Geographic Information as a source of spatial data

- 2.11 VGI and its Uses

Definition of VGI

Goodchild (2007) coined the term Volunteered Geographic Information in 2007. He used the term to refer to the extensive engagement of residents, often without proper training, in the creation of geographic information, which for many years had been reserved for government agencies. Contributors are usually untrained and contribute voluntarily; their results may or may not be accurate. Goodchild felt that this shift in the creation of geographic knowledge would surely have a great impact on the relationship between GIS and geography in general and the general public. Feick and Roche (2013) note that there is no consensus on the definition of geographic information (GI). VGI is very heterogeneous in nature and can include personal data such as geotagged photos, passively contributed information such as cell phone tracking or even quasi-scientific data such as locations of animal sightings and amateur weather station recordings.

VGI is produced through the process of crowdsourcing. Haklay et al. (2014) note that crowdsourcing is "the process of obtaining information from many contributors amongst the general public, regardless of their background and skill level" (p.7). VGI is thus a form of crowdsourcing of geographic data from non-expert individuals. Sui et al. (2013) note that in the past few years there has been a transformation in the way geographic data and knowledge is produced and distributed. There have been a wide variety of technologies including Web 2.0 that have changed the way data is shared. Fast and Rinner (2014) note that the geospatial Web 2.0 or Geoweb "is a collection of online location-enabled services and infrastructure that is engaging a wide range of stakeholders in mapping processes" (p.1279). There have also been different terms used to describe this phenomenon including VGI and crowdsourcing. The general idea revolves around the use of the Internet to create, share and analyse geographic information over multiple computing platforms including computers, tablets and smartphones. They also note that the use of Geoweb platforms by individuals can lead to the creation of VGI. There are also additional processes used in the creation of VGI, which will be discussed in this chapter.

Sui et al. (2013) note that new technology has made it possible for anyone to become a geographer. They note that VGI represents a major shift in the content, characteristics, creation and sharing of geographic data. New cyberinfrastructure allows partnerships to be formed between governments, NGOs, industry, business and citizens. These organizations work with citizens to create projects that allow people to contribute geographic data all over the world, to help those in need. The authors note that VGI produced through crowdsourcing can now be relied upon to engage a new mode of geographic knowledge which will create a more knowledgeable, efficient and sustainable world (Sui et al., 2013).

Feick and Roche (2013) note that there are differences in the way that VGI and authoritative data are produced. VGI is produced by a large number of individuals with varying levels of interest and ability, whilst authoritative data is produced as a result of activities by experts. VGI participants can be both producers and spatial data users, engaging in both roles at different times. The production and use of VGI are loosely organized and is not constrained by market forces or regulatory standards that authoritative GI is subject to. Authoritative GI is created by the government or private sector. Government GI is designed to streamline government operations, such as development control, whilst the private sector aims to fill the GI market to make a profit (Feick and Roche, 2013).

The creation of VGI is not subject to the same standards as when authoratitave GI is created. Authoritative data must conform to a variety of standards when it is created such as the National

Standards for the Survey of Canada Lands. This standard and others are described in section 2.22. WhileVGI is not subject to data standards, there are often suggestions on how to contribute or informal standards posted by the project organizers to help ensure data quality and that the data meets the FFU for the project. OSM Wiki notes various suggestions for how data should be created, while organizations such as the Humanitarian OpenStreetMap Team (HOT) list instructions on various projects to help ensure quality. These suggestions and informal standards are described in section 2.25.

VGI is created in a variety of ways by different contributors. For example, anyone can contribute to OSM. OSM data is created by individuals who can digitize buildings, roads and other features anywhere on a map of the earth. Creators can also add attributes to their creations and edit existing data. This process is described in section 2.12. Other forms of VGI are produced by individuals who share geographic data in an app, such as the City of Edmonton 311 app or those who pinpoint features on a map and provide comments, such as in a project by the City of Saint John (both examples described later in this section). VGI data is created for a variety of reasons. First of all, it provides free data, which can allow interested organizations or governments to gain input on how people feel about a certain topic, in a cost-effective manner. It is also created as an alternate source of data. For example, OSM provides data around the world, often in areas where authoritative data is limited (Barrington Leigh and Millard-Ball, 2017). In contrast, authoritative GI is created for government operations (such as zoning), so that governments can have an accurate set of data for their purposes. Alternatively, authoritative GI can be produced by the private sector to sell to governments or interested organizations, who require accurate data. The data creation process for VGI is driven by interested individuals who want to contribute to the breadth of data in their area or help with humanitarian mapping, both of which can be done in OSM. Also, VGI projects are driven by governments and organizations who want to gain citizen input on a particular topic. Interested citizens can have their opinions heard and governments gain valuable information. The data creation process for authoritative GI is driven by government need for local data, or if an opportunity exists for a private company to create data to fill a gap in the market.

VGI has some benefits over authoritative GI. It is worth noting that VGI can often be more current than authoritative data. For example, OSM is constantly updated and can contain more buildings or features than a government dataset that is a few years old. One example is demonstrated in Figure 4.3, which shows recent commercial buildings in Niagara Falls that were not included in the government dataset (commissions). This also means that VGI data quality can vary constantly as new data is created or existing data is edited. While VGI has a world of creators who can contribute data, governments and companies that create authoritative data often have limited resources. The data they produce may only be maintained once every year (or other specified period) and new versions of data are often infrequently released, meaning that the data quality and completeness remain static over time until a new version is released. As an example, Appendix 4 notes the metadata for the datasets used in this thesis, many of which note the timelines for maintenance and updates.

VGI can be used for a variety of important purposes. Teymurian et al. (2013) investigate how VGI can be used to improve public transit. They note three ways in which public transit users can improve the transit experience by using VGI. The first is information provision, in which users collect and share real time data such as traffic, the bus location and unexpected events so that others can better plan their trip. The second is planning, in which transit users can rank proposed transit plans, show the locations of preferred transit stops and share opinions about time tables. The third is monitoring, in which users can report problems, evaluate the system and propose solutions (often with geospatially referenced data) (Teymurian et al., 2013).

Horita et al. (2013) notes that VGI is often used in disaster management. Their investigation of literature found many examples of VGI being used primarily in the response phase, across a variety of natural disasters, with fire and floods being the most managed disasters using VGI. They note that the use of VGI in disaster management is growing and that social media was the primarily way of sharing VGI for these purposes. Roche (2011) noted that after the Haiti earthquake, a platform was set up to receive relief needs. Furthermore, basic mapping was performed in the area as there were no recent government maps. Volunteers from around the world pooled together to map out the area, providing the only current GI for the area. In many areas, government GI is not available and thus VGI provides the most accurate mapping data for these areas. Such examples demonstrate the immense social value of VGI. This ties in the work of the Humanitarian OpenStreetMap Team (HOT), which lists projects that VGI/OSM contributors can work on, in response to natural disasters around the world.

VGI is used by local governments as a way to gain valuable citizen input on a variety of issues. Many cities have developed apps for citizens to report issues. For example, the City of Edmonton 311 app, allows users to report potholes, roads/sidewalks needing snow cleared, floods, litter and vandalism. Apps such as this allow users to take pictures and provide location information relating to their issue. It also allows the city to gain information on what issues need to be addresses and which ones are important to local citizens. Municipalities are also using VGI to assist in a number of planning projects. For example, the City of Saint John, NB allowed residents to become involved in their Central Peninsula Neighbourhood Plan. Residents could pinpoint areas on a map, select a category (such as could be better, works great, what's missing) and provide comments. Projects such as these allow cities to gain valuable input from citizens about issues that are important to them.

VGI has some limitations which users should understand. VGI does not have strict standards to ensure quality when it is created. Feick and Roche (2013) note there is an absence of market forces and professional standards in the creation of VGI. The measurement of VGI quality is thus variable, whereas standard measures can be applied to authoritative GI. Those who create VGI for personal or limited use have little incentive to document data quality or provide metadata. Mature VGI projects, such as OSM, have more documentation and allow for inspection of individual edits. For example, OpenStreetMap Changesets allows users to see the history of OSM edits for a particular feature. It is worth noting that VGI quality is constantly changing over time. Some users may input data with many errors, while others may improve that data later on. All of these limitations are discussed in further detail throughout this thesis.

- 2.12 OpenStreetMap as an example of VGI

OpenStreetMap is one of the most successful examples of VGI. OSM was founded by Steve Coast, an MSc student in 2004. Around that time, ideas around crowdsourcing were gaining momentum. Coast had a simple idea: "if I collect geographic data about my area – where I have local knowledge – and you collect geographic data about your area – where you have local knowledge – then these can be combined, and we can begin to build a spatial database of a region. If this scales up to a larger *crowd* of people, then it is very possible to crowdsource the mapping of the entire world." (Mooney and Minghini, 2017, p. 38) The focus of OSM is not on cartographic outputs, but rather an editable spatial database containing geographic data and information from around the world. OSM has been growing in popularity for a variety of reasons. The availability of low-cost, high-quality GPS data allows consumers and citizens to collect geographic information on their smartphones or other devices and upload these data to OSM. OSM is made up of citizen contributors. That means that anyone from anywhere in the world can sign up and take part, from beginner to expert (Mooney and Minghini, 2017).

OSM has the aim of building and maintaining a free editable world map, thus ensuring that users are not restricted by copyright and license. OSM started by focusing on roads and streets, but now contains a large variety of geographic objects including buildings, land use and points of interest from around the world which have been mapped by thousands of volunteer contributors. OSM has changed the way spatial data is created and shared. It is no longer limited to cartographic experts. OSM allows users to contribute to and access real-time updated maps of the world. Furthermore, users can access the history of mapping activity for an area and collaborate with other OSM users (Mooney and Corcoran, 2013).

OSM data is made up of various elements. The OSM Wiki Beginners Guide 1.3, Understanding OSM Data (https://wiki.openstreetmap.org/wiki/Beginners_Guide_1.3), provides the following description of OSM data elements:

- Nodes are points used to indicate locations. Nodes are either separate or connected.
- Ways consist of a connected line of nodes. They are employed to create paths, roads, rivers and other line features.
- Closed ways form a closed loop and are used for areas.
- Areas consist of filled in closed ways. An area is often implied when creating a closed way.
- Relation can be used to make more intricate shapes, or to indicate elements that are related but not actually connected.
- Elements can contain tags. A tag is a key=value pair that portrays what the element is. For example, mapping a mobile phone store can be done by making a node and supplementing it with the following tags: shop=mobile_phone. name=John Smith's phone centre.

Mooney and Minghini (2017) describe ways as being polygons and polylines and relations as being a logical collection of ways and nodes. A way contains either two nodes (polyline) or three nodes (polygon). A node represents a point feature and its geographic coordinates, usually as latitude and longitude. Every OSM object must have a key/value pair which corresponds to an attribute or tag and is used to describe its characteristics.

Ballatore and Mooney (2015) note that any tag can be assigned to any object and that users are free to create their own tags. There are many tutorials and services designed to teach how to start using OSM and to explore tags. The taginfo service is one example that allows users to better understand the structure of tags and conceptualize a wide range of key/value pairs as well as see the spatial distribution of tags (taginfo). This service is constantly updated in near real-time and stores the tags from every object in the global OSM database.

When creating features on OpenStreetMap via the openstreetmap.org website, the following process was noted. After selecting an area of interest, the user can select to create a point, line or area. Building footprints are created using “area”. After digitizing the outline of the building, the user must add attributes. First the feature type must be selected from a list. House is used for houses. The user is invited to add a name, levels, height and address. The street name and city are available from a drop-down list. Additional fields can be added such as material and roof color. All of these attributes are indicated as tags. Only one tag is required for a building, thus a user can leave many fields empty if they do not know the information. Once uploaded the user receives a thank you message and a changeset number. It is worth noting that if building is selected as type, instead of house, commercial etc., then the attribute will only show building = yes. During the creation process, no indication of quality is present. The user is free to create errors when digitizing the building or adding attributes. A base map air photo is shown to guide the digitization process. It was noted in this example, that it was provided by Bing and that a recent apartment building was not shown on the air photo. The lack of oversight when creating features contributes to the lack of positional accuracy and lack of attribute accuracy.

2.2: Spatial data quality

This section explores spatial data quality by examining ISO spatial data quality standards, methods for assessing spatial data quality for authoritative GI as well as the many ways VGI can be assessed including contributor motivations and trust assessment.

- 2.21 ISO definition of Spatial data quality

The International Organization for Standardization (ISO) establishes principles for explaining the quality of geographic data. ISO has established a variety of data quality elements (see Figure 4) to assess spatial data quality. In their overview of data quality elements, the following are measured (see section 7.3 of ISO/TC 211 N 3521, 2013):

Completeness: The measure of the presence or absence of features and their attributes. Measures include omissions and commissions.

Logical Consistency: "The degree of adherence to logical rules of data structure, attribution and relationships" (p.9). Measures include: (from p.9 of ISO/TC 211 N 3521, 2013)

- conceptual consistency – adherence to rules of the conceptual schema;
- domain consistency – adherence of values to the value domains;
- format consistency – degree to which data is stored in accordance with the physical structure of the dataset;
- topological consistency – correctness of the explicitly encoded topological characteristics of a dataset.

Positional Accuracy: "The accuracy of the position of features within a spatial reference system" (p.10). Measures include: (from p.10 of ISO/TC 211 N 3521, 2013)

- absolute or external accuracy – closeness of reported coordinate values to values accepted as or being true;
- relative or internal accuracy – closeness of the relative positions of features in a dataset to their respective relative positions accepted as or being true;
- gridded data positional accuracy – closeness of gridded data spatial position values to values accepted as or being true

Thematic Accuracy: "The accuracy of quantitative attributes and the correctness of non-quantitative attributes and of the classifications of features and their relationships" (p.10). Measures include: (from p.10 of ISO/TC 211 N 3521, 2013)

- classification correctness – comparison of the classes assigned to features or their attributes to a universe of discourse (e.g. ground truth or reference data);
- non-quantitative attribute correctness – measure of whether a non-quantitative attribute is correct or incorrect;
- quantitative attribute accuracy – closeness of the value of a quantitative attribute to a value accepted as or known to be true.

Temporal Quality: "The quality of the temporal attributes and temporal relationships of features (p.10). Measures include: (from p.10 of ISO/TC 211 N 3521, 2013)

- accuracy of a time measurement – closeness of reported time measurements to values accepted as or known to be true;
- temporal consistency – correctness of the order of events;
- temporal validity – validity of data with respect to time.

Usability: This is based on the user's requirements. Usability can be evaluated by using all quality elements. In some cases, it may not be evaluated using the standard measures listed above. Usability should be used to describe the dataset's suitability for a particular application based on user requirements and knowledge, similar to the description of fitness for use.

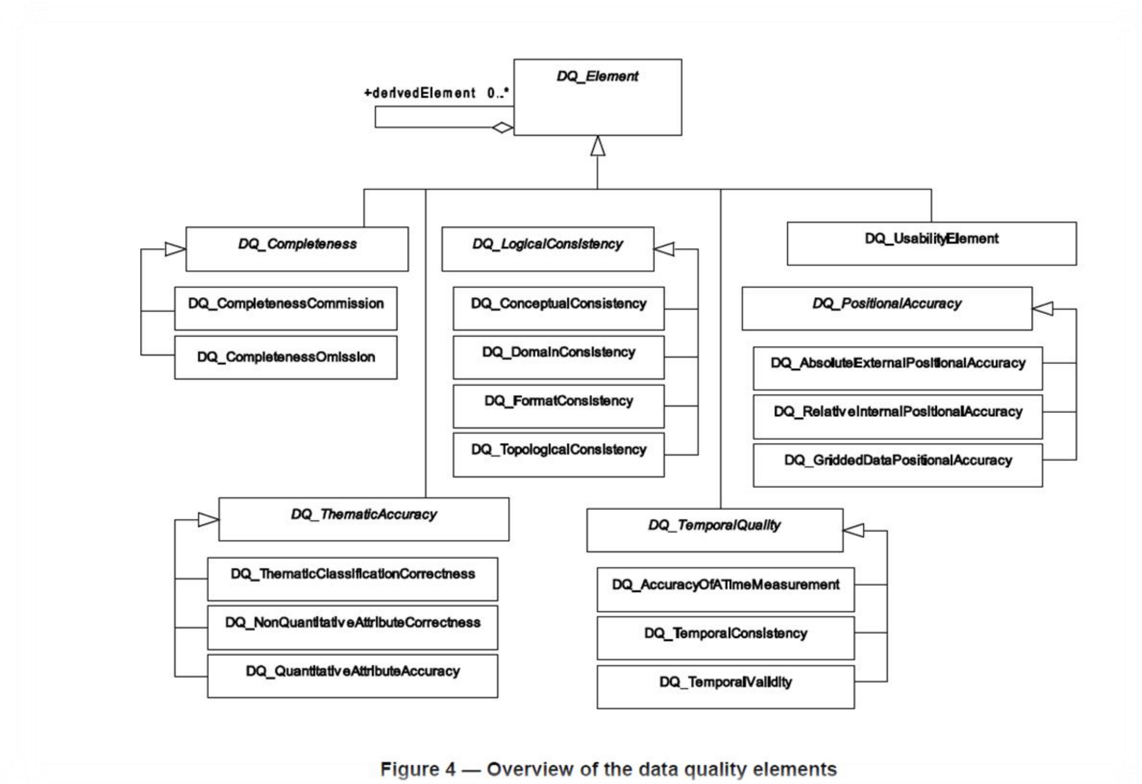


Figure 2.1: Source ISO/TC 211 N 3521, 2013 p.9

- 2.22 Canadian Data Quality Standards for authoritative GI

There are a variety of spatial data standards that authoritative GI must conform to when it is created. These standards are used to promote quality of data produced by experts and ensure consistency across datasets of the same type. In Canada, different standards apply depending on the type of data. Each of the standards listed below apply to only a certain type of data. This section will explore different spatial data quality standards in Canada that are applicable to the data used in this study. For example, parcel data must conform to the National Standards for the Survey of Canada Lands and all municipal data used is part of the Canadian Geospatial Data Infrastructure (CGDI).

The Canadian Geospatial Data Infrastructure (CGDI) is the Canadian national spatial data infrastructure (SDI) that was established in 1999. The CGDI does not host data, rather it is a set of standards, operational policies, technology, and framework data that serve as an infrastructure that promotes interoperable access to geospatial data and services across Canada. In 2012, a CGDI assessment framework was developed based on international SDI assessment models and known assessment methods. The 2012 framework includes 47 assessment methods including standards, policies, technology, framework data, collaboration and leadership. According to the Standards and Specifications performance outcome “A complete and performing CGDI means that there are common technical and data standards in place that allow diverse data sources, services, applications and systems to operate with each other within Canada and internationally” (KPMG, 2016, p. 5). The 2015 assessment found that there were some gaps in the monitoring of performance of each CGDI indicator in a timely manner. Overall, the 2015 assessment found that the CGDI meets most of the assessment methods including data sharing and integration and the coordination of data collection and quality control (KPMG, 2016).

The CanTopo Map Standards and Specifications 1:50,000 (2014) identifies the standards for the creation of maps at a 1:50,000 scale. It includes “a definition of each map feature, instructions on how each feature is compiled and a description of the cartographic symbols used to represent each map feature” (p. 4). Proper feature names in English and French are identified as well as the category they belong to. For each category, the definition as well as the appropriate colours, instructions and font to use are identified. Examples of the cartographic symbols to use as well as any historical notes are presented. A list of written is instruction is also included. Under the instructions for buildings, it is noted that buildings under 60 m in length are referred to as point buildings, excluding sheds and garages under 10 m x 10 m. Buildings 60 m and over in length are referred to as buildings to scale and are shown in their correct position and orientation where space permits. For point buildings there is a minimum line separation in 0.2 mm (pp.257-258).

The National Standards for the Survey of Canada Lands (2014) “provides professional Canada Lands Surveyors with the technical standards that apply to surveys undertaken on Canada Lands. The National Standards have been compiled to provide a common approach to boundary definition for all property rights systems (land registration and natural resource management regimes)” (p. iv). The standards note that all surveys must be geo-referenced to North American Datum 1983 (Canadian Reference System). At least two geo-referenced control points are required to achieve the accuracy standards. In terms of accuracy, the absolute accuracy based on the control points must be +/- 0.20 m or better at a 95% confidence interval. The minimum relative accuracy must be +/- 0.02 m plus 80 parts per million at a 95% confidence interval.

The American Society for Photogrammetry and Remote Sensing (ASPRS) provides positional accuracy standards for geospatial data. The root-mean-square error (RMSE) is “the square root of the average of the set of squared differences between data set coordinate values and coordinate values from an independent source of higher accuracy for identical points” (p. A5). The RMSE is used to calculate various measures of positional accuracy including horizontal and vertical accuracy. Survey checkpoints and ground control points are used. There are requirements for the placement of horizontal and vertical checkpoints such as being in easily visible points or on flat or uniformly-sloped terrain to minimize errors. Non-Vertical Vegetated Accuracy (NVA) and Vertical Vegetated Accuracy (VVA) are calculated at the 95% confidence interval using the RMSE and must meet the accuracy standards, represented as *X* (ASPRS, 2014). The Natural Resources Canada Federal Airborne LiDAR Data Acquisition Guideline (2017) notes that positional accuracy, must at a minimum meet the 2014 ASPRS guidelines. The NVA accuracy at 95% confidence interval is 19.6 cm (1.96*RMSE) and the VVA is 29.4 cm. All levels of government in Canada must conform to this guideline in the creation of LiDAR data. These standards are important to note for this study, as LiDAR data is often used to create municipal building footprints. It can also be present as a base map on OSM. It is good to note that horizontal accuracy (NVA), means that the position of municipal building footprints should not be more than 19.6 cm away from their actual location.

- 2.23 Methods for assessing spatial data quality for authoritative GI

Spatial data quality has been studied for over 30 years. Devillers et al. (2010) note that error models have been used for many years in different fields to measure small errors. Several methods have been proposed to measure spatial data quality including GIS measurements of the positional accuracy for points, lines and polygons. Additional methods have looked at measuring temporal uncertainties as well as the quality of remote sensing images. In the past decade, the importance of semantic quality has come to light. The sharing of spatial data from different sources with different categorizations of their attributes and different specifications for how the data are created highlights the issue of semantics. As semantics are different between datasets, the need for metadata describing the fitness for use of data is vital; however, it is often missing or incomplete (Devillers et al., 2010).

Devillers et al., (2010) note that spatial data quality research has had failures in that there is a weak connection between academic research and the daily users of spatial data. Users often assume that the data is perfect and software vendors do not communicate spatial data quality findings into their software. Increased awareness of spatial data quality is needed among the geospatial user community. Another issue is the terminology used to define spatial data quality. Different researchers, organizations and users have different interpretations of what quality terminology means. There needs to be an increased awareness of spatial data quality among data users. Users are becoming more aware of spatial data quality issue through everyday activities such as noticing the inaccuracy of addresses in Google Earth/Maps. Research methods relating to assessing non-spatial data could contribute to the research on spatial data quality (Devillers et al., 2010).

Devillers et al. (2007) and Gallagher et al., (2015) write about investigating the fitness for use of spatial data. As mentioned in section 1.1, FFU is concerned with users of geospatial data being able to understand how the dataset fits the intended use. It concerns the match between the data's characteristics and the user's requirements for a given task. Users of geospatial data must understand how the datasets fit the intended use. Material that describes data quality is often tough to understand, thus data quality is neglected by users and data can be misused. The fact that data is gathered at different times, by different organizations and with different standards enables the use of data for non-intended purposes when heterogeneous layers are overlaid. As such, Devillers et al. (2007) present a tool for managing heterogeneous data quality and assessing FFU of a given dataset. Gallagher et al. (2015) recommend the establishment of open data quality indicators that flag errors in the data.

Grira et al. (2009) explore the uncertainty in measuring spatial data quality. Spatial data quality management seems to be concerned with measuring errors rather than meeting the needs of the various users. While FFU is accepted as a means of assessing data quality, quality information is still addressed to a single usage and assumes that the user understands the measure of quality. Users are disregarded from the system design process and their needs are assumed. The authors discuss the idea of "perceived quality", stating that VGI users do not have the same understanding of data that experts do (Grira et al., 2009).

The ISO TC 211 N 3521, Annex E, Evaluating and Reporting Data Quality, provides some examples to evaluate and report data quality. These examples show how different standards and quality assessment methods are used evaluate data representing different types of real-world features, such as trees and roads. For example, feature types are noted along with attribute name, value type and value domain. For trees, those with a height of less than 1m are not to be recorded. Also, a maximum of 10% can be missing or in excess and a maximum of 20% can have the incorrect height. No tree features can be misclassified. For roads and houses, "condition" may have no value along with "name" and "number of occupants". Only those feature types and attributes defined in the data product specification can be present in the dataset. For transportation network and buildings, a maximum of 2 features can be missing, in excess or misclassified. The quality evaluation process for all feature types listed consists of specifying data quality units, specifying data quality measures, such as excess or missing item, specifying data quality evaluation procedures and determining the output.

- 2.24 Assessing VGI quality

There are a variety of ways to assess VGI quality. This thesis focuses on measuring the quality of building footprints from OSM, which is one of the most popular forms of VGI. The following section introduces ways to assess the quality of VGI datasets. These methods can also be applied to OSM. Specific methods for assessing OSM building footprint quality are introduced in the next section. This section focuses on evaluating VGI quality by understanding contributor trust as well as the need for metadata.

As VGI can be created by anyone, it is difficult to know the ability of a contributor or the quality of their contributions. As such, Haklay et al. (2010) note that no assumptions can be made about the quality or background knowledge of data volunteers. As it can be difficult to determine the quality of contributors, data may contain errors.

Goodchild and Li (2012) suggest that VGI can often be more accurate than government sourced data. The authors propose three methods for assessing VGI; the *crowd-sourcing, social and geographic* approaches. The crowd-sourcing approach assumes that a group of editors can validate data and correct any errors that an individual can make; thus, more volunteers are better. One volunteer could correct an error made by another, however, there can sometimes be some disagreement as to what the correct type of feature is. Also, obscure features may not attract enough volunteers to ensure accuracy. The social approach relies on select groups of trusted contributors to act as moderators of the data. Individuals who make many accurate contributions would be given a higher level of trust. In OSM, the Data Working Group (DWG) comprised of eight members deals with any dispute relating to features. The geographic approach compares the purported geographic fact with the wide body of local knowledge. It is governed by the syntax of what can and cannot occur at a given location. It relies on spatial dependence, in that a purported fact about a location must be consistent with its geographic context and other facts that are known about the same location.

It is suggested that when more volunteers contribute to a given dataset, they will notice and fix any inaccuracies. Haklay (2010) notes that it is possible to count the number of contributors per square kilometre and use that as an inference of quality and fitness for use. It is thought that positional accuracy can be improved by increasing collaboration across multiple contributors. The authors compared the number of contributors per area to its positional accuracy. They suggest that the relationship is not linear and that above 15 contributors per square kilometre, the positional accuracy is very good below 6m. They also note that the first 5 contributors provide the largest improvement to positional accuracy. The heterogeneity of datasets means that they should be evaluated at a local, not global scale. It is also worth noting that the amount of data contributed by each person leads to a difference in data quality.

Foody et al. (2015) note that VGI can be acquired rapidly and freely, thus it has enormous potential to aid in mapping activities. It is worth noting, that there is often no way to distinguish between different volunteers and thus assign a level of trust to their contributions. The limitations of trust caused by volunteer uncertainty can limit the practical value of VGI. For VGI to realize its full potential, a rating system for the accuracy of a contributor's edits must be available. This rating could prove useful to provide feedback to help users improve their skills and understanding. When many edits are made, a common method is to follow the majority view. This can cause problems when one person correctly labels a case, while many others incorrectly label it. Seeking more volunteers may reduce the quality of a dataset as the contributions may be of poor quality or inaccurate. This can dilute the potential of useful data added a single, skilled member. The authors state the importance of three questions. First, can volunteer labelling accuracy be characterized from the data they provide? Second, does the number of cases contributed by a volunteer relate to the quality of contributions? Third, can information about volunteer quality be used to better mapping applications? Their study found that while the number of volunteers is important for quality, it depends on other factors as well. They found that the number of cases contributed by a volunteer does not relate to the quality of the contributions (Foody et al., 2015).

Kalantari et al. (2014) write about metadata issues with VGI. Metadata is “a set of data which describes and gives information about other data” (p.37). As an example, in OSM metadata can refer to attributes which are assigned to a building, such as name and address. While VGI availability is increasing, VGI often does not contain metadata, making it difficult for users to discover data. The lack of metadata means that VGI is incomplete and may be inaccurate. The authors’ example of looking at four entries for a hospital in Melbourne determined that three of them had the incorrect address or locational position. On a local scale, this lack of metadata can cause major accuracy issues and as such, the authors suggest that creators should be able to describe their VGI.

- 2.25 OSM Quality Standards

While users are not mandated or even instructed to follow quality procedures when creating OSM data (as was noted in section 2.12), the OSM wiki (wiki.openstreetmap.org) has numerous guidelines for ensuring quality that interested users can read. To start, the OSM wiki notes that there is no assurance of quality, but that if one person inputs an error, the other 99.9% of users can fix it. This is an example of Linus’ Law, which states that when there are more contributors, the quality of data will increase, as noted by Haklay et al. (2010). Haklay et al. (2010) also note that while more contributors might notice and fix bugs, no assumptions can be made about the knowledge of an individual contributor. It states that users must judge for themselves the quality of the data. The Data Working Group (DWG) is a group of invited members who deal with issues of vandalism, serious disputes and copyright issues. They can temporarily block users and redact information. Users can contact the DWG if they note acts of vandalism, major border changes or copyright and the offending user does not fix the issue. That being said, the DWG does not enforce quality measures on everyday OSM edits (OSM wiki).

The OSM wiki (https://wiki.openstreetmap.org/wiki/Main_Page) notes that OSM base map aerial imagery comes from a variety of sources that they have permission to use. Bing and Mapquest are the most common sources. The resolution and quality vary and it is not updated regularly. In certain areas, local and national governments have donated high quality imagery or framework vector data such as Natural Resources Canada's CanVec data and Statistics Canada's Building Canada 2020 building footprint project (https://wiki.openstreetmap.org/wiki/WikiProject_Canada/Building_Canada_2020). The OSM wiki lists various quality assurance tools that can be used to lead to better OSM data quality. These include bug reporting tools, error detection tools, visualization tools and monitoring tools, assistant tools and tag statistics. It is worth noting that it is up to the individual user to manually use these tools to evaluate quality. OSM wiki also notes that there are varying degrees of accuracy. It notes errors with GPS accuracy as well as in aerial imagery such as straight roads appearing curved. It provides suggestions to ensure accurate topology such as having many closely-spaced nodes to represent a winding road. In terms of completeness, OSM wiki notes that comparison with external data is often used. Also, for internal comparison within OSM, cross-locational comparison or feature density can be used. Cross-locational comparison may be accurate in Europe, but less so in North America, where similar sized cities are not likely to have similar levels of completeness (OSM wiki). OSM wiki actually lists academic articles on OSM completeness (including Barrington Leigh and Millard Ball, 2017) for those users who want to measure it. It also notes that the map is never complete and that completeness varies.

OSM wiki describes taxonomies for tagging specific types of features. By navigating through the features page (https://wiki.openstreetmap.org/wiki/Category:Features), the user can select the type of feature (subcategory), listed alphabetically to see the tagging guidelines. For example, under the Health subcategory, amenity tag pages are provided for different health providers which describe how to tag a doctor's office, pharmacy or numerous other types of health service providers. Other types of subcategories, such as transport or sport, will have different tagging guidelines and these guidelines can vary by location. For buildings, generic tagging instructions, such as building=house are noted on the buildings page (https://wiki.openstreetmap.org/wiki/Buildings).

The OSM wiki provides a number of good practices and editing standards and conventions which users *can* follow. Among these are don't map historical or temporary events, mapping roads appropriately (# of ways) if they are straight or curved and don't use abbreviations for naming. The OSM wiki How We Map page (https://wiki.openstreetmap.org/wiki/How_We_Map) notes that OSM contributions should be truthful, legal, verifiable and relevant. It notes that OSM values community cohesion over data perfection. Overall, while there are numerous *suggestions* on how to accurately map in OSM, none of them are formally enforced before a user submits data edits. Also, users would have to invest their own time in searching for and learning these guidelines. This may help to explain why attribute accuracy and positional accuracy are lacking in many areas. While experienced users may pay more attention to these, many users are probably not aware of the concept of "quality" when they are creating features. As noted in section 2.23, quality control measures in a VGI/OSM environment are often based on the number and quality of contributors. Additionally, comparisons to reference datasets or the use of specific tools are required to assess quality. This is different from quality control for authoritative data, which must conform to specific guidelines, as noted in section 2.22.

While OSM does not mandate quality standards for data creation, certain organizations have informal standards that they use to ensure that the contributions meet the required FFU. The HOT Tasking Manager provides instructions based on the needs of the project. One example is for Cyclone Kenneth, the instructions note how to create square and circular buildings. It also notes to create the full outline of buildings as accurately as possible, even if part of it is covered by trees. They note that while many buildings are close together, users should not let them touch. It also notes that buildings should be tagged as "building", unless the user actually knows what type of building it is. This particular project notes that many roads are already mapped, but that any roads created should follow the East Africa Tagging Guidelines. These guidelines, which appear on OSM Wiki, note how roads should be tagged, as well as public transport. It asks users to consider the importance of roads and the types of transport used in this area, which may be different from other areas in the world. There also exist tagging guidelines for other areas such as the UK and China, based on the unique features that exist in those areas.

Missing Maps promotes the mapping of natural disasters and links to projects found on the HOT Tasking Manager. It also allows people to host a mapathon for a specific event. Missing Maps allows users to become validators. Validations allow experienced users to analyze others' edits and helps improve the quality of data. HOT tasks consist of squares that users are instructed to complete. Once completed, these squares are then validated by validators to ensure that all mapped areas are reviewed by experienced mappers and any errors or omissions are fixed. This is an example of both the crowdsourcing and social approaches as noted by Goodchild and Li (2012). The crowdsourcing approach is used as many contributors map and review the same area. The social approach is used as validators are assigned a higher level of trust.

- 2.26 Assessing Quality of OpenStreetMap Building Footprints

This section introduces specific methods that can be used to evaluate the quality of building footprints from OSM. While the methods in the previous section can also be used to evaluate OSM datasets, they were primarily qualitative as they focused on the characteristics of contributors. In contrast, the methods in this section are primarily quantitative. To begin, it is important to note that OSM datasets can often be compared to reference datasets with an aim to measure data quality. When comparing datasets, it is often assumed that the reference dataset is of higher quality (Haklay, 2010).

Fan et al. (2014) evaluated OpenStreetMap building footprints in Germany. Evaluations were done to measure completeness, semantic accuracy, position accuracy and shape accuracy by comparing to reference data. There has been an increase in the use of OSM in recent years. 3D buildings have been used in many studies, but these rely on quality building footprints to begin with. The authors suggest have larger buildings will have a higher percentage of overlap with the reference data while smaller and higher building will have less overlap. When measuring positional accuracy, they noted an average offset of 4.13 metres, with a range between less than 1 cm and 15m. Overall, they note that OSM building footprints were nearly identical to those in the reference dataset and that offset is caused by the distortion and limited resolution of the base map in OSM.

Fan et al. (2014) note four elements for measuring building footprint accuracy. Completeness refers to the number or area of building footprints compared to the reference dataset. Semantic accuracy measures if buildings in OSM are indeed present in the real world. It also measures the geometric relations between buildings in both OSM and the reference dataset. Tests have been developed in this thesis to measure these relations. Also developed, was a test to measure commissions, which are buildings in OSM that are not present in the reference dataset. Fan et al's. (2014) definition of semantic accuracy differs from other's (including Devillers et al's., 2010) definition of semantic quality. Semantic quality is a measure of ensuring that consistent semantics (language) are used across various datasets. It focuses on creating standardization in the way features are created and described.

Fan et al. (2014) also mention position accuracy which measures how well the coordinates of the OSM building footprint relate to its actual position on the ground. Furthermore, shape accuracy measures the shape similarity between OSM building footprints and those in the real world or reference dataset.

Hecht et al. (2013) mention three ways to evaluate completeness of building footprints in OSM. Unit-based measurements include comparing the number of buildings and the area of buildings between OSM and the reference dataset. The authors note that an object-based method, known as the centroid method should also be used to increase the accuracy of the evaluation. This method measures whether the centre of the reference building falls within the OSM building. This is designed to measure whether the OSM buildings have proper overlap with the reference dataset. I have implemented all three completeness tests in my methods.

Many authors talk about using generalization to improve the quality of building footprints. These methods are designed to eliminate complex angles in buildings which may not exist in real life. Lokhat and Touya (2016) propose a squaring method that squares angles to the nearest 45 degrees. They performed this using GeOxygène.

Pászto et al. (2013,2015) mention ways to measure and improve and quality through generalization tools. They mention that the simplify buildings tool in ArcGIS can be used to reduce building complexity. Also, programs such as FRAGSTATS 4.1 and Shape Metrics Toolbox for ArcGIS can be used to perform shape metric calculations. That being said, the Shape Metrics Toolbox, originally designed for ArcGIS 9, has malfunctioned in testing when using the updated ArcGIS 10 version. There exist a number of shape metrics that can be evaluated. It is worth noting that this concept relates to landscape ecology and many tools are designed to measure landscape “patches”, usually rasters and thus are not useful in the research context of this thesis.

In terms of measuring positional accuracy, one useful method proposed by authors is the Hausdorff Distance. Filippovska and Kadab (2010) mention how many studies have used Hausdorff Distance to measure positional accuracy. The Hausdorff Distance is essentially the maximum distance between two surfaces. Min et al. (2007) propose a method to further modify Hausdorff Distance, known as the Extended Hausdorff Distance, which demonstrates the versatility and usefulness of this measure.

2.3: VGI Contributors and their Motivations

- 2.31 VGI Contributors

There are a variety of people who produce and contribute to VGI. It is worth noting that there are different types of contributors and contributions. There are also a variety of motivations for contributing to VGI. Editing OSM features is a popular type of VGI contribution, yet the number of edits and level of editors varies greatly from one location to another.

Coleman et al. (2009) note that there are important questions to ask when determining how an organisation should use VGI, how they should assess the credibility of produsers and how they can attract new produsers. A produser is someone who both contributes to and uses VGI data. The authors suggest that there are 5 types of VGI contributors: Neophyte, Interested Amateur, Expert Amateur, Expert Professional and Expert Authority. These range from people with interest in the subject, but no knowledge (Neophyte) to those who are starting to research the subject (Interested Amateur) to those with expert knowledge (Expert Amateur) to those who rely on their knowledge for a living (Expert Professional, Expert Authority). These classifications, however, are not sufficient to describe the wide range of contributors. The authors identify 8 motivations for contributing to VGI, some of which apply to OSM including: altruism, professional or personal interest, intellectual stimulation, social reward, enhanced personal reputation and most importantly for OSM, pride of place. All of these reasons may be considered motivations for individuals to participate in mapping parties. Furthermore, there are some instances where contributors have negative motivations for contributing such as intentional mischief or to promote their agenda. Despite the success of OSM and VGI, there remain skeptics. The authors note three lessons about VGI contributions. First, that contributions may not be actual data, but rather just an update to attributes. Second, that contributors want recognition and third, that contributors want to see their contribution used. They also note that while many volunteers are willing to contribute information about their local area, few are willing to make contributions over a longer time period or for other areas. It is thus important to create ways in which to ensure contributor will remain interested in contributing in the future (Coleman et al., 2009).

Budhathoki and Haythornwaite (2013) note that large numbers of contributors are needed to create and update OSM data, but that repeat contributions are also needed. Those who receive recognition and feel like part of a community will be more likely to contribute again. Contributor communities can provide motivation for new and continued contributions. The authors note that there are both “heavyweight” and “lightweight” contributors. Heavyweight contributors are part of a community, they are knowledgeable and their contributions are judged. Lightweight contributions are simple, random contributions. The authors state many motivations for OSM contribution which include learning, reputation, fun, altruism and community participation. In a study, they found that the majority of contributors were young males with a full-time job and college education, living in Europe. They determined that casual mappers believe two things: that it is important to provide free digital maps, and that such map data should be available for free only for non-commercial applications (Budhathoki and Haythornwaite, 2013).

Mooney and Corcoran (2012) conducted a study on heavily edited objects in OpenStreetMap. These are objects that have been edited 15 or more times. They note that of all the heavily edited objects in the UK and Ireland, 87% of the edits were made by only 11% of the 4128 contributors. Additional nodes were added to the objects in 79% of the edits. Only 0.4% of all objects in the study area were considered to be heavily edited. It was found that in 64% of tag edits, tag values were reverted back to a previous value and in 32% of tag edits, a new or updated tag value was assigned. In 4.1% of the objects, the name attribute was changed 3 or more times. The authors note that future studies should consider the lineage and history of OSM data as part of the quality assessment. Larger numbers of heavily edited objects should be studied to see if there is a correlation with the author’s findings (Mooney and Corcoran, 2012). This study highlights the characteristics of more popular objects in OSM. One can note that not all contributors give an equal contribution to edits and that to have more accurate objects it is likely that you will need to have highly active contributors.

- 2.32 Collaboration and Recognition

Investigating ways to improve OSM contributions will lead to increased completeness and accuracy of OSM data. In order to promote greater levels of contribution, it is important to understand the motivations behind users' contributions. In order to attract and maintain contributors to OSM, it is important to understand the collaborative nature of OSM. Chen (2017) notes the desire for collaborating and discussing among users is a motivation in itself. VGI contributors who view their contribution as part of a collective effort or common goal are more likely to contribute (Kuznetsov, 2006).

Due to the collaborative nature of OSM, it is important that contributors feel recognised. In interviews with contributors, Chen (2017) found that receiving thank you or positive feedback encouraged them to contribute more. It is worth noting that there are limited functions in OSM that allow users to give positive feedback on others work. Hamari and Koivisto (2015) suggest that adding a liking feature would allow users to give positive feedback on others' contributions which would help satisfy the social recognition motivation and foster a sense of community. Additional features to allow better social interaction can help encourage more contributions.

- 2.33 Fixing Errors through Self-Policing and Self-Efficacy

One way that OSM controls accuracy is through self-policing behaviour. Chen (2017) suggests that the nature of competition leads contributors to pay attention to others work, which exposes errors to more people allowing them to be corrected. Another aspect of self-policing is a sense of ownership, which is a motivator in itself. While contributors create data for others to use, they feel a sense of ownership for their creations. They will pay attention when others make changes to their work. This sense of ownership also extends beyond an individual's contributions to the greater local area with which they are knowledgeable. This sense of ownership allows contributors to be better engaged with data that they created as well as data in their local area, which helps to improve accuracy (Chen, 2017).

Self-efficacy was identified by Budhathoki and Haythornwaite (2013) as a motivation for OSM contributions. They suggest that mappers are interested in the idea of improving maps in their local area or areas they are knowledgeable about. It was suggested that showing incomplete or inaccurate areas to those with local knowledge was a great motivation for both amateur and expert contributors who have a desire to see their contributions being used. Seeing errors is suggested to help contributors overcome their inhibitions and become confident that their contributions can help fix the errors. Providing positive feedback to novice contributors further encourages them to contribute their local knowledge. A desire to improve OSM in all areas is also a motivator for contributors. Such data improvement goals, in general, can be used to motivate contributors by showing them errors or incomplete areas (Budhathoki and Haythornwaite, 2013).

- 2.34 Additional Motivations and HOT Contributors

Although OSM does not offer financial incentives to its contributors, Chen (2017) noted that some companies such as Mapbox use OSM to provide geospatial services to clients and hires people to work on OSM datasets, thus monetary rewards such as salary or career advancement can influence some contributors.

Coleman et al. (2009) note that altruism is an important motivator for some contributors who create content for the benefit of others without the need for personal gain. Altruism has become increasingly popular in response to humanitarian disasters. Dittus (2016) noted many new active contributions after natural disasters, in coordination with projects organised by Humanitarian OSM Team (HOT). These types of contributions can help organisations and governments respond and provide quality mapping to areas where it did not previously exist.

Dittus (2016) notes that 80% of Humanitarian OpenStreetMap (HOT) contributors have no previous OSM experience and 90% have less than 5 days of experience. In studying the activity of contributors, 50% worked for at least 65 minutes total, 20% work for at least 3 hours total and 5% work for 18 hours or more. Among HOT contributors, 30% contribute to a second project and 5% contribute to 6 or more projects. Dittus notes that highly publicised disaster events can be key to recruiting more contributors and that contributor numbers go up after these events. He also notes that newcomer retention can be a problem, with most contributors stopping after a few days.

Dittus (2016) also mentions newcomer experiences with HOT OSM contributions. He suggests that it is important to keep tasks simple and not to overwhelm contributors with complex task requirements. While HOT mapathons have been successful at attracting contributors, Dittus suggests that these events are simply attractors for those who were already prepared to contribute. An important motivator is letting people know that their work will have an impact. He states that it is important to share what has been done and what the maps have been used for. Dittus notes that few contributors were "reactivations" or people who were dormant for 60+ days and then recontributed.

- 2.35 Encouraging Contributions

The studies presented in this section demonstrate the various reasons why contributions occur, ways that errors are resolved, contributor motivations and the characteristics of HOT contributors. By understanding these topics, one can make suggestions for ways to encourage more contributions from more contributors. First off, as collaboration is a motivation, there needs to be ways for users to interact with each other. Having projects listed on social media platforms and allowing users to discuss their contributions with others can help establish a sense of collaboration. Another way to encourage repeat contributions is by recognizing a contributor's efforts. Adding a positive feedback or like tool on OSM would make contributors feel valued are more likely to make further contributions. As self-policing and self-efficacy are important factors in OSM contributions, these should be used to encourage more contributions. As a sense of ownership applies to a contributor's area, making them aware of edits or other contributions in the area will encourage more scrutiny, thus improving accuracy. Another aspect is a desire to improve one's local area. Having a platform where users could show errors or a lack of data in their area would help encourage other local users to contribute to that area to improve it.

Altruism is an important motivation for contributors. Having natural disaster mapping projects publicized in multiple areas could help encourage contributions. While the HOT Tasking Manager has a list of projects, having these projects listed on other mapping social media platforms can help attract more contributors. Based on the study by Dittus (2016), many HOT contributors stop contributing after a short time. To attract further contributions from them, it is important to keep tasks simple. Also, the importance of their contributions and the uses of the mapping should be publicized on the HOT website and on other mapping social media pages. This will help encourage repeat contributions, but also attract new contributors as they can read about how their contributions will help others. Overall, based on all of the studies reviewed, the number one thing to do is to inform people of the numerous projects they can contribute to and to do so in as many ways as possible.

2.4: VGI/OSM for Planning

This section demonstrates the importance of VGI and OSM for planners by exploring the need for geospatial data at the municipal level, the value of VGI for planning the importance of ensuring the quality of VGI for planners' and citizens' use.

- 2.41 The need for geospatial data at the municipal level

There is a need for geospatial data at the municipal level in order to accomplish a variety of tasks. Planners rely on geospatial data in order to have a digital inventory of buildings, roads and infrastructure in their municipality which are used to perform planning analysis, guide policies and create maps for planning projects and community consultations.

Planners would be interested in a variety of VGI to help them with planning analysis and public participation in planning. One example of such is Atzmanstorfer et al's. (2014) study of a system allowing citizens to report local issues in Quito, Ecuador. VGI layers such as roads, buildings, parks, trails and paths can help build a digital inventory of a municipality's infrastructure especially when municipal open data has not been created. Barrington-Leigh and Millard-Ball (2017) noted that this is especially valuable in developing countries. Public participation can be increased through VGI by allowing citizens to share their thoughts, report issues on apps and map features or issues in their municipality. Examples of this are noted in section 2.11 such as the 311 app in Edmonton and the citizen mapping project in Saint John. Planners can use this VGI to keep track of new infrastructure, perform analysis and improve public services.

Barrington-Leigh and Millard-Ball (2017) note that OpenStreetMap (OSM), which started in 2004, has grown rapidly over the years. OSM started with a focus on streets but has since grown to include buildings, land uses, points of interest and other geographic features. Their study on the completeness of road networks found that most Western countries have a relatively complete road network in OSM. Planners and citizens now have a large variety of data to use from OSM.

Drummond and French (2008) note that the use of spatial analysis and manipulation by GIS closely aligns with the needs of planners. The increase of available spatial information and open source GIS has allowed a variety of public interest groups to perform GIS analysis to provide relevant information to planners about their needs. GIS is becoming more mainstream and is no longer limited to expert users. As such, GIS analysis is now essential for planners to convey information to the public (Drummond and French, 2008). While GIS is now essential for planners, in some communities, there is a lack of government spatial data and GIS software. The availability of OSM and open source GIS software thus presents an opportunity for planners in small communities lacking resources to access and analyse spatial information in their community.

- 2.42 Value of VGI for planning

VGI and OSM can be used for a variety of planning purposes. This is especially useful when a municipality has not created its own open data. The vast amount of data available means that multiple types of analysis can now be performed. In smaller communities and low-income countries, there is a lack of resources to create this data and thus VGI presents an amazing opportunity to use data at the local level. Furthermore, as VGI is always being updated, it allows for more up to date data.

Goodchild and Li (2012) note that individuals without expertise in cartography can now create maps of their local area using their local knowledge. In some cases, volunteers can create more accurate maps of their local areas than a distant government agency. OSM has become one of the most successful alternatives to government sourced data. The authors suggest that VGI can often be more accurate than government sourced data. Haklay (2010) notes that crowdsourcing data can result in a large cost reduction for the enterprise that profits from it. Government data is often slow to be updated and suffers from budget constraints. OSM allows users to contribute to data that may not otherwise be available or up to date. Coleman et al. (2009) note that governments can use VGI data in high use areas to keep local attributes and data up to date.

Barrington Leigh and Millard Ball (2017) note that OSM can be used for research in economics, urban planning and environmental studies as well as to analyze transportation networks. OSM can provide many benefits, especially in low-income countries where government data may be unavailable. It is worth noting that completeness may be lacking in some countries, which can make analysis difficult or inaccurate. OSM completeness is greatest in low and high densities. Urban areas with lots of contributors are likely to be complete as well as interurban roads that traverse rural areas. Smaller towns and villages are most likely to have incomplete data. Smaller countries are generally more complete as well as those with high levels of internet access. GDP was found not to have a significant impact on OSM completeness. OSM road data can be used for transportation behaviour modeling and local climate emission modeling as well as many other types of research in areas where authoritative data is not available. Road length per capita is a good indication of economic development and transportation patterns and monitoring OSM road network changes can help indicate these patterns. OSM contribution requires internet access and a good education. That being said, income does not necessarily have an effect on OSM completeness. Many low-income countries have complete data and also many places have had intense mapping efforts following humanitarian disasters, such as Haiti. Overall, China is the least complete area, due to government restrictions (Barrington-Leigh and Millard-Ball, 2017).

Boeing (2017) presents a new tool, called OSMnx that allows users to collect and analyze OSM street network data in a simple way. OSM data is often more complete than the official government supplied data from which it is created. OSM has a more complete street network that includes defined paths, trails, bike lanes and alleys that may not be identifiable in government sourced layers. While tools currently exist to extract data from OSM, there are compromises such as size limitations, oversimplification and replicability. The new OSMnx tool is a free Python package that allows users to download political/administrative boundaries, building footprints and road networks from OSM. Users can specify a large variety of place geometries and can combine nearby areas together in a seamless download package that can be saved in a variety of formats. The tool can perform many types of network analysis functions including querying street/path type and shortest path calculations based on road type, speed limit etc. It can also calculate many types of road network statistics. OSMnx is more accurate as it properly calculates nodes and non-planar roads, thus accounting for elevation changes such as tunnels and bridges. Downloaded files can be used in a variety of GIS programs. The simple automation of the collection and analysis of road network and other data from OSM using this tool can save researchers a lot of time and allows for easier analysis. As road network data is important for planning, an easy to use tool like this, combined with the vast data in OSM, could help planners easily gather and analyse data to aid in policy decisions such as road improvements, new bike lanes etc. In the future tools like these may have a great impact on planning analysis and will allow greater use and understanding of the vast data network of OSM (Boeing, 2017).

In terms of public participation, VGI represents an amazing opportunity to better engage with local citizens. Cowan (2013) states that VGI allows members of the public to participate directly in the use, creation and sharing of spatial information that is relevant to personal or community concerns. VGI can also be used as an alternative to ineffective traditional public participation methods. Atzmanstorfer et al. (2014) note that geospatial web platforms, social media and VGI have unlocked a new era for public participation GIS (PPGIS). In Quito, Ecuador, they developed and tested a social Geoweb platform, *GeoCitizen* that merges Geoweb technologies and social media into a tool that allows citizens to collectively report observations, discuss ideas, solve and monitor problems in their local area.

Nikšič et al. (2017) note that VGI can be used to monitor citizen preference. Such uses can include allowing users to identify preferred areas of interest in a city. Another example looked at tracking bicycle riders to determine which paths were most used and where new paths should be created. This type of citizen preference information can prove valuable to inform planners of the usage rate of community services and infrastructure and where new investments are needed. This type of data also provides real-time information to help with determining future trends, as opposed to using static datasets. A variety of studies have been performed in various cities including Ljubljana, Slovenia and Odense, Denmark to monitor citizens' urban transportation patterns thus determining where congestion occurs and in which areas improvements should be made (Nikšič et al., 2017).

Fonte et al. (2017) note that the thematic richness of OSM can have great potential for developing land use and land cover (LULC) maps. There are some issues in the conversion of OSM data into standardised LULC classes. The authors developed an automated method to convert OSM data into LULC classes which can be used for a variety of planning tasks requiring land use maps. Their method resulted in the product being delivered as an ESRI shapefile, for the effective use of GIS land cover analysis by planners. Their application will be released under an open source licence to allowing anyone to create an LULC map from OSM data and monitor change over time.

Kelly (2007) notes that spatial data had become invaluable in the fields of urban and rural development, planning and management. A spatial data infrastructure (SDI), which is a collection of spatial data available for all levels of government, non-profit and academia, has been developed by numerous local and regional governments. There have been challenges with the integration of data and the need for data standards. Numerous users now require up to date spatial data to aid in decision making such as emergency and natural resource management. Spatial information can help solve problems in cities. For example, in Lagos, Nigeria, rapid urbanization has led to decreases in the quality of the environment and difficulty managing land in the region (Osei et al., 2006). Local government has not kept up with tracking the change or making effective policies for land administration. In rapidly growing cities, especially those without adequate government data, up to date spatial information is needed to have the appropriate information for planners to make decisions about land management and expansion policies (Osei et al., 2006). In such cases, OSM can provide useful spatial data for planners that is updated frequently and is often more accurate and complete than government data.

- 2.43 Quality requirements of VGI for planners' and citizens' use

In order for VGI and OSM to be useful for planners and local citizens, a certain level of quality must be obtained. Many studies have examined the completeness of OSM as well as identifying features that may be missing. There may be a variety of reasons for varying accuracy and solutions to improve accuracy in those areas. In many areas, OSM data is said to be complete and accurate and thus it presents a useful data source.

Brinkhoff (2016) used OSM features to denote built-up areas. He notes that the "landcover" key in OSM has few values. Many countries have poor "landuse" coverage features. Buildings are far more numerous in OSM; however, many are coded simply as "yes". As buildings are features that require a tag, "yes" if the most often used tag to describe a building when the creator does not know the building type. He notes that some OSM built-up features have low accuracy. He states that building footprint quality in OSM is sufficient. Built-up feature coverage varies dramatically by region. Germany is very well digitized in terms of OSM built up features while Mexico City is not. While determining built-up areas from OSM data is feasible on a global scale, there still exist variations in coverage, features and attributes that can make quality analysis difficult.

Haklay (2010) evaluated the quality of OSM in London by comparing the positional accuracy of motorways. A buffer-zone method is a common way to compare positional accuracy of objects. Completeness was also measured by comparing the total length of road datasets for an area. It was found that the total length of the OSM dataset was 69% of that of the reference Meridian dataset. Omissions of data in a certain area can also indicate completeness. Derived and remote areas may be missing features when compared to a reference dataset. The study found that OSM data was accurate to about 6m and that major roads had up to 100% coverage. Having a diligent group of participants in an area can lead to a very complete and accurate dataset, while other areas may not be as accurate or complete. A research question from this analysis is at what point is the information become good enough for cartographic output and GIS analysis? Positional accuracy can vary greatly from over 70% down to 20% and the errors are not randomly distributed, indicating overlooked areas.

Authors such as Goodchild and Li (2012) and Barrington Leigh and Millard Ball (2017) have noted that VGI and OSM can often be as accurate as or even more accurate than government sourced data. This is especially true in low-income countries and rural areas where there is a lack of government resources to produce data. Barrington Leigh and Millard Ball's (2017) worldwide completeness study found that OSM completeness was at around 83% as of January 2016; however, their completeness measurements do not account for attribute or positional accuracy. Of the 185 countries, 77 are more than 95% complete. The findings demonstrate that researchers, planners and policymakers in most areas can rely on the completeness of OSM for their studies. In many places, OSM is now the most complete and accurate data available, even for local governments and thus its completeness is essential for it to be relied upon for planning and development.

What all of the studies demonstrate is that the accuracy of OSM and VGI varies around the world. That being said, many areas have a high level of completeness and accuracy and thus the data is suitable for analysis. In his own studies, the author of this thesis has noted that many areas in Canada lack completeness and accuracy. Also, attributes are also vague or non-existent for many OSM features. These finding will be discussed in more detail in this paper, however, this demonstrates a need for continuous improvement of OSM data in Canada in order for the data to be useful. Ways to improve upon this data will also be examined.

2.5: Summary

The literature review demonstrates the unique importance of understanding VGI, spatial data quality and OSM to be able to extract the enormous potential for planners and citizens. VGI data availability is rapidly expanding and so are its potential uses. This brings with it problems in ensuring the quality of this data for users. Therefore, one must understand spatial data quality and apply different methods of evaluation to ensure quality. In terms of OSM, the data provides a variety of uses for planning. The literature, however, is often focused on other uses for OSM as well as individual quality measurements and thus presents an opportunity to study the usefulness of OSM for planning and the creation of a simple, multi-dimensional quality evaluation of OSM datasets.

The literature identifies a set of needs which will be addressed in this thesis. There are a lack of studies focusing on measuring the quality of building footprints, particularly from OSM. The study performed in this thesis adds to the literature by measuring the quality of building footprints from OSM. There is also a need for OSM quality evaluation in different areas of the world. Most studies have focused on evaluating OSM quality in Germany or the UK, likely as these countries have a high level of completeness. This study evaluates OSM quality for Canadian cities, which is an important addition to the body of literature. Additionally, the methods presented by other authors have been difficult to comprehend for the average person or planner. The methods presented in this thesis are designed to be understood and replicated by planners and everyday GIS users.

Chapter 3: Methods

3.1: Methods Overview

This thesis investigates the evaluation of the quality of building footprints from OpenStreetMap (OSM). It aims to combine existing methods and new ones into an easy to understand and easy to implement solution. The problem with the current evaluation of building footprints is the diversity of methods, lack of aggregation of methods and the complexity of methods. Furthermore, existing studies on OSM quality often focus on roads or focus on ways to improve quality rather than measure it. In terms of diversity, different studies usually focus on an existing specific method or introduce a new one. For example, Hecht et al. (2013) measured the completeness of building footprints from OSM using three different methods: area of buildings, number of buildings and the centroid method. Mooney and Corcoran (2012) measured the number of edits for OSM objects. Fillippovska and Kabad (2010) measured positional accuracy using Hausdorff Distance. What these studies lack is an aggregation of methods to measure multiple aspects of quality. Proposed methods are often complex to understand, involving formulas and tables that are not relatable to municipal staff, citizens, students or others who may benefit from using building footprint data. For example, Fan et al. (2014) use a turning function to measure tangent angles, which many people would likely not understand. Min et al. (2007) explain their methods using numerous formulas involving non-negativity and triangle inequality. In terms of studies done on OSM quality, many have focused primarily on measuring road quality like those from Boeing (2017) and Barrington-Leigh and Millard-Ball (2017). Numerous studies focus on ways to alter or improve OSM data rather than measuring its quality. For example, Lokhat and Touya (2016) propose a squaring method for buildings and Pászto et al. (2013,2015) suggest the use of generalization tools to improve the quality of data.

There has been much focus on the evaluation of spatial data quality in general. In relation to OpenStreetMap (OSM), many studies including those by Coleman et al. (2009), Budhathoki and Haythornwaite (2013) and Goodchild and Li (2012) have focused on contributor motivations and the contributor to quality relationship. As noted above, in terms of quality evaluation of specific features, most studies tend to focus on the quality of road datasets from OSM. Studying the research has found a limited number of studies focusing on measuring the quality of building footprints, however, these tend to focus on a specific metric or rather a way to improve the quality of building footprints, usually through generalization. The study in this thesis aims to develop a simple to use set of models for measuring building footprint quality which incorporate some of the methods and ideas from the literature as well as combining them with new methods to achieve a comprehensive evaluation of building footprint quality.

This study is unique in that it combines many different quality measures together. Furthermore, in the literature, it was noted that many of the tests had not been performed on buildings footprints as overall there are limited studies on building footprints. This may be due to the fact that roads were initially seen as being more important data. This study will focus on Canadian cities, while most existing studies on OSM quality have focused on other countries, notably Germany (such as Fan et al., 2014) or the UK (such as Mooney and Corcoran, 2012). There are limited studies on OSM quality in Canada, especially related to building footprints. It is believed that this study adds a valid contribution to the literature by performing new tests on building footprints in Canada and by developing an easy to use model which can be replicated and used in other areas by researchers and planners. Furthermore, there will be an investigation into the reasons behind varying levels of quality and the usefulness of OSM for planners; topics which have rarely been discussed in the existing literature. These topics will be discussed in the discussion chapter.

The purpose of this study is to evaluate the quality of building footprints from OpenStreetMap (OSM) for a sample of Canadian Municipalities. This is done by comparing them to a reference dataset of building footprints produced by the municipality. Models are developed using Model Builder in ArcGIS to perform a variety of tests on the datasets. The results can thus be aggregated for the entire municipality or tests can be done only for certain measures or only for certain areas as desired. The evaluation of the results aims to compare quality across municipalities and investigate trends and ways to improve quality of building footprints in OSM. Further tests are also performed on OSM datasets without comparing them to a reference dataset in order to determine quality when no reference dataset is available or being used.

This study uses quantitative methods for the evaluation. Municipalities are evaluated by a standardised set of tests which produce numerical results in the new layers' attribute tables. This type of method allows for easy comparison between different areas as results will be-numerical and can be compared. The discussion focuses on reasons behind the differences in quality including an examination of qualitative factors such as contributor motivation and knowledge of local areas. Furthermore, the importance and usefulness of OSM data and ensuring its quality for planners will be discussed.

In addition to the model, there will be a report on the number of building footprints in each municipality over a set time period (2010-2018). There will be a report on when contributions occurred and in which areas of the city they occurred in. Determining when contributions occurred is important as it can help one to understand the reasons behind the contributions. Are there external events that contributed to contributions during a set time period? The reasoning behind contributions as well as ways to improve contributions are highlighted in the discussion chapter.

3.2: Data Collection and Study Areas

Datasets for this study were gathered from OSM and municipal datasets. Municipalities studied must have building footprint, zoning/land use and parcel shapefiles available either through an open data portal or from the University of Waterloo Geospatial Centre. This is in order to allow all of the quality tests to be performed. It is worth noting that many of the tests can be performed without all of these datasets, however, for this study, only areas with all of these datasets were chosen to get a greater picture of quality and to ensure consistency among results. These datasets were downloaded from the municipality's (local or regional) open data portal or these datasets were obtained from the University of Waterloo. Metadata pertaining to the municipal open data used can be found in Appendix 4. Building footprints and roads were downloaded for each study municipality from OpenStreetMap using the HOT OSM Export tool. The area for the downloaded building footprints is slightly larger than municipal boundaries to ensure all data is received. The OSM dataset is later clipped to the same boundary as the municipal dataset, usually by using the zoning layer.

In the study performed in this thesis, geospatial municipal data was used that must conform to a variety of standards as noted in section 2.22. Looking at the metadata for each of the datasets used provides some insight (see Appendix 4). For example, parcel data is provided by external providers and gathered from sources such as the Ontario Land Registry Office or the BC Land Title System. The creators of this parcel data must meet the National Standards for the Survey of Canada Lands. Building footprints were derived from orthoimagery, LIDAR data and aerial photographs, which must meet the Canadian and ASPRS guidelines. Zoning layers are based on the city's zoning by-law and usually conform to parcel boundaries. The city of Brantford notes that its datasets conform to the ISO 19139 metadata standard. The data used in this study forms part of the CGDI and thus must conform to their common standards and policies.

For the evaluation of completeness over time, datasets were captured for each municipality for each six-month period from January 2010 to July 2018. These files were gathered from James McCarthy at the Mapping, Analysis and Design Lab at the University of Waterloo. The Osmosis tool for OSM was used which allows for the extraction of data from OSM planet history files in order to give a dataset from a given period in time. It also notes the contributor's name for each feature. The analysis of temporal quality is important because it shows when the building footprints were added in each municipality. Additionally, one can observe what types of buildings were added at a given time and in what area they were added. Having attributes about the contributors allows one to see how much they have contributed and how many contributors there are for a given area and time period. The explanation behind the measurement of temporal quality in this thesis is presented at the end of this chapter.

This study will focus on 10 Canadian municipalities with populations between 30,000 and 100,000 people as of the Statistics Canada 2016 census. Study areas will align with municipal (city) boundaries, defined by Statistics Canada as Census Subdivisions (CSD). This means that for cities that are part of a larger urban area, only the individual city will be studied. The cities selected are found in four provinces: Ontario, British Columbia, Alberta and New Brunswick. This allows for comparison of quality across various regions of Canada for similar sized cities. There will be an investigation of potential reasons for regional disparities. Cities in other provinces were not chosen mainly due to lack of data; however, some of the tests (but not all) could still be applied to OSM data from other cities.

The reason for choosing cities in the 30,000 to 100,000 population range is because these represent medium-sized urban areas in Canada. Some of the selected cities are geographically isolated, while others are part of or near a larger urban area. It is worth noting if geographic location relates to quality. Mid-sized cities were chosen as they represent the greatest opportunity for the use of OSM data. In Canada, there exist many mid-sized cities, thus there are many people who could benefit from this data. According to the Statistics Canada 2016 Census, there were 95 CSDs with populations between 30,000 and 100,000 and only 54 CSDs with populations over 100,000. These 95 CSDs represent a total population of 5 273 968 people. Appendix 1 shows the list of mid-sized CSDs in Canada and their population and land area (study cities are highlighted). Larger cities generally have better municipal data available and are usually more complete on OSM due to a larger number of local contributors and interest. It is also more difficult to report on results for larger cities and they may have neighbourhood differences. That is why they were excluded, and mid-sized cities were chosen, where the quality information can more easily be aggregated and understood. The use of this information can be important as many mid-sized cities do not yet have municipal data. This was noted when searching for municipal data in mid-sized cities, many provinces (and their municipalities) are behind when it comes to the provision of municipal open data (as shapefiles), including Manitoba, Prince Edward Island and Nova Scotia. Quebec has a general lack of information available in English. All Quebec open data is accessed through Données Québec and parcel information is not available; zoning is also not available for most cities. Cities were chosen from the four provinces of Ontario, British Columbia, Alberta and New Brunswick as these provinces are ahead when it comes to the provision of open data.

List of Cities:

1. Halton Hills, Ontario
2. Niagara Falls, Ontario
3. Kamloops, British Columbia
4. Prince George, British Columbia
5. Chilliwack, British Columbia
6. Saint John, New Brunswick
7. Lethbridge, Alberta
8. Grande Prairie, Alberta
9. Stratford, Ontario
10. Brantford, Ontario

Halton Hills is located on the edge of the Greater Toronto Area (GTA) in Ontario and had a 2016 population of 61161 and a land area of 276.27 km^2. It is composed of two primary communities, Georgetown and Acton, with the rest being primarily rural area.

Niagara Falls is located in the densely populated Greater Golden Horseshoe (GGH) in Ontario and had a 2016 population of 88071 and a land area of 209.73 km^2. It is located on the American border and is a world-famous tourist destination. The population is fairly clustered with rural area in the south.

Kamloops is located in central British Columbia (BC) and had a 2016 population of 90280 and a land area of 299.25 km^2. Like other BC communities, the population is clustered in the central area, on both sides of the Thompson River and is surrounded by rural area.

Prince George is located in North Central BC and had a 2016 population of 74003 and a land area of 318.26 km^2. The population is clustered together and surrounded by rural area.

Chilliwack is located in the Fraser Valley in Southern BC and had a 2016 population of 83788 and a land area of 261.65 km^2. The population is spread out on both sides of the Trans-Canada Highway and also includes rural areas.

Saint John is located in Southern New Brunswick on the Bay of Fundy and had a 2016 population of 67575 and a land area of 315.96 km^2. The population is quite spread out and includes some rural areas.

Lethbridge is located in Southern Alberta and had a 2016 population of 92729 and a land area of 122.09 km^2. It is primarily urban with population on both sides of the Old Man River.

Grande Prairie is isolated in North-Western Alberta and had a 2016 population of 63166 and a land area of 132.73 km^2. Like many cities in Alberta, it is primarily urban.

Stratford is located in South-Western Ontario approximately 35 km west of Waterloo and had a 2016 population of 31465 and a land area of 28.28 km^2. The population is entirely urban and fairly dense. Stratford attracts many tourists each year to its famous Stratford Festival featuring numerous plays.

Brantford is located in Southern Ontario approximately 25 km west of Hamilton and had a 2016 population of 97496 and a land area of 72.44 km^2. The population is entirely urban and is fairly dense.

Table 3.1: Study Municipalities

City ↓ Metric →	Population (2016)	Area (km^2)
Halton Hills	61161	276.27
Niagara Falls	88071	209.73
Kamloops	90280	299.25
Prince George	74003	318.26
Chilliwack	83788	261.65
Saint John	67575	315.96
Lethbridge	92729	122.09
Grande Prairie	63166	132.73
Stratford	31465	28.28
Brantford	97496	72.44

Figure 3.1: Map of Study Municipalities

3.3: ISO Quality Metrics for Building Footprints

Based on the ISO/TC 211 Standard 19157 (2013), the measures of data accuracy which can be applied to building footprints were identified. The definitions of these measures were interpreted from the ISO/TC 211 Standard 19157 document.

Positional Accuracy (section D.4) has a variety of measures:

- Mean value of positional uncertainties: This is the difference between the measured position and the true position for a given number of points, such as building points or building centroids.
- Mean value of positional uncertainties excluding outliers: This measure excludes outliers (values above a defined threshold) from the calculation.
- Number or rate of positional uncertainties above a given threshold: Identifies only those points that are over a given threshold (certain distance) from the true value.

Thematic Accuracy (section D.6) has a variety of measures:

- Number of incorrectly classified features: Identifies features (buildings) with incorrect attribute class. Ex: residential building classified as commercial.
- Misclassification rate: The number of incorrectly classified features based on total number of features.
- Number/rate of incorrect attribute values (non-quantitative attribute correctness): The count of all features with incorrect attribute names. Rate identifies number of incorrect values over total number of attributes.
- Quantitative attribute accuracy: The measure of correct quantitative attributes. Ex: correct street number.

The ISO Standard 19157 also identify data quality elements which can apply to building footprints (section I.4):

- Commission: A feature that is present in a dataset that is not present in reality.
- Omission: A feature that is present in reality but missing from the dataset.
- Topological Consistency (geometric relationships): Correctness of the adherence to topology rules. Ex: Buildings must not overlap.

- Positional Accuracy: The accuracy of features in relation to their position on Earth.
- Thematic Accuracy: The correctness of quantitative and non-quantitative attributes and the classification of features and their relationships.

The models in this thesis have been developed to test a variety of measures based on the ISO/TC 211 Standard 19157, although not all the above measures are tested. The next sections will explain the metrics used in the models and the individual tests they perform.

3.4: Metrics for Model with Reference Dataset

Commissions

Commissions are an indicator of semantic accuracy and are thus important to measure. Fan et al. (2014) note that semantic accuracy indicates if buildings in OSM are indeed present in the real world. The model isolates commissions in the OSM feature class. These commissions are important to identify as they can represent errors in the OSM dataset or new buildings that have not yet been added to the reference dataset. ISO section 7.3 includes commissions as a measure of completeness as it represents the presence of (excess) features.

Metrics of Completeness

Completeness: 3 tests

The three completeness tests are designed to get a comprehensive understanding of completeness. The model uses both unit-based (area and number of features) and object-based (centroid method) to get a thorough assessment of completeness. The unit-based methods compare the total area of buildings between the OSM and reference dataset, as well as the total number of features in both datasets. The centroid method determines if the centroid of the reference dataset's building is within the outline of the OSM building. These measures are based on those indicated by Hecht (2013) who states that while unit-based measures are important, by including the object-based centroid method, one can increase the accuracy of the evaluation.

Zonal Completeness: 2 tests

Zonal Completeness is important to measure to get a proper representation of what types of buildings are being represented in the OSM dataset. The model measures zonal completeness by number of features in each zone as well as total area of features in each zone to see which zoning categories are most complete when it comes to the OSM dataset for a particular city. The number of features and total area of features in each zone are compared between the OSM and reference datasets. Zonal completeness is important, especially for planners, as it allows one to see the relationship between land use and completeness. This study hypothesizes that commercial and institutional zones will have a greater level of completeness than residential zones. Many authors such as Coleman (2009) and Budhathoki and Haythornwaite (2013) have suggested that contributors will often focus on areas that they are knowledgeable about and that many contributors are motivated to improve local knowledge. This is why commercial and institutional zones that attract many visitors may have more contributions than residential zones that are only frequented by neighbourhood residents. Understanding which zones have greater completeness can allow planners to focus on improving contributions in less complete zones.

In order to measure zonal completeness, a spatial join is required between the building footprint dataset and the zoning dataset. This is done for both the municipal and OSM building footprint datasets. This gives a count of the number of building footprints that are in each zoning category and the area of all building footprints in each zone. This is done using the “within” clause. This means that building footprints that are partially within two or more zones will be counted in each zone that they are within. That is why the number of building footprints in all zones is reported in the tables as being higher than the total number of building footprints in the municipality. The other option would be to use a “completely within” clause, meaning that only building footprints that are completely within a zone would be counted, leading to a lower number of building footprints being reported. This was not used because of the positional inaccuracy of the building footprints, as many are not completely contained with a parcel.

Due to the unique nature of zoning categories across municipalities, the names of the zones are different across municipalities. For some municipalities, there were so many zones, that certain zones of the same type were aggregated into one new zone. For example, if there was “commercial node”, “shopping centre commercial” and “commercial” they would all be aggregated to “commercial”. This means that it is difficult to compare zonal completeness across municipalities as the extent of the zoning categories is different. For some municipalities, the zones were not aggregated, whereas others with many zones were. For the analysis, general comments are made about completeness by zone type between municipalities. For Halton Hills and Grande Prairie, a “not specified” category was removed with an insignificant land area. For Stratford, the “not specified” category includes 3 large parcels that consist of fields and trees. It was not removed due to the total land area.

Metrics of Positional Accuracy

Positional Accuracy: 2 tests

The model measures positional accuracy with two tests. The Near distance between centroid points is an important measurement as it represents the offset or placement inaccuracy of the OSM dataset. It measures the distance between the centroids of the OSM building and the building in the reference dataset. One potential fault of this method is that buildings without a 1:1 relationship can contain multiple centroids in one of the feature classes. This will often increase the Near distance when 2 OSM centroids are both referring to the same reference centroid. A greater mean Near distance can indicate that the OSM dataset contains spatial relation errors. This measure is useful as the Near distance is reported for each OSM feature. That means that one can determine the positional accuracy for a selected building or group of buildings.

Another test measures whether buildings footprints are contained within a parcel. This will measure whether the OSM building footprints are in the correct location as building footprints should not cross parcel boundaries. In some cases, such as row houses or commercial blocks, the building footprint will cross multiple parcels. This is not necessarily an error as block buildings are often divided by ownership and parcel. These buildings will be identified in the output feature class and one can determine if an error exists. The primary purpose of this method is to identify single buildings that are skewed in their geographic position and cross over their parcel boundary. The ISO notes that internal positional accuracy is the “closeness of the relative positions of features in a dataset to their respective relative positions accepted as or being true” (p.10). In this case, the reference dataset is accepted as having the true position. Many studies such as those by Fan et al. (2014), Fillipovska and Kadab (2010) and Haklay (2010) have measured and reported positional accuracy for either buildings or roads. That being said, the methods of measuring positional accuracy done here are different than those used in such studies.

3.5: Relations between Datasets

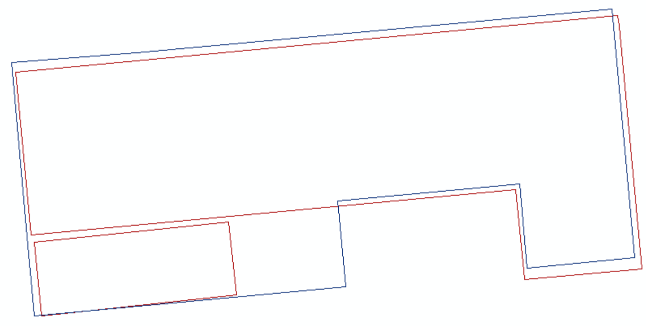
The following types of relations between datasets can be identified by using by the model with the reference dataset. Examples are provided as well as the tools used to identify them or the feature classes that they will show up in.

Relation types (from Fan et al., 2014): Examples from Niagara Falls

1:1 – a building is represented in both datasets

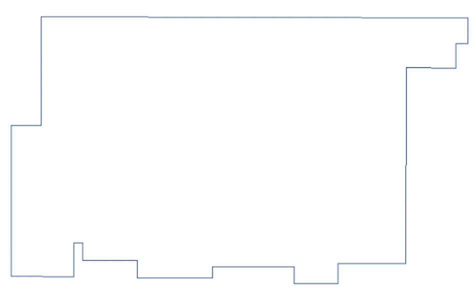
1:1 – select by location (OD building footprints that intersect OSM building footprints, export features) then (OSM building footprints that intersect intersected OD building footprints, export features) The result will give all features with a 1:1 relationship but may also give some with 1:n and n:1.

1:n – an OSM building is represented as multiple buildings in the reference dataset



1:n – these features will show up in the intersected feature classes, however, it is difficult to pinpoint them. One way to do so is to use the near tool (OD building footprint intersected points to OSM building footprint intersected points, geodesic distance) Then look for NEAR FID shared by two or more features, this indicates that 2 OD features are near or represented by only 1 OSM feature.

1:0 – an OSM building does not exist in the reference dataset (Commission)

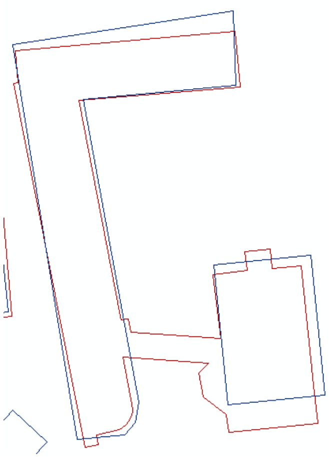


1:0 - Select by location, OSM features that intersect OD features, then change the selection to get commissions

0:1 – A building in the reference dataset does not exist in the OSM dataset

0:1 - Select by location, OD features that intersect OSM features, then change the selection to get omissions

n:1 – a building in the reference dataset is represented by multiple buildings in the OSM dataset



n:1 - these features will show up in the intersected feature classes, however, it is difficult to pinpoint them. One way to do so is to use the near tool (OSM building footprint intersected points to OD building footprint intersected points, geodesic distance) Then look for NEAR FID shared by two or more features, this indicates that 2 OSM features are near or represented by only 1 OD feature.

n:m – a group of multiple buildings are incorrectly recorded in both datasets

Picture not available from datasets

n:m – these are not calculated and are usually non-existent

3.6: Metrics for Model Without Reference Dataset

Metrics of Attribute Accuracy

Attributes of building footprints are important in order to understand what type of building it is and where it is located. The first test will identify if a building has the street name (address) identified. The second test will identify if a building has been given a name. It will also give an indication as to the type of business (for commercial buildings). Buildings with names are usually commercial, institutional or apartment residential. The third test will identify if the building type is identified, rather than simply stated as "yes". Together, these three tests give a comprehensive picture of attribute or rather thematic accuracy. Kalantari et al. (2014) note that attributes are a form of metadata, which is often missing from VGI. Their study showed that a lack of metadata can lead to incorrect positions and descriptions of buildings; thus, it is important to have as many attributes as possible. ISO notes that "the accuracy of quantitative attributes and the correctness of non-quantitative attributes" (p.10) are part of the definition of thematic accuracy.

Metrics of Positional Accuracy

Do any buildings touch roads? Do any roads touch buildings?

The first test identifies if buildings overlap with roads, thus identifying inaccurately created buildings. In some cases, a service road or footway will lead directly to a building. This is why the second test is important. It identifies and isolates any roads that touch buildings. The attribute table will identify the type of road. This way, service roads and footways can be excluded from the list of errors. The purpose of the tests is to note overlap between buildings and actual roads while identifying in new feature classes which buildings and which roads contain errors.

Topology

Do any buildings overlap?

This test will identify overlapping buildings within the OSM dataset. In many cases, the building was created twice in the same place. In other cases, two overlapping buildings will have different shapes and thirdly, some buildings may have a sliver overlapping with a neighbouring building.

These three types can be identified through visual inspection. In all of these cases, some form of building creation error exists, thus they are important to identify and ensure that buildings follow this topology rule. This topology rule is one of the ArcGIS topology rules for polygons and is also under the umbrella of the ISO definition for topological consistency.

3.7: Description of Models

The models were developed to easily measure a consistent set of spatial data quality measures for building footprints. The tools used in the model help achieve new layers and attributes that are used to measure quality, either of the OSM dataset itself or in comparison to the municipal dataset. These new layers created from the model can be visualized to see the results and comparison between datasets. The attribute tables give quantitative results of the test for each building of which the mean is reported for the entire municipality. The individual values for each feature can in some cases indicate a level of quality on a feature by feature basis, allowing for a study of quality for a smaller area. The goal of the models is that they are replicable. The tools are in place and thus the individual datasets for each municipality can thus be “plugged in” to the model for easy comparison across municipalities as well as future use of the model to measure improvements in quality over time.

Before performing the two main models, a quick model was developed and is used to project all datasets to the same UTM projection for the city as well as the clip datasets to the same city boundary. This preparatory model is not necessary if all of the data is already projected and clipped to the appropriate boundary.

The first model compares the accuracy of an OSM building footprint dataset with that of a reference dataset. The tests perform a variety of accuracy calculations including completeness, commissions, positional accuracy and attribute accuracy. There are some limitations to this model, such as the lack of shape accuracy metrics, which are discussed later. The second model works to evaluate OSM building footprint quality without using a reference dataset. Positional accuracy, attribute accuracy and topology rules can be determined for the OSM dataset. It is worth noting that these tests are not as comprehensive as those where a reference dataset is used; however, when a reference dataset is available, these tests should also be used to gather a more comprehensive picture of data quality.

The first preparatory model involves the projecting and clipping of datasets. First, the Project tool is added to the model to project the OSM building footprint and roads datasets to the local UTM zone. If the municipal datasets are not in the local UTM zone, they are also projected. Second, the Clip tool is added and the projected OSM datasets are clipped to the municipal boundary, usually by clipping it to the same extent as the zoning dataset. Third, the Dissolve tool is added to dissolve the zoning dataset by zone. This creates one multipart polygon for each zone, which will allow for easier analysis as the user only has to look at one feature instead of 1000s of individual parcels representing each zone. The outputs from this model are the municipal building footprint, OSM building footprint, parcel, zoning and OSM roads datasets projected into the local UTM zone, covering the same boundaries. Also, a dissolved zoning dataset is produced, which is needed to count the number of buildings in each zone. Having datasets in the same projection is necessary for planning analysis to reduce errors caused by different projections. Making sure the datasets cover the same geographic area is also important to reduce errors related to incorrect counts.

A variety of tools are used in the models. Spatial Join allows for a count of features within other features. Make Feature Layer creates a layer from a dataset, which is needed for use in the model. Select Layer by Location allows features to be selected based on their location (within another feature, intersects, contains another feature etc.). Copy Feature allows creates a new layer based on the previous tool’s output (ex: results from a selection are exported as a new layer). Feature to point converts polygon (buildings) to points that are located in the centroid of the building. The Near tool determines the distance between the centroids of two feature classes. The Frequency tool creates a table listing the number of occurrences of each attribute (2 or more indicates n:1 relation).

The first model performs a number of tasks against the reference datasets to determine various quality measures. First, a Spatial Join is used to get the count of building footprints in each dissolved zoning class. This is done for both the OSM building footprint dataset and the municipal building footprint dataset. Next, Make Feature Layer is used to create feature layers for both the OSM and municipal building footprint datasets as well as the parcels dataset. Feature layers are required for use in the model.

To determine commissions, Select Layer by Location is used to find those OSM buildings that do not intersect municipal ones. It is also used to create layers for municipal building footprints that intersect OSM building footprints and municipal building footprints that have their centroid in OSM building footprints (centroid method of completeness). Next, Select Layer by Location is used to get OSM building footprints that are not completely within a parcel. Next, select layer by location is used on the OSM building footprints that intersect the municipal building footprints (that intersect the OSM building footprints). Copy Feature is used to create new feature classes for all the selections, so that users can view the results. Next, Feature to Point is used on the intersected municipal and OSM building footprints. This is so the Near tool can then be used on the point feature classes to determine the distance between the centroids of both feature classes. This is then exported as a new feature layer. Next, the Frequency tool is used on this point feature layer to determine n:1 relations (which municipal building footprints are represented by 2 or more OSM building footprints).

The outputs from this model include both an OSM building join layer and a municipal building join layer, which give the counts and areas of each building footprint type in each zoning category. These layers are used to determine zonal completeness. Additionally, a layer is created showing only those buildings that are possible commission errors for viewing on the map. A layer showing those buildings that are not contained within a parcel is also produced. A layer showing municipal buildings that have their centroid in an OSM building is created. A layer is created showing municipal building footprints that intersect OSM building footprints. This layer is then used to output a new layer that shows OSM building footprints that intersect this layer (1:1 relation). This is necessary to ensure that commissions are not included in the mean near distance calculation. Two point feature classes are created for both of these (municipal and OSM) intersected buildings. They are then used to create a mean near distance point feature class with mean near distance shown in the attribute table. With this feature class, the user can view the centroid points and the near distances in the attribute table. Finally, a frequency table is created showing the frequency for each objectid. As noted, planners can use these resulting layers to calculate measures such as object-based completeness, positional accuracy, zonal completeness and commissions. While this study aggregates measures on a city-wide basis, having these layers showing exactly which buildings are affected by each test, allows planners to perform analysis on a neighbourhood or individual building level. They would simply have to clip the output layers to their desired study area or select individual buildings of interest.

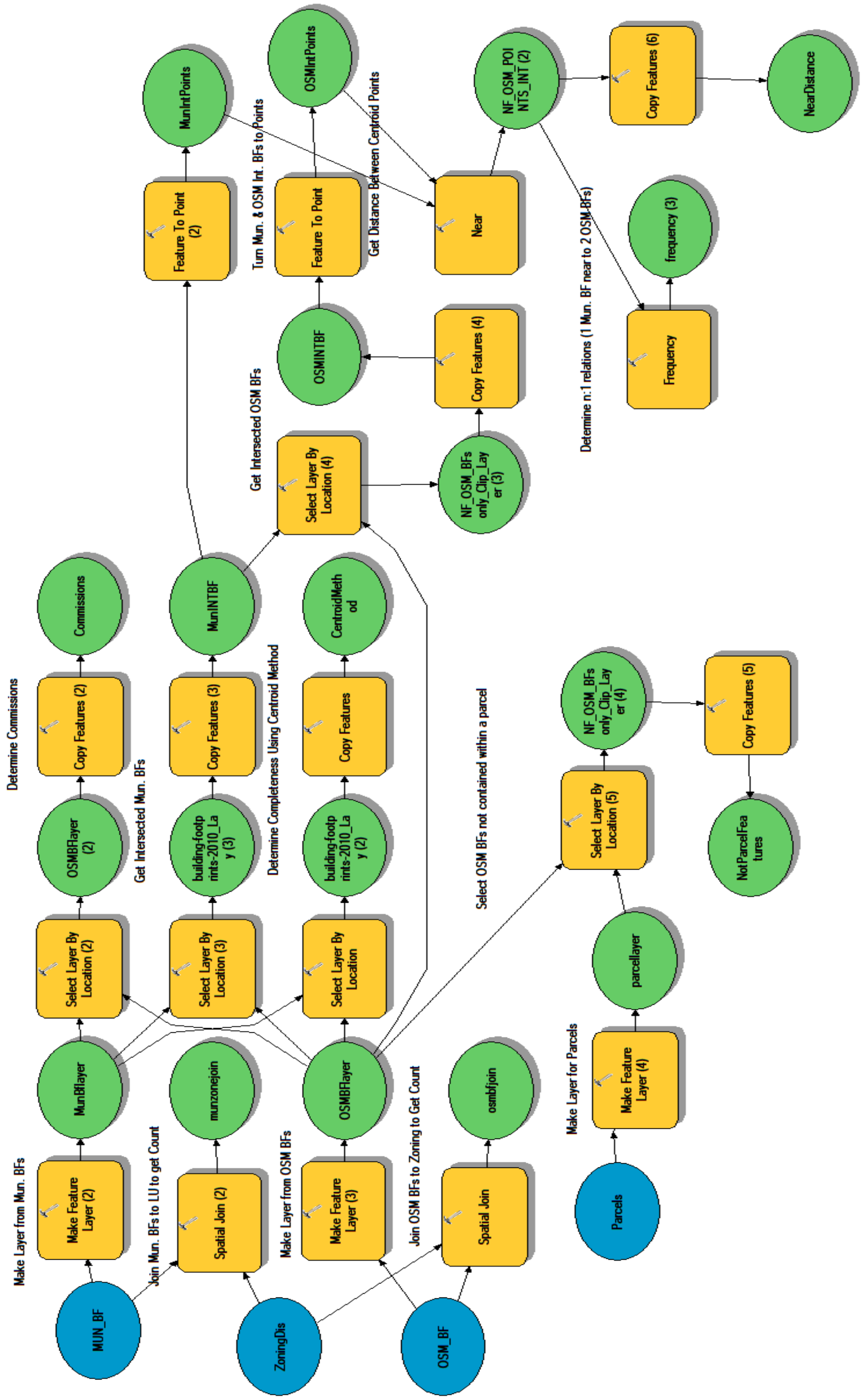


Figure 3.2: Model With Reference Dataset (For Niagara Falls)

The second model performs a variety of tasks to measure quality without the use of reference datasets. First, make feature layer is used on both the roads and building footprint datasets from OSM. Next, the intersect tool is used on the building footprint dataset to determine overlapping buildings. Next, select layer by location is used twice to determine which buildings touch roads and which roads touch buildings. Results are copied to new feature classes. Next, select layer by attribute is used three times on the building footprint dataset to determine buildings with a name, those with a street name identified and those with the building type identified. The identified results are copied to new feature classes.

The outputs from the second model include an OSM layer showing overlapping buildings, a layer showing buildings that touch roads and a layer showing roads that touch buildings. Three additional layers are created showing OSM buildings that have a name, those that have a street name and those that have a type identified. Planners can use these layers to determine which buildings and roads contain potential topological errors. They can also see which buildings have each type of attribute. Again, this allows for quality analysis on a neighbourhood or individual building level.

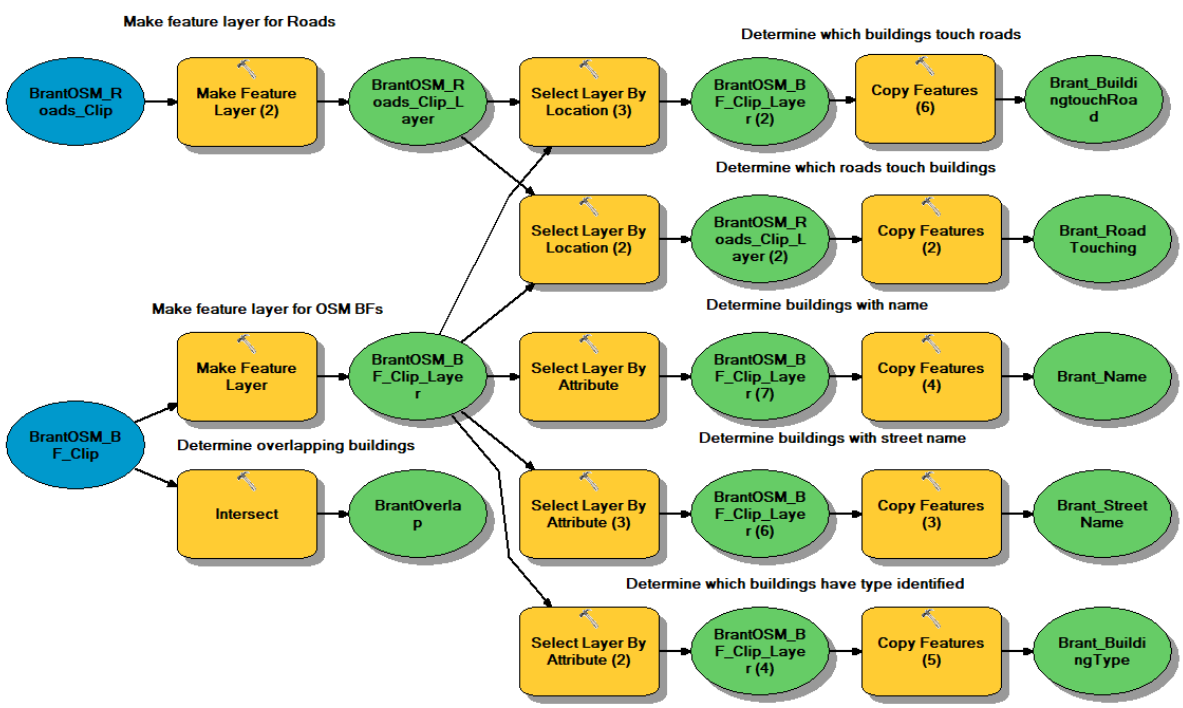


Figure 3.3: Model With No Reference Datasets (For Brantford)

3.8: Monitoring of Completeness over time

This study will report on the number of building footprints in each municipality over a set time period. The number of building footprints in each municipality will be recorded every six months from January 2010 to July 2018. In order to do this, all of the timed datasets are added in a map document for each municipality, one at a time. These datasets are then sorted in chronological order. By opening the attribute table for each dataset, the number of building footprints can be reported at each time interval. Once this is done an analysis of where the building footprints are being created is done. For this, a road feature class and a Census Subdivision (CSD) boundary feature class are used as references. This helps determine which areas are seeing increases in the number of building footprints over time and if they are located along major roads. Google Maps is also used as a reference to determine the type of area in which the building footprints are being added. I will report on the number of contributors for each municipality for July 2018. This is done by dissolving the July 2018 dataset by contributor name. This result is posted in the comparison table and at the end of each municipalities' descriptive paragraph. While it would be useful to report the number of contributors for each time period, this would be ineffective and time consuming. First of all, there were not many contributors during the first few time periods. Secondly, as a link is being made between the number of contributors and the *current* level of quality, only the most recent data for contributors should be used.

Determining when contributions occurred is important as it can help one to understand the reasons behind the contributions. Are there external events that contributed to contributions during a set time period? The reasoning behind contributions as well as ways to improve contributions are highlighted in the discussion chapter. Many authors, including Goodchild and Li (2012) suggest that having more contributors will lead to increased accuracy as more editors will fix any errors in the data. The study will thus see if there is a link between accuracy from the tests (performed in November 2018) and the number of contributors for each city (as of July 2018). It is worth noting that for most of the study municipalities, few building footprints were added between July and November 2018. The only exception is Brantford, where the number of building footprints increased from 1716 to 2591, thus there may be additional contributors.

3.9: Methods Summary

The methods used in this study are comprehensive in nature and allow for a thorough evaluation of building footprint quality for OpenStreetMap building footprint datasets for mid-sized cities in Canada. The models and tests are useful and effective at measuring various data quality standards identified by the ISO and the academic community. The models are replicable for others to use in Arc GIS. They can be used for building footprint datasets in any city or area. In addition to the models, the aspect of temporal quality and completeness over time is presented through an analysis of the city's datasets taken from various time periods. This useful information allows researchers to see in which areas the contributions have occurred. The results gathered through these methods allows for insight into the reasoning behind varying quality by location, the usefulness of building footprints for planners and ways to improve completeness and accuracy of OSM data, all of which are presented in the discussion chapter.

Chapter 4: Results

This chapter will present the results of the study. The first section presents the results from the models for each study municipality. A table indicating the raw numbers for each metric is presented first. The next section compares the results by city. This starts with a comparison table of key metrics across all study municipalities. Next, a written description and comparison of the results is presented. The final section in this chapter investigates the results from the completeness over time study. This section starts with a table indicating the number of building footprints for each six-month period from January 2010 to July 2018 for each city. Also, the table notes the number of OSM contributors for each city in July 2018. This chapter ends with a written description of the trends in the number of building footprints over time for each city. The reasons for varying accuracy across municipalities are discussed in chapter five.

4.1: Results by Municipality

Explanation of Table

Table 4.1 presents the raw numbers for each metric that was measured by the models for each study municipality. Firstly, the total area of all the municipal building footprints, as well as the OSM building footprints, are presented in m^2 (rounded to the nearest m). Next, the total number of municipal and OSM building footprints is stated. This is followed by the number of municipal building footprints that have their centroid in an OSM building footprint (centroid BFs). The total number of commissions is reported next as well as the total number of OSM building footprints that are not completely within a parcel. The total number of N:1 relations based on the "frequency" measure are then reported. This means that one municipal building footprint is represented by more than one OSM building footprints. In other words, multiple OSM centroids are referring to the same municipal centroid. Finally, the total number of OSM building footprints that contain attributes for name, street name and building type are listed.

Table 4.1 – Results by Municipality (Raw Numbers)

City → Metric ↓	Halton Hills	Niagara Falls	Kamloops	Prince George	Chilliwack	Saint John	Lethbridge	Grande Prairie	Stratford	Brantford
Area of Mun. BFs (m^2)	4442648	6915216	7872540	6449148	10059062	5437781	8677734	4788628	2598237	7966261
Area of OSM BFs (m^2)	942373	1797235	6044322	1707919	1247773	4560696	1066206	2790936	2229359	3326482
# of Mun. BFs	20417	39556	38139	27860	43149	36678	58235	30666	12670	36168
# of OSM BFs	757	2120	28066	1415	1037	21055	916	13127	6289	2591
Centroid BFs	786	1794	27942	1348	956	19132	781	12629	5791	2481
Commissions	59	98	309	25	42	661	4	88	541	35
BFs Not in Parcel	199	465	7610	662	240	5421	410	4062	2159	896
N:1 Relations	14	68	119	55	55	490	52	11	47	37
BFs w Name	88	169	738	119	144	447	347	176	133	144
BFs w Street	75	137	563	26	57	5282	173	30	95	897
BFs w Type	265	663	12907	181	322	1564	514	137	1493	1255

4.2: Comparison of Metrics by Municipality

- 4.21 Comparison Across Municipalities

Explanation of Comparison Table

The comparison table (4.2) provided on the following page is a summary of a municipality's overall level of quality across different tests and can be used to compare quality across municipalities. The metrics that are compared include the completeness by area, completeness by the number of building footprints and completeness using the centroid method. The percentage of building footprints that are commissions and that are not contained within a parcel are also presented. Next, the mean near distance between building footprint centroids (OSM and municipal) are presented. The number of roads and building footprints that touch are presented next as well as the number of building footprints that overlap. Finally, attribute accuracy measures are presented including the percentage of OSM building footprints that have a name, street name and building type identified.

Table 4.2 - Comparison of Quality Metrics Across Municipalities

City → Metric ↓	Halton Hills	Niagara Falls	Kamloops	Prince George	Chilliwack	Saint John	Lethbridge	Grande Prairie	Stratford	Brantford
Area Comp. %	21.21	25.99	76.78	26.48	12.4	83.87	12.29	58.28	85.8	41.76
# of BF Comp. %	3.71	5.36	73.59	5.08	2.4	57.4	1.57	42.81	49.64	7.16
Centroid Comp. %	3.85	4.54	73.26	4.84	2.22	52.16	1.34	41.18	45.71	6.86
% Commissions	7.79	4.62	1.1	1.77	4.05	3.14	0.44	0.67	8.6	1.27
% BFs not in Parcel	26.29	21.93	27.11	46.78	23.14	25.75	44.76	30.94	34.33	34.58
Mean Near Dist. (m)	4.45	2.86	2.06	4.6	5.11	2.12	6.79	1.57	2.05	3.18
Road/BF Touch	1	2	2/1	1	0	1	0	0	0	0
# Overlap	4	12	32	2	16	26	0	0	22	2
% w Name	11.62	7.97	2.63	8.41	13.89	2.12	37.88	1.34	2.11	5.56
% w Street	9.91	6.46	2	1.84	5.5	25.09	18.89	0.23	1.51	34.62
% w Type	35	31.27	45.99	12.79	31.05	7.43	56.11	1.04	23.74	48.44

When looking at and comparing the results from the models, one can notice which cities stand out in terms of spatial data quality of their building footprint OSM data. By analyzing these results certain assumptions can also be made. In terms of overall spatial data quality, Kamloops scores the highest having the highest completeness by number of building footprints and centroids of municipal building footprints in OSM building footprints. The city also scores well in the categories of positional accuracy and the building footprint attributes with type of building, but not so well on the other attribute categories. Saint John and Stratford also stand out for their high levels of completeness, particularly by area of building footprint coverage. Grande Prairie also does well overall but has the worst level of attribute accuracy. Brantford presents the most balanced level of spatial data quality with reasonable scores across all tests.

The worst city overall is Lethbridge, which has the lowest completeness and worst positional accuracy. Surprisingly, the city has the best attribute accuracy. Chilliwack also scores poorly overall. The remaining cities of Halton Hills, Niagara Falls and Prince George are mid-pack on the tests but overall have a low level of spatial data quality.

When looking at positional accuracy Lethbridge scores the worst with a high percentage of building footprints not contained within parcels and a high mean Near distance. Prince George also scores poorly in the categories. Niagara Falls has the lowest percentage of building footprints outside of parcels and Grande Prairie has the lowest mean Near Distance suggesting that the position of OSM building footprints compared to their municipal counterparts is more accurate.

In terms of attribute accuracy, Lethbridge scored the best overall, with the highest scores in building footprints with name and building footprints with building type. Brantford also did well and so did Chilliwack, despite having a low overall score. Grande Prairie did the worst in all three tests. There appears to be a link between completeness of building footprints and attribute accuracy. Cities with low levels of completeness (fewer building footprints) have higher attribute accuracy. This is seen with Lethbridge and Chilliwack which both have higher levels of attribute accuracy. Cities with higher completeness (more building footprints) have less attribute accuracy. This is seen with Grande Prairie, Stratford, Saint John and Kamloops which all score poorly in at least two of the three attribute accuracy tests. Perhaps the reasoning is that with more building footprints, it becomes harder to provide attributes for them. This really should not be the case if attributes are added when the building is created. If attributes are added afterward, then it would be more difficult and time consuming to add attributes to a higher number of buildings.

- 4.22 Distribution of Quality within Municipalities

When looking at the variations in quality within municipalities, certain trends are apparent. First of all, city centres have a higher level of quality then outskirt areas. City centres have the highest levels of completeness across the three completeness measures. Also, building footprints in the city centre are more likely to have attributes. Building footprints with a name are usually commercial and are in city centres. The street name and type attributes can also be found for building footprints outside of the city centre, however many building footprints with these attributes are found within the city centre. Figure 4.1 demonstrates building footprints with attributes in Saint John (high overall completeness). These trends generally apply to all of the study municipalities. In municipalities with low completeness, the building footprints that are there are usually those in the city centre only. Residential area building footprints are usually missing from municipalities with low completeness including Chilliwack, Lethbridge, Halton

Hills, Niagara Falls and Prince George. Figures 4.7 to 4.25 (odd numbered) presented later in the chapter show the distribution of building footprints in each municipality as of July 2018.

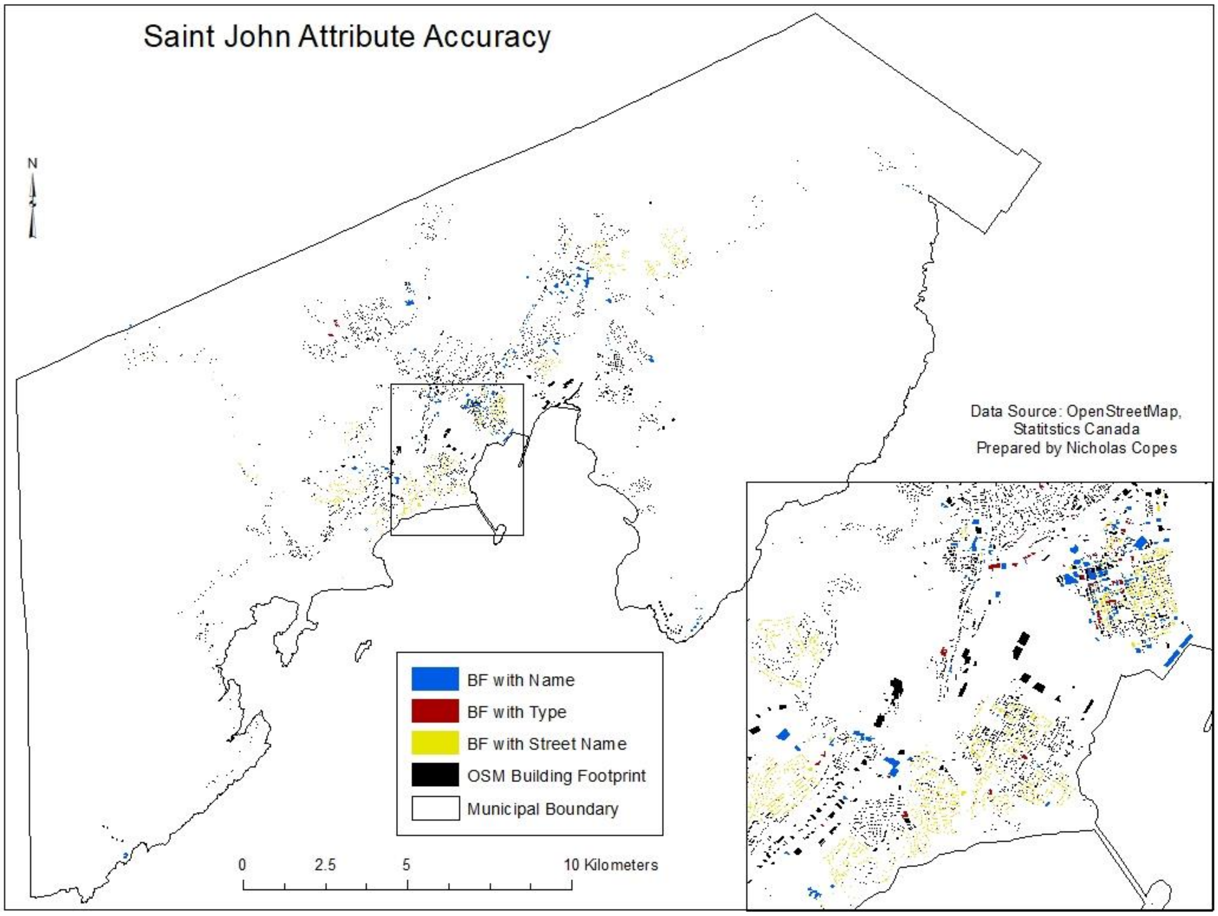


Figure 4.1 – Attribute Accuracy in Saint John

In terms of other quality measures, patterns are not as apparent. For example, positional accuracy measures such as buildings not within parcels and Mean Near Distance do not show local variations. Buildings not located within a parcel are found across all areas of each municipality and the Mean Near Distance is a result of this poor positional accuracy, as municipal building footprints were digitized to fit within the municipal parcels. Figure 4.2 shows buildings not within parcels dispersed across the municipality of Grande Prairie. One can notice that inaccurate building footprints are located across the municipality in both the core and outlying residential areas.

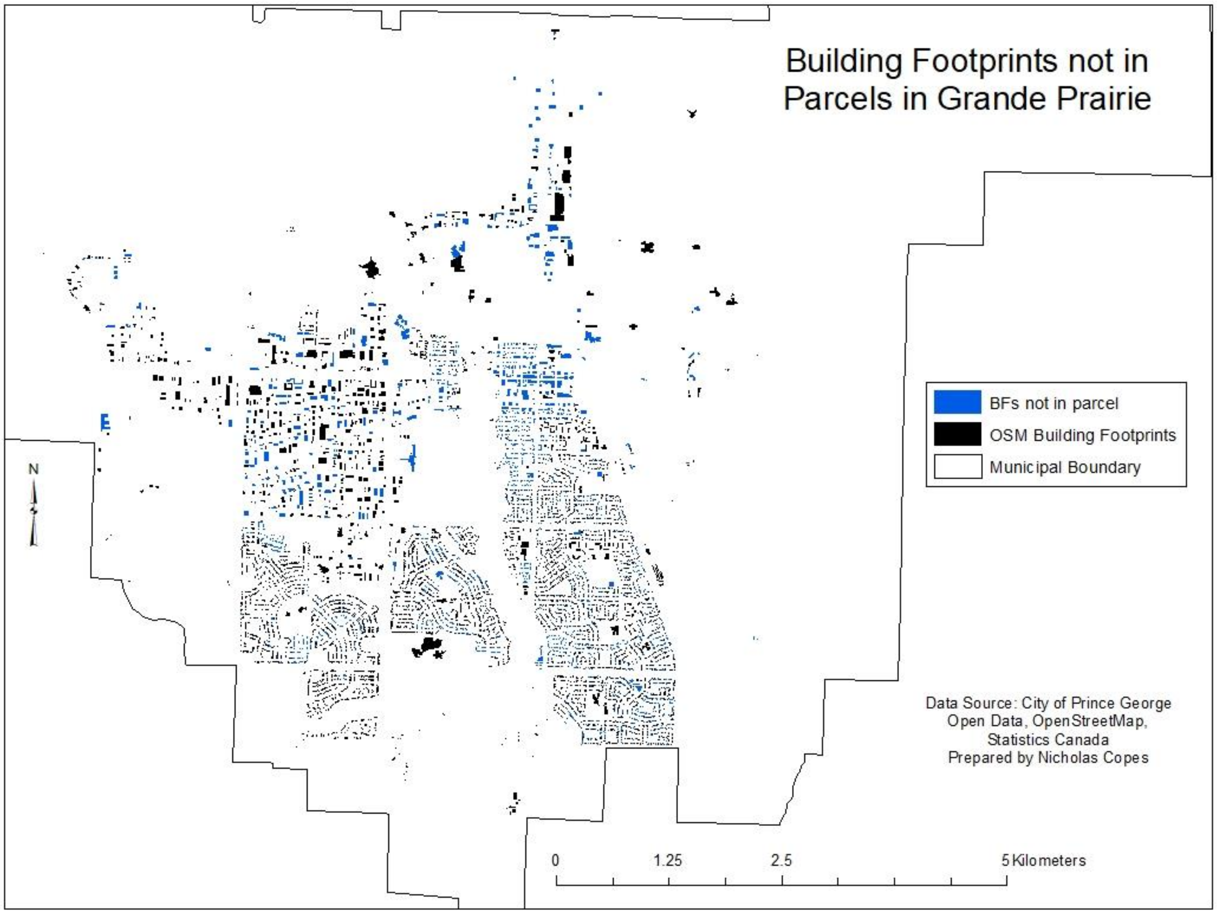


Figure 4.2 – Building Footprints Not Within Parcels in Grande Prairie

- 4.23 Evaluation of Quality Metrics

The quality metrics chosen for this study were based on metrics identified through the literature review as well as those that were simple to implement within the ArcGIS model builder environment. It is worth examining some of the results to see if the quality metrics were useful for their intended purpose. Table 4.2 demonstrates the results of each quality metric across the study municipalities. One metric of note is the percentage of building footprints not contained within a parcel. The values range from 21.93% to 46.78%. These values seem high and suggest that the positional accuracy of many buildings is incorrect. Many authors have noted that reference data, such as parcels are of higher quality than OSM data. As noted in section 2.22 parcel data in Canada must conform to the National Standards for the Survey of Canada Lands. The likely reasoning for the high values for this metrics would be the positional offset of OSM buildings. When compared to municipal buildings, a greater percentage of OSM buildings are not contained within a parcel. For example, in Prince George, 20.9% of municipal building footprints are not contained within a parcel versus 46.78% for OSM building footprints. In Kamloops, 10.26% of municipal building footprints are not contained within a parcel versus 27.11% for OSM building footprints. It is worth noting that municipal buildings footprints created from LiDAR data must conform to ASPRS guidelines. These guidelines differ from those for parcels, so there are still some municipal buildings that are potentially “incorrect”, however a greater number of municipal buildings are contained within parcels and thought to be of higher quality. OSM data on the other hand has no verification of positional accuracy when buildings are created, as noted in section 2.24.

Other metrics that should be evaluated are those related to completeness. As a measure of quality, completeness is important. In order to use the data for analysis, it must first be present. Cities with a high level of completeness present greater opportunities to use the data for analysis, especially in areas outside of the city core. Also, it was noted that cities with a high level of completeness scored better in measures of positional accuracy. In this study, Kamloops, Saint John and Stratford had some of the lowest values for building footprints not contained within a parcel and mean near distance, whilst having the greatest levels of completeness. In contrast, Lethbridge had the lowest completeness, but the highest percentage of building footprints not contained within a parcel and mean near distance. It is likely that cities with more buildings have more contributors and that those contributors contribute more features. This validates the common assumptions that more creators in an area will fix errors and that contributors with more edits are more trusted. In terms of the use of three measures of completeness, Hecht et al. (2013) noted that this provides a more accurate assessment. The centroid method of completeness isolates OSM building footprints that are more positionally correct as they contain the centroid of the municipal building footprint. OSM building footprints that do not meet this measure should be noted as inaccurate likely due to their position or size. While a percentage overlap function would also be of use, one would have to define the percentages that are acceptable (which could vary) and isolate those inaccurate buildings, whereas the centroid method is consistent and automatically excludes inaccurate buildings.

In terms of the importance of each metric, there is no defined weight for each one. That being said, completeness, especially by number of building footprints is though to be the most important measure. This is because having the buildings present is necessary for analysis. While measures of positional accuracy are also important, it was noted that cities with high completeness also had higher levels of positional accuracy. Scoring highly in both of these categories will rank a municipality higher up. Having a complete inventory of buildings in the right location means they can be used for analysis without too much worry. Attribute accuracy is also important for qualitative analysis. That being said, municipalities with higher attribute accuracy tended to have few buildings, so that negates the usefulness. Measures of topological accuracy such as overlap and roads and buildings that touch are important as they represent errors in the creation of data. In this study, however, these measures were insignificant as the reported cases were non-existent or very low in all municipalities.

It is worth noting that the quality of OSM data is constantly changing as new buildings are added and existing buildings are modified. Due to the nature of OSM data, completeness could remain the same over a certain time period, while measures such as positional accuracy could either improve or decrease depending on the nature of edits being performed on existing buildings. While having many editors will likely improve the quality of existing features over time, they may get worse (through an incorrect edit) before getting better. For example, Touya et al. (2017) noted that the point locations of subway station entrances in Paris varied considerably in terms of the actual distance from the entrance and that many duplicate (incorrect) points existed. Additionally, attributes may be added or changed over time. It is worth noting that evaluating OSM data quality is different than evaluating authoritative GI. Authoritative data must conform to set standards as noted in section 2.22 and is released at a set time. This means that the quality of authoritative data is constant over time, until an updated version is released. While the quality measures used in this thesis are important, one should evaluate the quality of OSM data at the time they intend to use it, as it changes over time. The results presented in this thesis note the quality at a specific point in time and the results may de different if the tests were performed again with more recent data.

- 4.24 Evaluation of Models

One of the objectives of this thesis was to develop simple to use models to evaluate the spatial data quality of building footprints from OpenStreetMap. The models used performed certain quality tests well, while other quality tests had certain issues associated with them. To begin, the measurement of completeness worked well. Two of the completeness measures, being the area completeness and completeness by number of building footprints were determined without using the models, simply by rereferring to the attribute tables of both the municipal and OSM building footprint datasets. The centroid method calculation was performed by the model and worked well by isolating the municipal building footprints that had their centroid in an OSM building footprint. Other tests that worked well was the isolation of commissions and the isolation of building footprints that were not contained within a parcel. Both of these were exported as new feature classes. Overlapping buildings were easy to identify as were the three categories of attribute accuracy. Buildings with a name, street name or building type were each extracted as a new feature class.

While the models worked well overall, there were some issues that were identified with certain tests. First off, while commissions were easily identified, it is up to the user to determine if a commission is really an error or simply a newly added building. Figure 4.3 shows commissions that are in fact new commercial stores.

Figure 4.3 – Commissions in Niagara Falls

Another issue with the models is with assigning building footprints to the zoning categories. Many buildings, particularly those from OSM, lie within two or more zoning categories. As such, the model is set to include them in all zoning categories in which they lie. The other option would be to exclude buildings that do not lie completely within a zone. Due to the nature of the data, this option was not chosen as it would exclude too many buildings from the zonal completeness calculations. A likely reason for buildings lying in more than one zone is the fact that many OSM buildings are not contained within a parcel. This may be due to simple geographic inaccuracy when the building footprints were digitized. The parcels were likely digitized around the municipal building footprints, which are usually offset from the OSM ones. Another issue is for townhouses. Sometimes they are digitized as a single building that crosses multiple parcels. Figure 4.4 demonstrates the issues with OSM building footprints lying in multiple zones and multiple parcels.

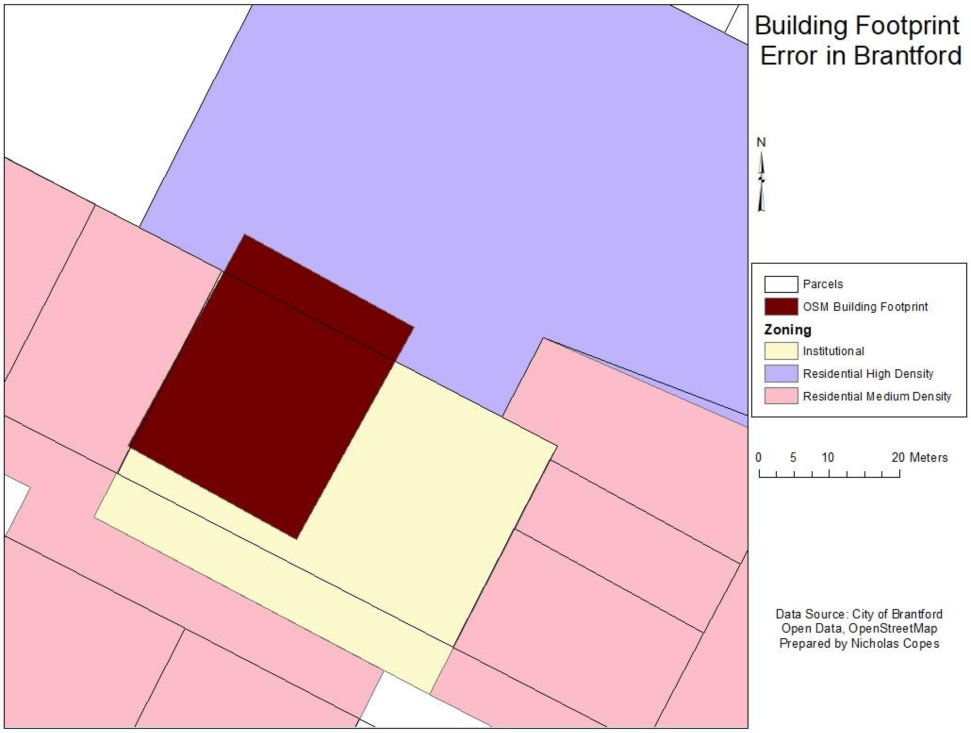


Figure 4.4 – OSM Building Footprint Location Errors

Another issue worth noting is with the relation between datasets. Many of the tests were designed to be used for building footprints with a 1:1 relation, meaning that one OSM building footprint is represented by one municipal building footprint. That being said, there exist some buildings with a 1:n or n:1 relation, meaning that they are represented as one building in one dataset but as two or more buildings in the other dataset. The issue with this is that it will affect the mean near distance calculation. The mean near distance calculates the distance between centroids, however, if one dataset has two or more centroids that correspond to only one centroid in the other dataset, the distance will be calculated to both those centroids. This will increase the mean near distance. The frequency table is designed to identify the number of n:1 relations. An example of this error can be seen in figure 4.5.

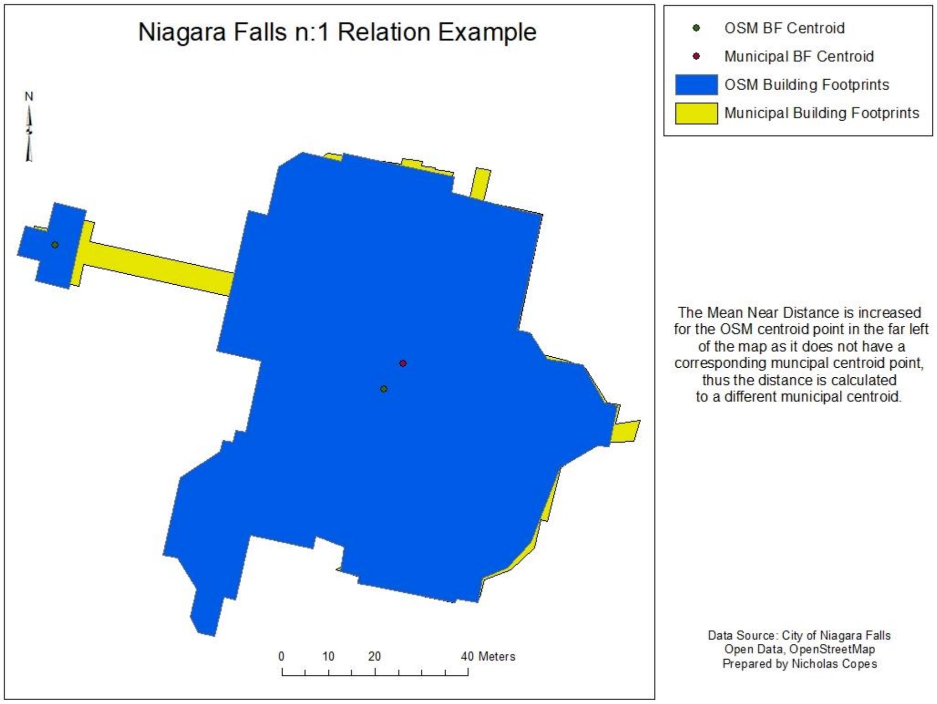


Figure 4.5 – n:1 Relation in Niagara Falls

- 4.25 Comparison of Zonal Completeness

While each municipality has different names for their zones and a different number of zones, there are still some similarities and differences when comparing the zonal completeness across the study municipalities. First of all, commercial zones almost always had the highest level of completeness or were very complete across all study areas. In areas where institutional was a category, it usually scored high on zonal completeness as well. Residential areas were among the lowest scorers in zonal completeness. One trend noted is that in municipalities with a high-density category, this category scored much higher than low density categories. Other categories that scored low across the study municipalities were rural and agricultural, with the exception of Stratford, which had a high agricultural completeness.

There were also some differences in zonal completeness across the study municipalities. First off, the municipalities of Kamloops, Saint John, Stratford and Grande Prairie had much higher levels of completeness that other municipalities and scored well across many zone types including residential. Industrial zones were one that varied in completeness across municipalities. In Halton Hills and Lethbridge industrial zones had very low completeness. In Niagara Falls, Prince George and Chilliwack, industrial zones had moderate completeness between 21 and 35% based on number of building footprints. Kamloops, Saint John, Grande Prairie and Brantford had high industrial completeness between 55 and 91% based on number of building footprints. Refer to Appendix 2 for zonal completeness tables by city. Appendix 3 contains maps of zones for each municipality.

- 4.26 Completeness by type of building

The type of building can be an important factor in determining an area's accuracy and completeness. Generally, contributors focus on commercial and institutional buildings before residential ones. Areas with higher concentrations of these buildings are often more complete and accurate as many people are familiar with local businesses and institutions. Residential buildings are often mapped by people who do not live in them (or nearby) and thus sometimes they are incorrectly mapped. Common errors can include touching houses, semi-detached house shown as a single house etc. Getting homeowners informed and interested in OSM can help them to map their own homes, which will lead to increased completeness and accuracy.

This study hypothesized that commercial and institutional zones would have a greater level of completeness than residential zones. This hypothesis turned out to be true. In general, all of the study municipalities had a reasonable level of completeness for commercial and institutional buildings. This study also predicted that city cores would have higher levels of completeness. This turned out to be true, as usually these areas are occupied by commercial and institutional zones. In many cities, however; residential completeness remains low. When looking at the completeness over time, commercial and institutional buildings were usually added before residential ones. Even in cities with high residential completeness, the residential areas were usually added last. Another interesting discovery is that universities are often some of the first buildings to show up in a city. This was the case in Prince George and Lethbridge where the university buildings were the first ones added. See figures 4.7 to 4.25 for information on when and where (in each municipality) buildings were added.

4.3: Completeness over time

Explanation of tables and figures: Table 4.3 shows the number of building footprints in each of the ten study municipalities from Jan 2010 to July 2018 for each six-month period. Also, the number of OSM authors for each city as of July 2018 is stated. Figures 4.6 to 4.24 (even numbered) show a line graph that represents the progression of building footprints over time as a percentage of the total number of municipal building footprints. Figures 4.7 to 4.25 (odd numbered) show a map that represents the locations of building footprints over time in each city. The legend is colour coded to show when the building footprint first appeared. An inset map shows the city centre or other significant area.

Table 4.3: Building Footprints over time 2010-2018

City → Date ↓	Halton Hills	Niagara Falls	Kamloops	Prince George	Chilliwack	Saint John	Lethbridge	Grande Prairie	Stratford	Brantford
Jan 2010	36	5	0	13	69	3	0	0	0	0
July 2010	38	19	0	13	69	40	0	0	0	0
Jan 2011	214	75	0	13	69	106	0	0	173	10
July 2011	217	79	71	311	76	121	58	0	173	29
Jan 2012	241	79	9520	311	127	414	65	0	174	35
July 2012	251	75	10700	300	174	8216	66	105	181	36
Jan 2013	278	494	11256	774	179	19057	66	237	186	83
July 2013	437	520	12576	867	180	19188	66	323	190	759
Jan 2014	463	577	14709	946	180	19289	66	447	269	766
July 2014	490	637	14799	949	189	19286	149	448	414	795
Jan 2015	512	648	14799	950	187	19287	258	541	692	818
July 2015	534	719	17848	962	193	19289	513	691	794	858
Jan 2016	554	969	19036	977	210	19293	535	730	810	866
July 2016	605	1033	19013	1106	238	19306	606	736	818	1017
Jan 2017	725	1612	23937	1153	248	19306	828	13142	825	1355
July 2017	776	1848	25110	1164	376	19307	905	13143	843	1435
Jan 2018	822	2097	27621	1392	480	21062	910	13150	6190	1710
July 2018	821	2109	27331	1410	951	21089	912	13150	6270	1716
# Authors (07/2018)	31	43	54	41	29	52	32	42	40	31

When looking at each city, certain trends can be spotted in the expansion of the building footprint dataset from OSM. Halton Hills demonstrated the most consistent growth in the number of building footprints over time, however, the overall number remains low. In 2010, all the buildings in Halton Hills were located in Georgetown (the main community). Starting in 2011, buildings started showing up in rural areas. In 2012, a few buildings started appearing in Acton (the secondary community). For July 2013, many new buildings appeared in Georgetown and Acton, although they were still centrally located, non-residential buildings. Further buildings were added since then, primarily in other parts of Georgetown. Residential completeness remains low. In July 2018, there were 31 OSM contributors.

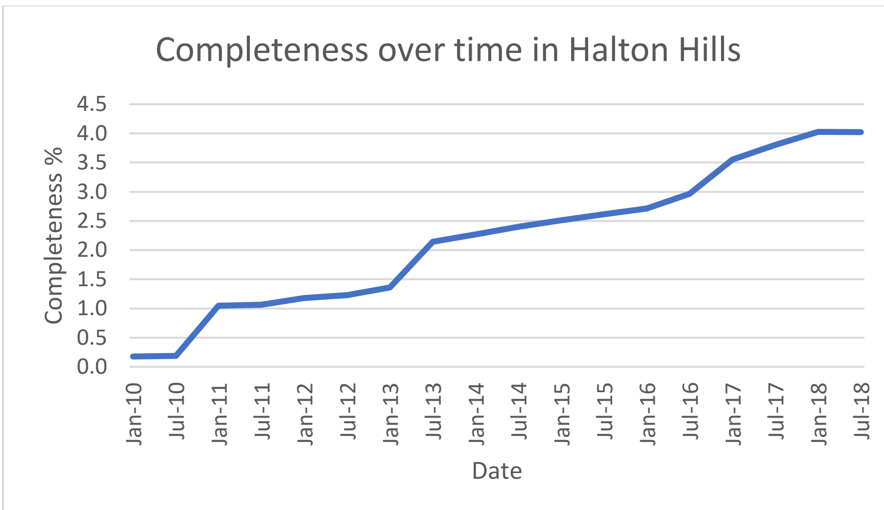
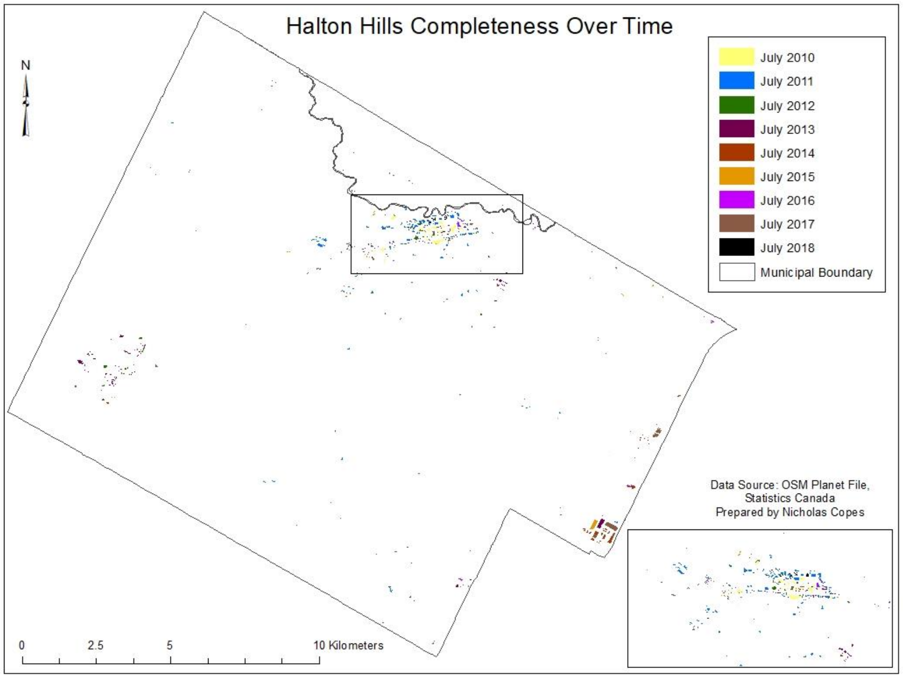


Figure 4.6 – Completeness over time in Halton Hills Graph



Figures 4.7 – Completeness over time in Halton Hills Map

Niagara Falls saw a jump in the number of building footprints in late 2012 from 75 to 494 and again in late 2016 with a jump from 1033 to 1612. Moderate growth occurred in between those times and after the second growth spurt. In 2010, the few buildings present were located near the tourist area. In 2011, some industrial buildings in the South-West part of the City were added. In Jan 2013, many more buildings were present throughout the city. For Jan 2017, many smaller buildings were added primarily along major streets in the tourist area. Residential completeness remains low. In July 2018, there were 43 OSM contributors.

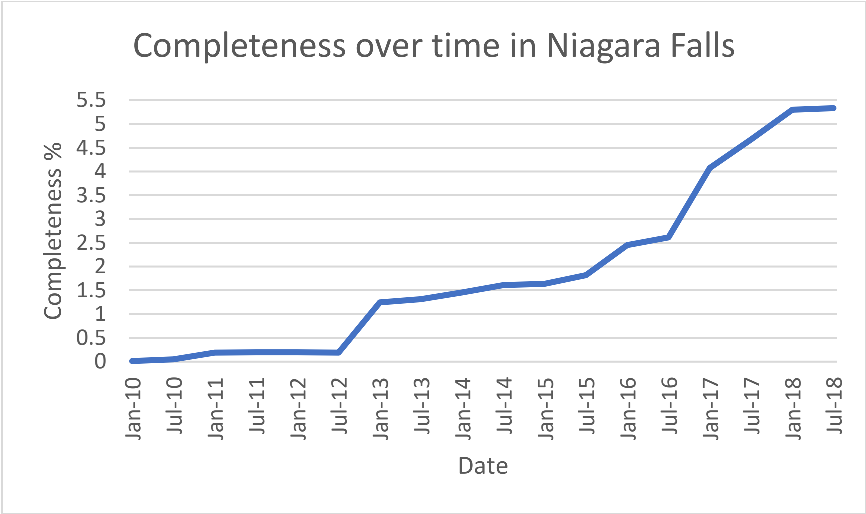


Figure 4.8 – Completeness over time in Niagara Falls Graph

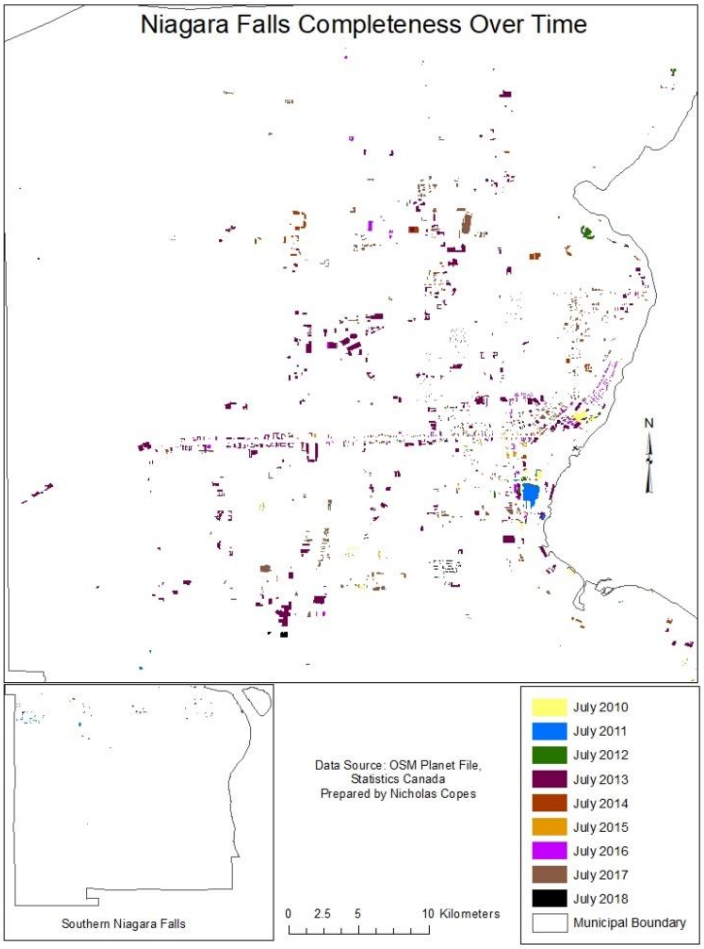


Figure 4.9 – Completeness over time in Niagara Falls Map

Kamloops was late to receive any building footprints but saw a massive growth early on in late 2011 receiving 9520 building footprints by Jan 2012. The number of building footprints remained stagnant during 2014. Another large spurt of growth occurred in late 2016 bringing the number from 19013 to 23937. Moderate growth has continued since then. In July 2011, there were few buildings scattered around Kamloops. For Jan 2012, a massive increase in building footprints occurred primarily in centralized corridors. For Jan 2014, a new residential cluster of buildings was added in the west side (Brocklehurst). For July 2015, two new residential areas in the south received new building footprints (Aberdeen and Juniper Ridge). For Jan 2017 many new buildings were added, primarily in North Kamloops and Valleyview. Further buildings on the outskirts have continuously been added since 2017. In July 2018, there were 54 OSM contributors.

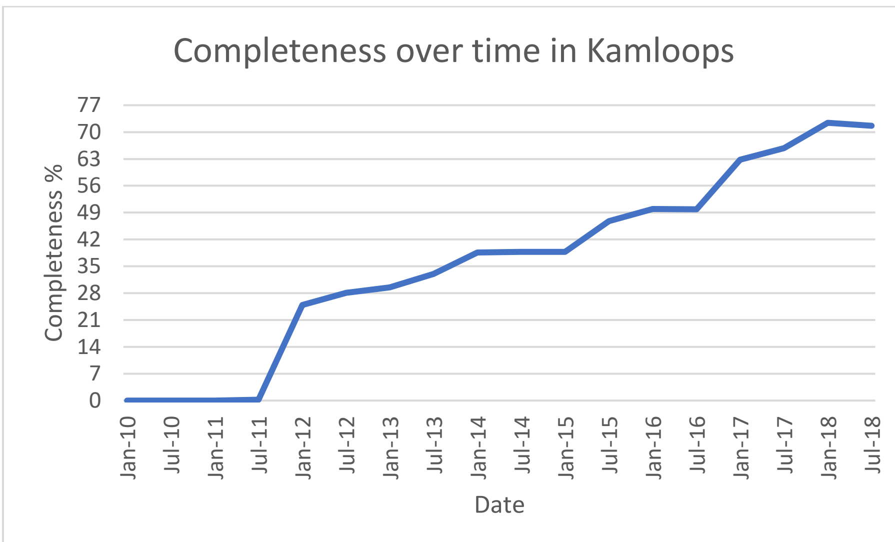


Figure 4.10 – Completeness over time in Kamloops Graph

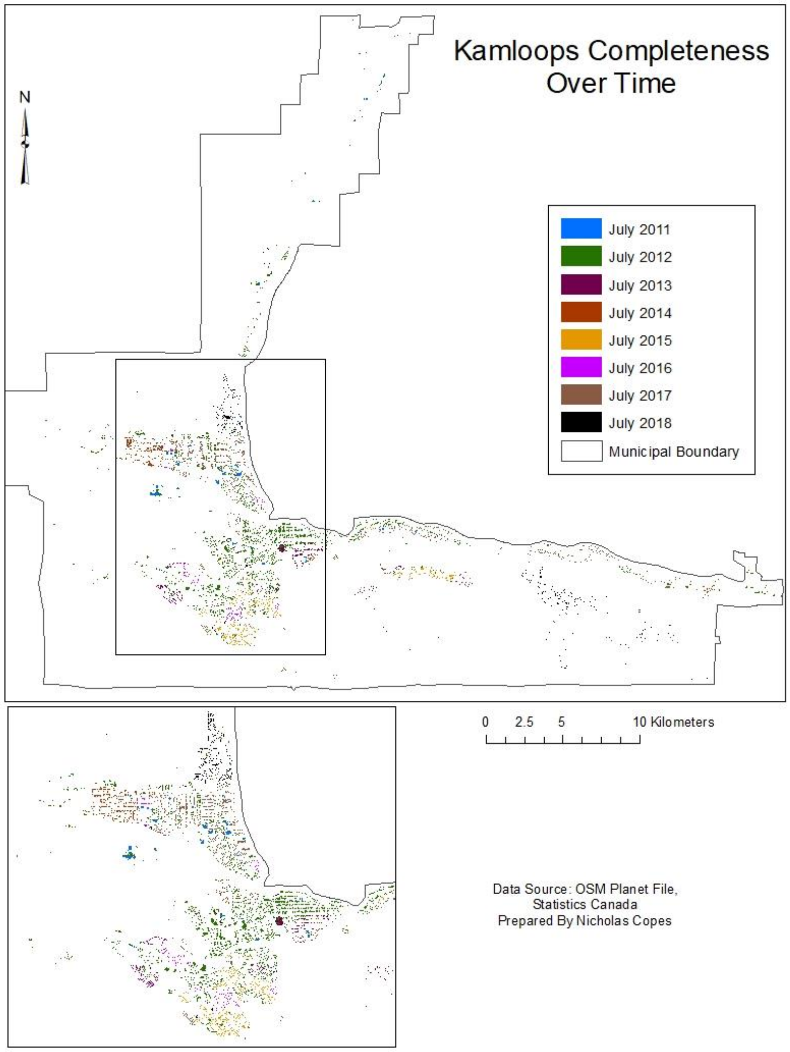


Figure 4.11 – Completeness over time in Kamloops Map

Prince George saw growth in building footprints in late 2012 bringing the number up to 774 by Jan 2013. The number of building footprints remained stagnant during 2014 and 2015 after which moderate growth occurred. The first 13 buildings in Prince George were all located at the University of Northern British Columbia. More buildings were added for July 2011 mainly along major roads. Many new buildings downtown were added for Jan 2013. One small residential area was added for July 2017 along the Nechako River. For Jan 2018, a new industrial area was added. In July 2018, there were 41 OSM contributors.

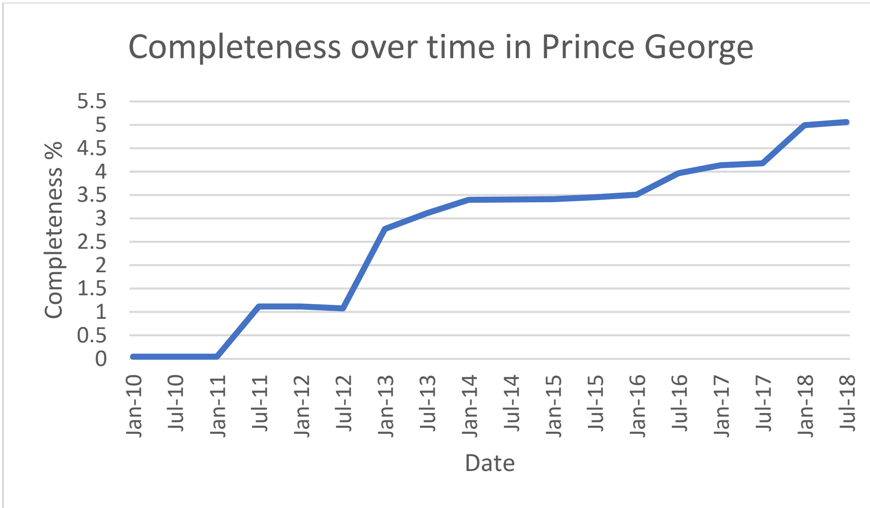


Figure 4.12 – Completeness over time in Prince George Graph

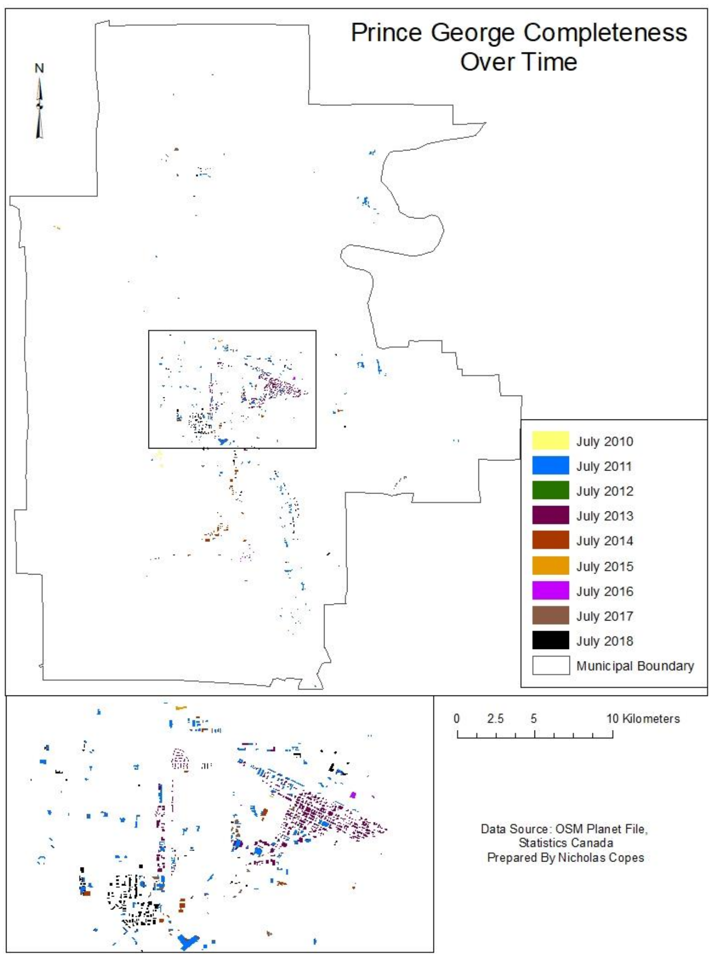


Figure 4.13 – Completeness over time in Prince George Map

Chilliwack was late to receive growth in building footprints with only a small amount remaining stagnant until 2015. After July 2015moderate growth occurred with a large spurt of growth occurring in 2018. In 2010, there were a few randomly scattered buildings. For July 2011, some retail buildings were added. In 2012 a few more buildings were added in the central area. July 2016 saw a few more scattered buildings added. July 2017 saw more buildings added in the central area. July 2018 saw the addition of many commercial buildings along the highway as well as a few buildings in rural areas, notably near Ryder Lake. In July 2018, there were 29 OSM contributors.

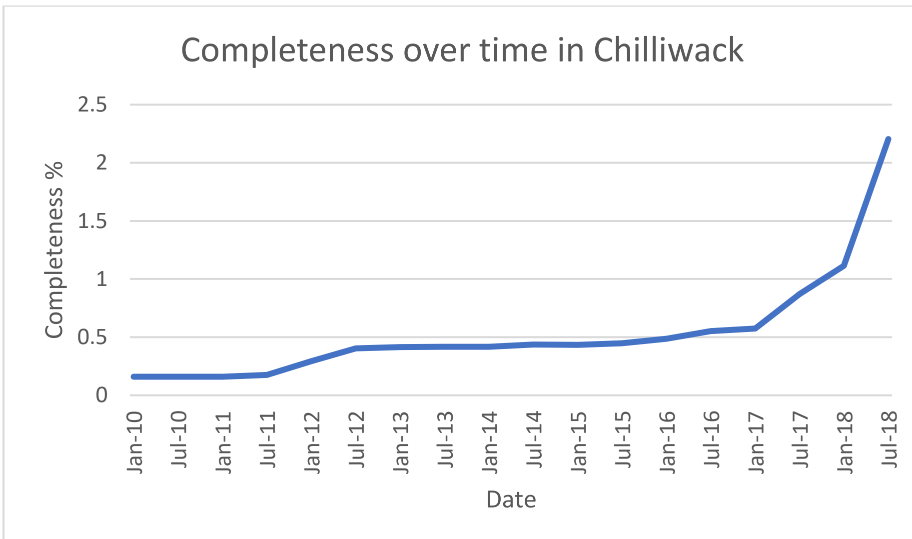


Figure 4.14 – Completeness over time in Chilliwack Graph

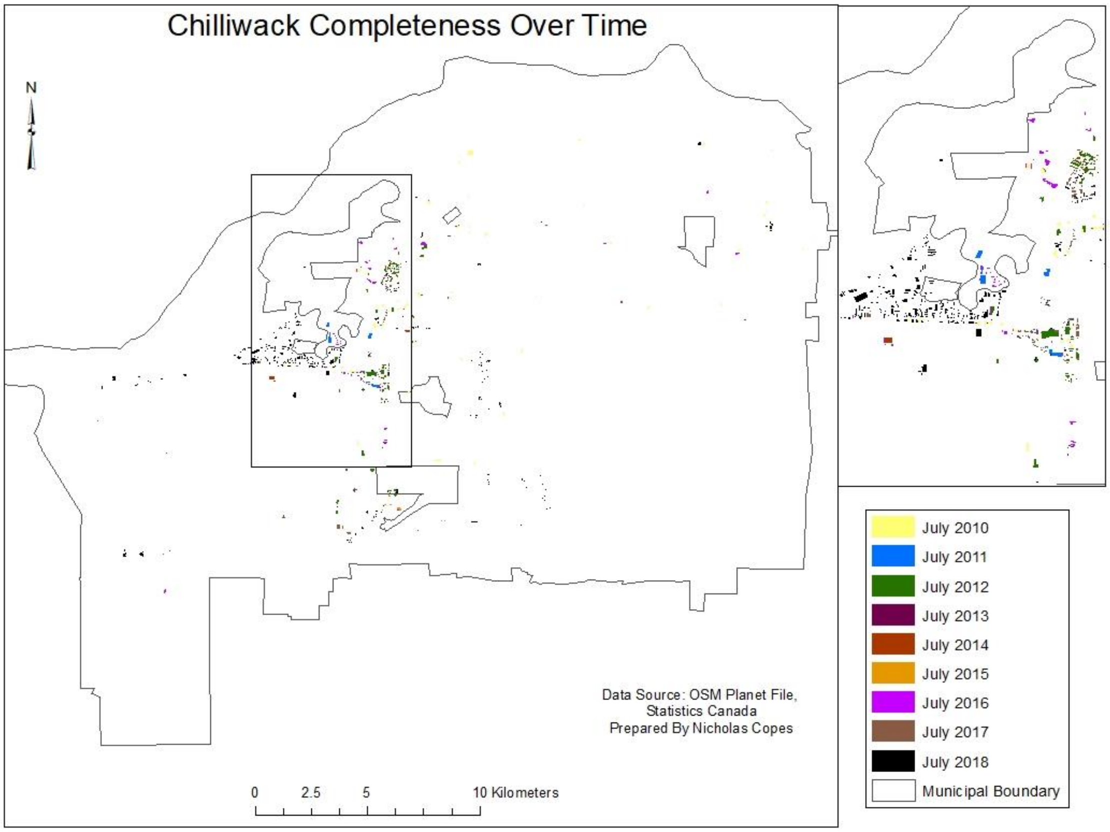


Figure 4.15 – Completeness over time in Chilliwack Map

Saint John received an early growth spurt of building footprints throughout 2012 increasing the number from 414 in Jan 2012 to 19057 in Jan 2013. The number of building footprints remained stagnant after that until late 2017 when the number of building footprints increased to 21062 for Jan 2018. This increase may be a result of Saint John being on the Canada OSM Tasking Manager as medium priority, starting in late 2017. In Jan 2010, Saint John only had 3 buildings. A few more were added near the central area during 2010 and 2011. For July 2012, the downtown and Saint John West areas became mostly complete along with a few residential areas. For Jan 2013, central areas became more complete and many new residential areas were completed around the city. For Jan 2018, new buildings in rural areas were added, most notably in the community of Harbourview in the South East. In July 2018, there were 52 OSM contributors.

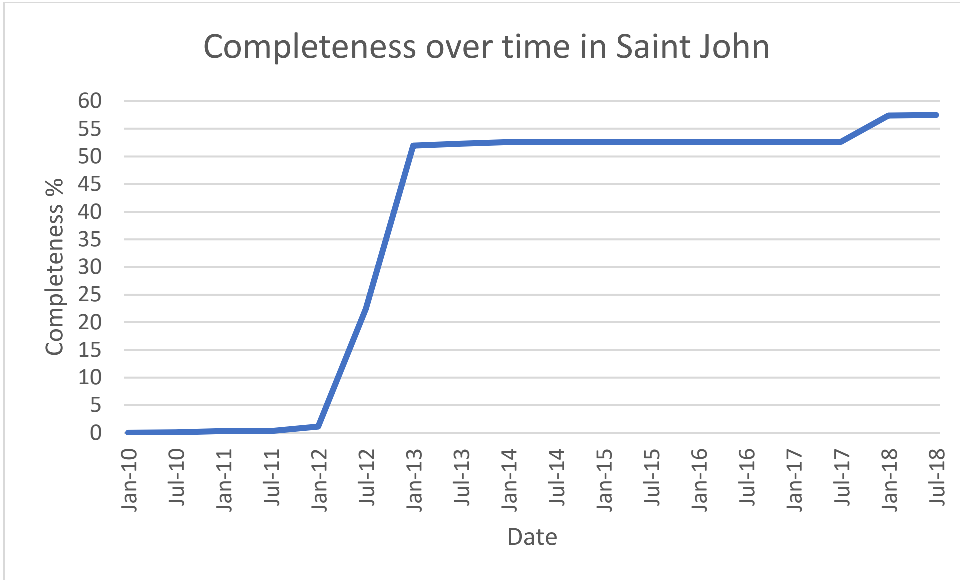


Figure 4.16 – Completeness over time in Saint John Graph

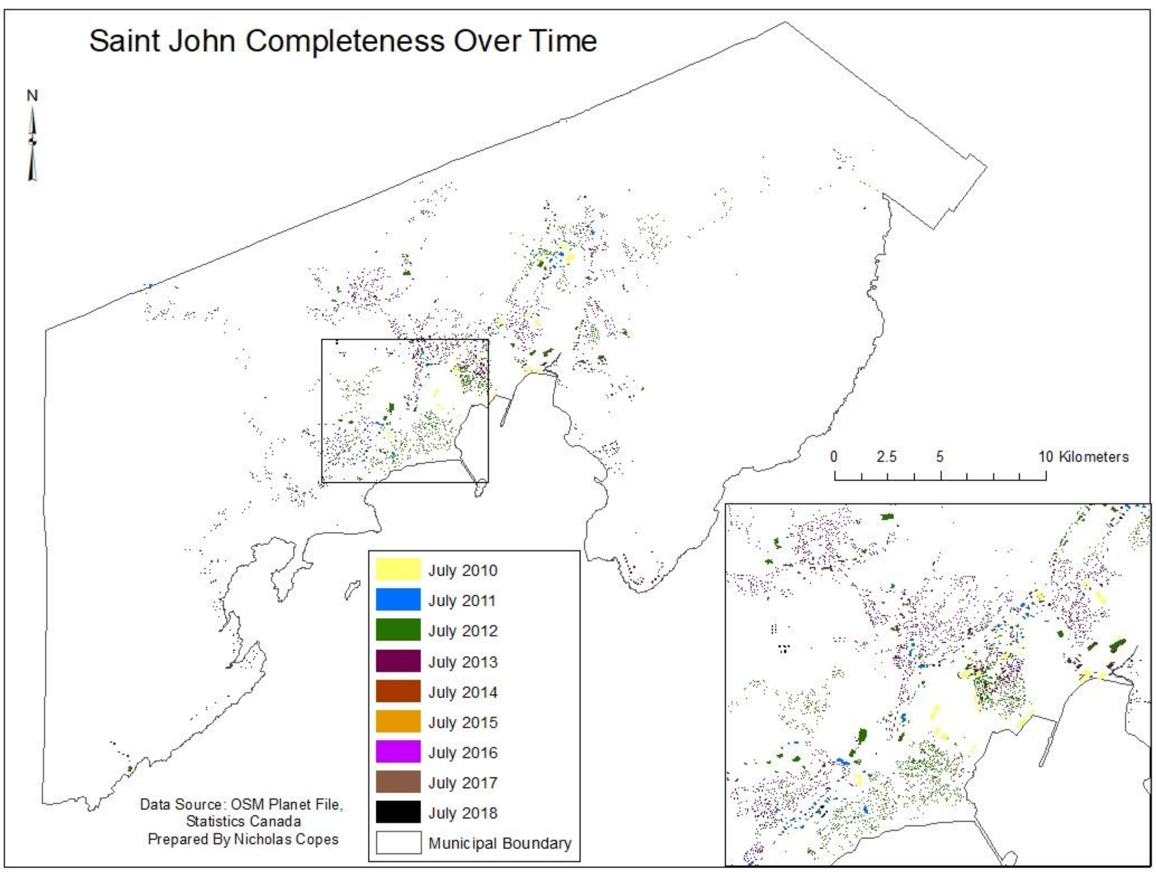


Figure 4.17 – Completeness over time in Saint John Map

Lethbridge had a very low number of building footprints and did not start to see growth until 2014 after which moderate increases occurred over time. Lethbridge had a few buildings starting in July 2011, all on the west side of the Old Man River, notably the buildings at the University of Lethbridge. A couple of big box stores in the South East were added for Jan 2012. A shopping mall and more big box stores were added for July 2014. More buildings started showing up downtown in 2015. Since then sporadic buildings have been slowly added around the city. Residential completeness remains low. In July 2018, there were 32 OSM contributors.

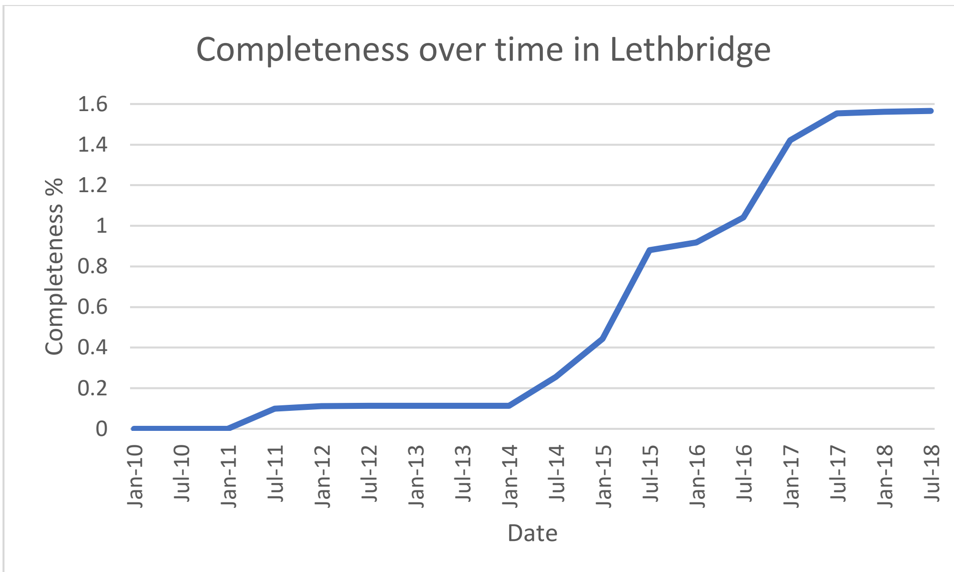


Figure 4.18 – Completeness over time in Lethbridge Graph

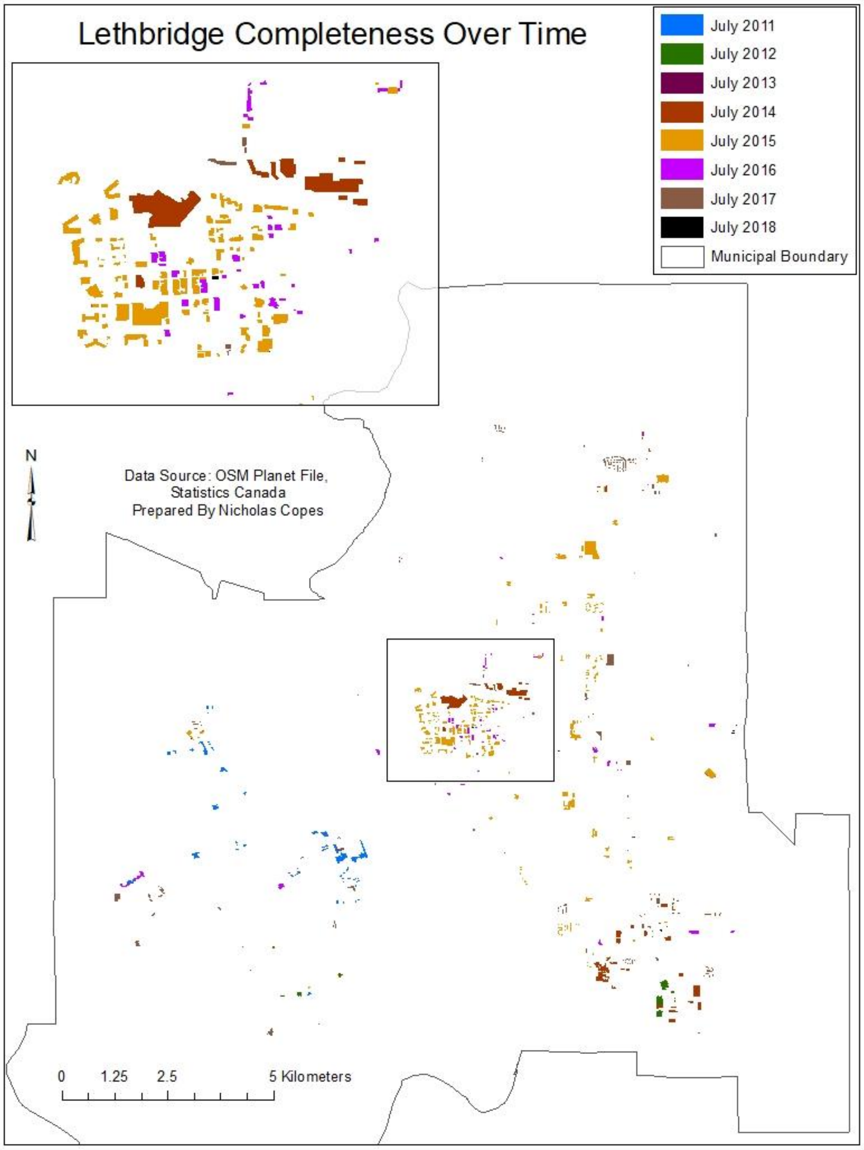


Figure 4.19 – Completeness over time in Lethbridge Map

Grande Prairie was the latest to receive any building footprints but had consistent growth from 2012 to mid-2016 when a massive growth spurt brought the building footprints from 736 in July 2016 to 13142 in Jan 2017. The number of building footprints has remained stagnant since 2017. In July 2012, there were a few residential buildings near downtown. During late 2012, 2013 and 2014, many industrial and commercial buildings added. July 2015 saw an addition of more commercial buildings around the city. For Jan 2017, numerous residential areas were added, almost all of them in the southern part of the city. While many residential areas are complete, the residential areas north of downtown as well as those east of the railroad tracks remain virtually untouched. In July 2018, there were 42 OSM contributors.

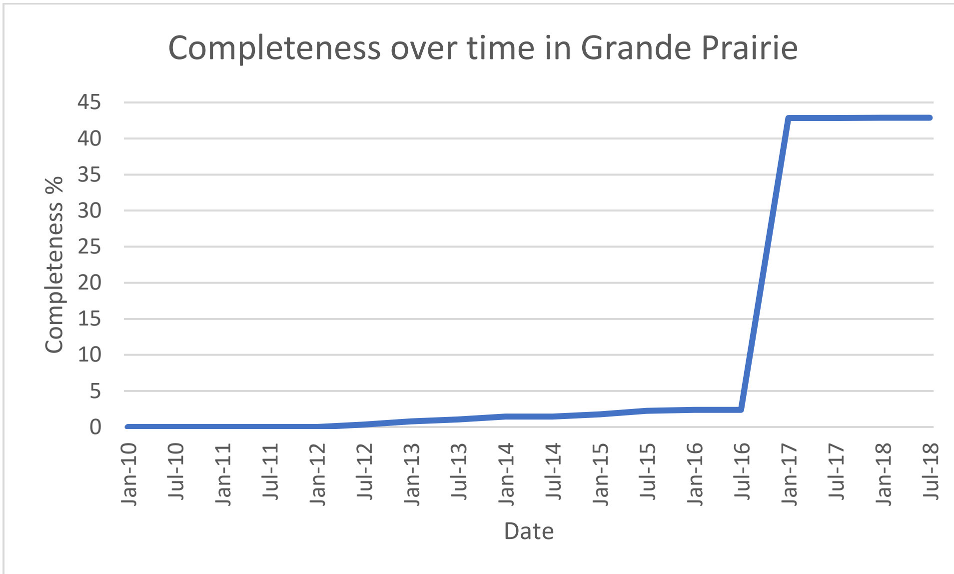


Figure 4.20 – Completeness over time in Grande Prairie Graph

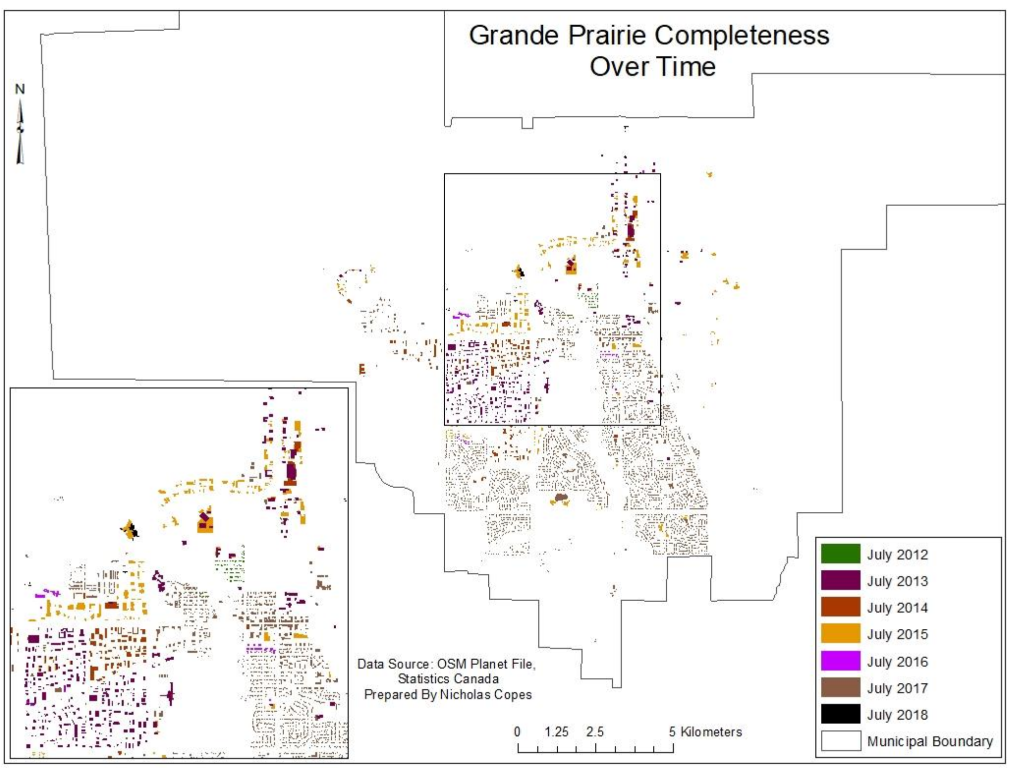


Figure 4.21 – Completeness over time in Grande Prairie Map

Stratford demonstrated a slow growth over time in the number of building footprints before receiving a large growth spurt in late 2017 bringing the number of building footprints from 843 in July 2017 to 6190 in Jan 2018. Like Saint John, this growth in late 2017 may also be attributed to the city appearing on the OSM Canada Tasking Manager. In addition, students from the University of Waterloo took part in mapping parties to increase building footprint completeness in Stratford. In 2011, Stratford had a few commercial buildings along main streets. For Jan 2014, some industrial buildings in the south were added. From 2014 to 2017 many new buildings were slowly added along the main streets. For Jan 2018, many new residential areas were added but remain only partially complete. In July 2018, there were 40 OSM contributors.

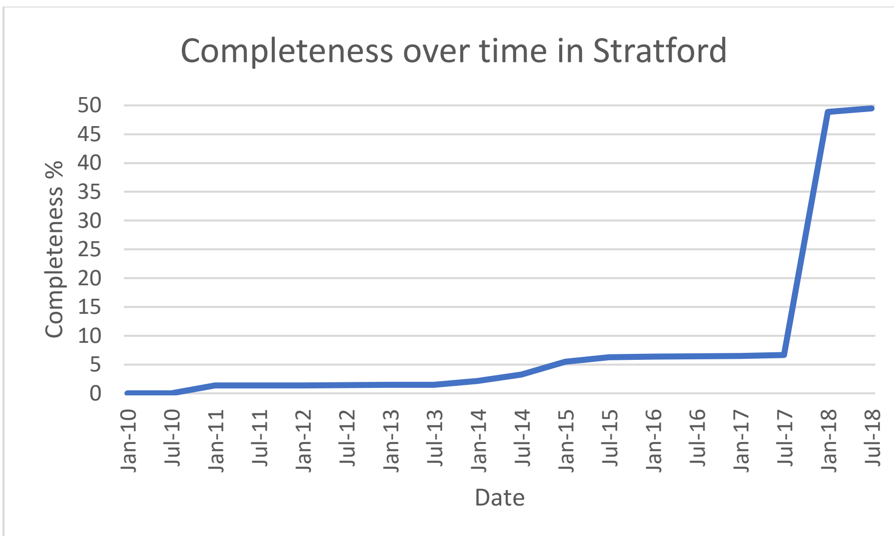


Figure 4.22 – Completeness over time in Stratford Graph

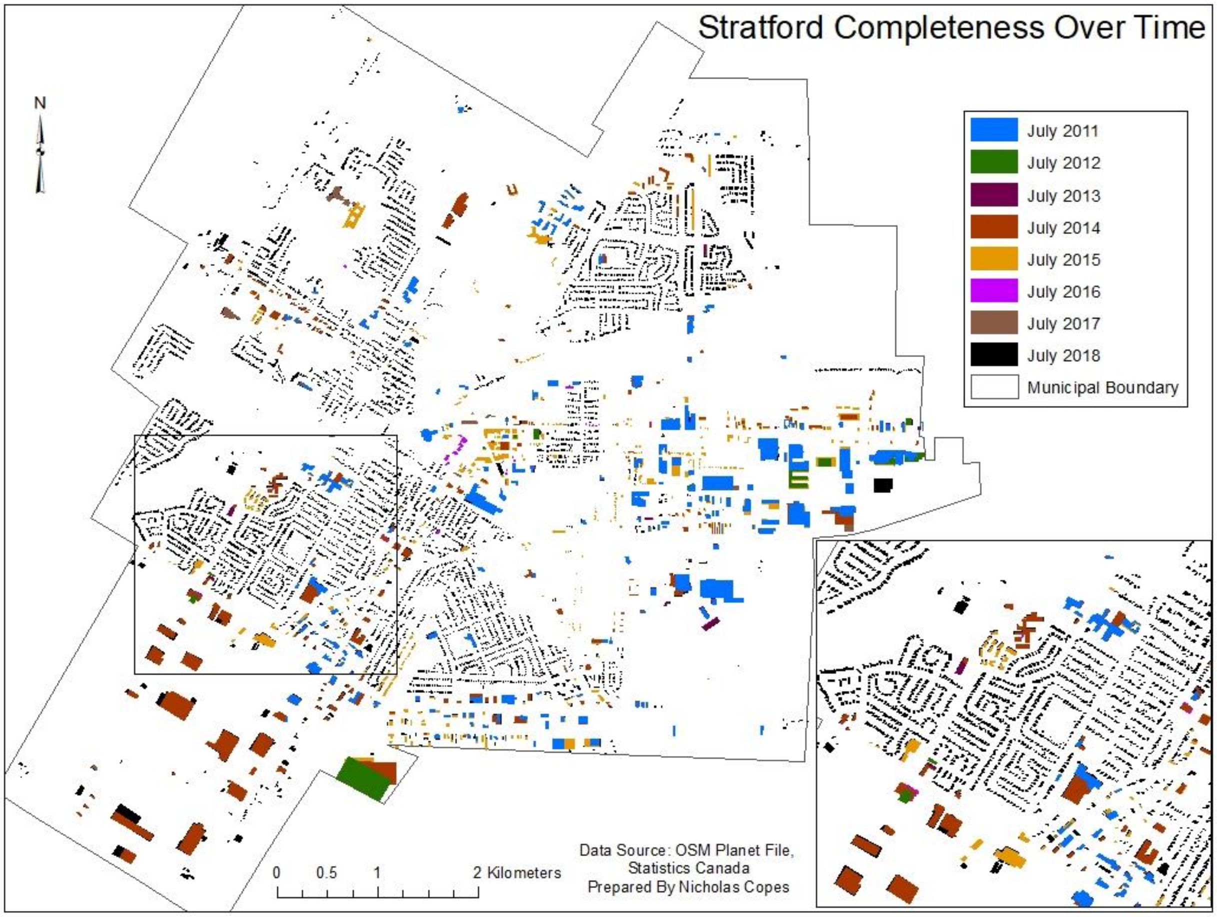


Figure 4.23 – Completeness over time in Stratford Map

Brantford received a mild growth spurt of building footprints in 2013 and since then has had consistent growth. Notably, the number of building footprints in Brantford was at 1716 in July 2018, however the number of building footprints downloaded on Nov. 5, 2018, to be used in this study's models was 2591 indicating the city has received the most recent growth in building footprints; in contrast the number of building footprints taken (Nov 5) for all other study cities was similar to the July 2018 number. In Jan 2011, Brantford started out with a townhouse complex and a school. A few commercial buildings were added before July 2013, when many more commercial, institutional and industrial buildings had been added. More buildings were slowly added over time. For Jan 2017, many more industrial buildings were added in the east. A few small pockets of residential buildings were added since then. In July 2018, there were 31 OSM contributors.

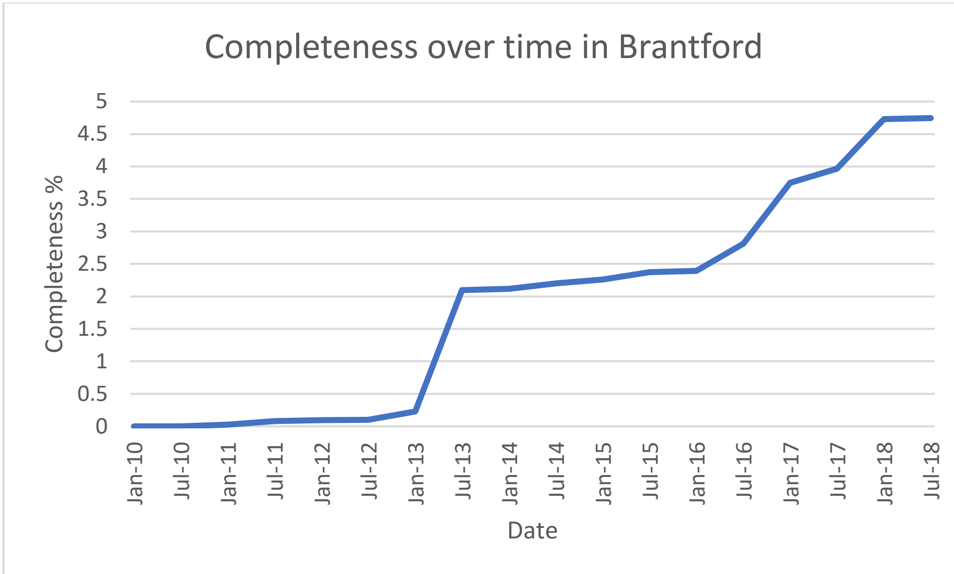


Figure 4.24 – Completeness over time in Brantford Graph

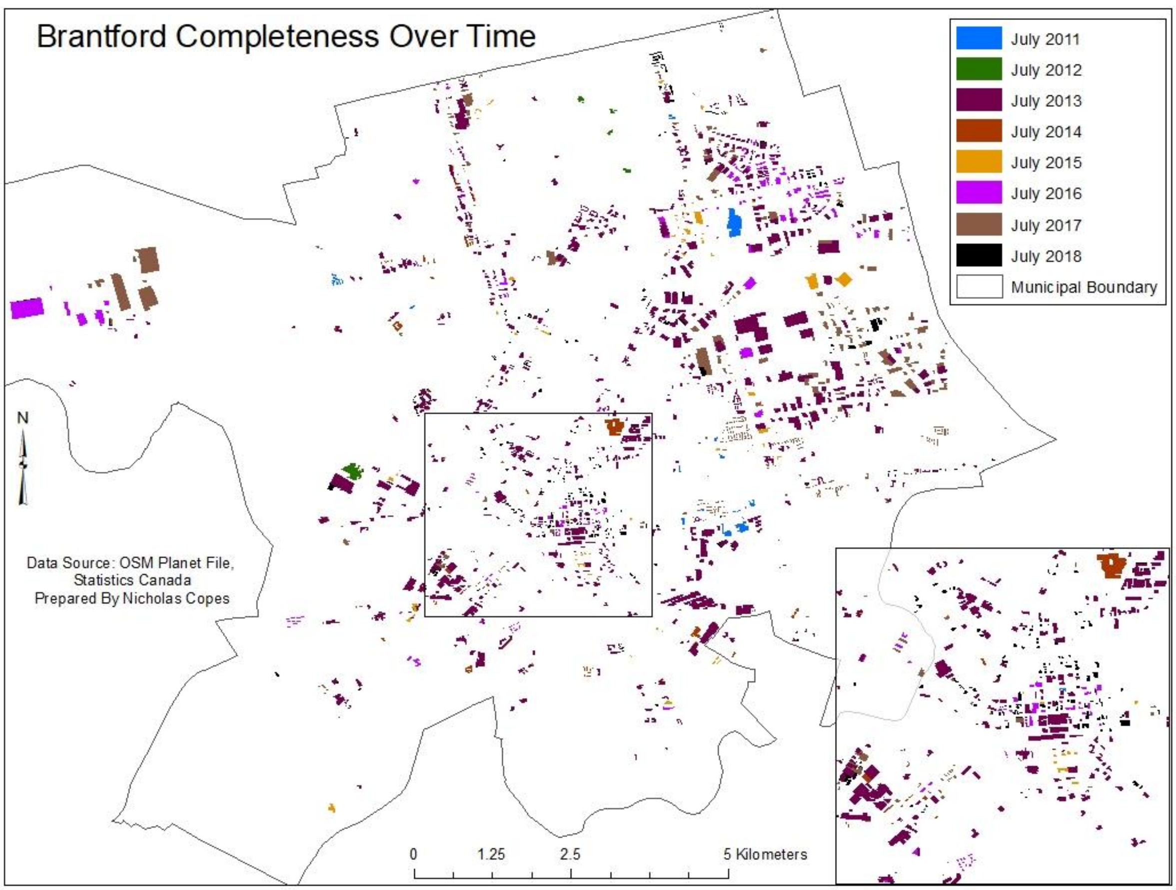


Figure 4.25 – Completeness over time in Brantford Map

Chapter 5: Discussion

This chapter explores the reasons for varying accuracy of OSM data based on factors such as a city's location and perceived importance. Examples from the results section will be included. An introduction to various crowdsourcing projects and their effect on a city's completeness and quality are then discussed. The next section of this chapter discusses the uses of OSM data for planners. Understanding what type of data planners need and ensuring the proper level of quality of this data is important. By having high-quality OSM data available, there are numerous benefits and opportunities for planners which will be discussed.

5.1: Reasons for Varying Accuracy

- 5.11 Local Knowledge and Interest

The accuracy of OSM data can vary across cities for numerous reasons. The amount of local interest and knowledge is a key factor in determining the level of completeness and accuracy. Areas with more people who are knowledgeable and interested in VGI can increase local interest to contribute. Areas with lots of geographers, planners and students can sometimes be more complete. Also, areas with more advanced government data and open data available can spur increased contributions and accuracy as reference data already exists.

- 5.12 Perceived Importance of Cities and Regions

As noted above, the perceived importance of an area can influence completeness and accuracy. This is not only true for cities on the tasking manager, but for many around the world. Barrington-Leigh and Millard-Ball (2017) note from their worldwide OSM study that completeness is greatest in low and high densities. Urban areas with lots of contributors are likely to be complete, while smaller towns and villages are most likely to have incomplete data.

Cities that attract many visitors and have many tourist attractions will often have high levels of OSM data and accuracy. This may be due to the fact that more people have been to the area and are familiar with the buildings there. This goes along with the common assumption that people want to map places that they are familiar with (as noted by Coleman, 2009 and Budhatoki and Haythornwaite, 2013). When looking at OSM, tourist areas and sites are usually mapped to a high level of completeness with many contributors. It might also be assumed that cities in more important areas, such as in large urban metropolises or near a large city might have a higher level of spatial data quality, while more remote cities would be ignored. In addition to perceived importance on a local level, regional variation can also exist. Some might assume cities in a certain region (such as a province) would have similar levels of spatial data quality. This may be based on the level of regional (provincial) data initiatives or the presence of regional geospatial organizations or groupings of contributors. It is worth noting that many of the studies based on regional and urban vs rural completeness as well as completeness of tourist cities were done in the UK, Ireland and Germany (such as Mooney and Corcoran, 2012 and Fan et al., 2014). The results from these studies may not hold true in the Canadian context of OSM.

The results from this study do not coincide with the assumptions on perceived importance. Niagara Falls is a great example. A popular tourist city located in the densely populated Greater Golden Horseshoe; yet it has a low level of spatial data quality. While the tourist area contains many building footprints, the rest of the city is incomplete. This is not unlike other cities which will often have the commercial core buildings but not the rest of the city. Further refuting these assumptions, Kamloops had the highest level of spatial data quality. Kamloops is not a well-known tourist destination, nor is it located in a densely populated area. Grande Prairie is also notable, as it contains a reasonable level of spatial data quality despite its geographic remoteness, hundreds of kilometers away from any important city. No explanatory variables were studied to determine why certain cities are more complete and have better spatial data quality than others in Canada. While studies have been done in Europe, in Canada more research is needed before one can make links between cities of the same size, in the same region or tourist destinations and the level of completeness. OSM wiki notes that cross-locational comparison can be accurate in Europe, but less so in North America. If cross-locational comparison were accurate in Canada, then for this study, one would expect all of these similar sized cities would have similar levels of completeness. While certain factors such as mapping parties can increase completeness in an area, which will be discussed in the next section, some factors influencing spatial data quality remain behind the scenes. It is worth noting that while the assumptions about perceived importance were false in this study, it is not a statistically significant sample and that results may vary depending on the area studied.

When looking at regional variation, it becomes difficult to spot a pattern. In this study, no links were found when comparing spatial data quality on a provincial or regional level. This is because the different regions studied had cities with both high and low overall levels of spatial data quality. In British Columbia, Kamloops scored first among all study cities, while Chilliwack scored poorly. In Alberta, Grande Prairie had a decent level of spatial data quality while Lethbridge scored last. In Ontario, Stratford had a high level of spatial data quality while Halton Hills and Niagara Falls scored much lower. Based on this study's findings, more research is needed before one can determine a certain level of spatial data quality for a city based solely on the region or area where it is located. It is worth noting that the results from this study do not represent a statistically significant sample and that regional links may be different in other areas of the world.

5.2: Crowdsourcing

- 5.21 Statistics Canada's Crowdsourcing Project and OSM Tasking Manager

Statistics Canada's Crowdsourcing project aims to compile complete building footprints and attributes on a national level. So far, Ottawa/Gatineau has been successfully mapped and other communities will be collaborated with in the future. The Building Canada 2020 initiative grew from the Statistics Canada initiative and has the aim to map all buildings in Canada by 2020 (WikiProject Canada). The OSM Canada Tasking Manager is a platform that lists current mapping projects in Canada and encourages user contributions in various communities. Currently, there is a rather *unique* list of communities involved in mapping projects. Users are invited to complete *squares* which are then validated (OSM Tasking Manager). These types of initiatives are proving successful at gaining more complete OSM building footprint coverage. That being said, more needs to be done to raise awareness and motivate people to contribute.

As noted above, the perceived importance of an area can motivate increased levels of OSM contribution. New and experienced contributors may check the OSM tasking manager to determine which areas are most important to contribute to. The OSM tasking manager lists certain mapping projects and ranks them by priority. Generally speaking, higher priority listings will attract more contributors and leads to increased contributions for that project. It is still somewhat of a mystery as to the determination of what projects are put on the list and the ranking of priority. On the OSM Canada tasking manager, there is currently a unique list of seemingly random communities that have priority. In monitoring this list, North Battleford, Saskatchewan has been given high priority and has now received 100% completeness. Strangely, other communities on the list (including bigger cities) have yet to attract many contributions. Based on the study in this thesis, appearance on the OSM Canada tasking manager can influence a city's completeness and quality. Two cities in this study, Saint John and Stratford appear on the OSM Canada tasking manager and both had high levels of completeness and quality. There is also the HOT Tasking Manager (worldwide) allows contributors to map areas affected by recent natural disasters and assigns urgency to each case.

- 5.22 Mapping Parties and Contributors

Another factor influencing completeness and accuracy of OSM data is the presence of mapping parties. A mapping party is a form of crowdsourcing consisting of a grouping of contributors who all contribute to the same area during the same time period, thus greatly improving completeness. As part of GIS Day 2017, the University of Waterloo hosted a mapping party for Stratford, Ontario. Furthermore, students in a GIS course were also focused on mapping in Stratford as part of the course. As a result, Stratford has a very high level of completeness, more so than many similar-sized cities in Canada. Mapping parties should be encouraged around the world. They should be included in GIS, geography and planning courses at universities and colleges as they provide a great way for students to learn about and contribute to VGI and help improve the amount of free geospatial data available.

Many people, even those with geographic knowledge are not aware of OSM. Promotion on social media and other platforms viewed by large numbers of people (not just those in the OSM community) is necessary. Also, mapping parties are a good way to increase completeness, as seen in Stratford. One idea is to incorporate OSM mapping projects into more university geography and planning courses across Canada. This will teach students about OSM and VGI and allow people with knowledge and interest in geography to contribute to their local area (either University City or their hometown).

Different levels of government can help contribute to OSM. Municipal governments can contribute their own datasets to OSM to improve completeness, such as uploading municipal building footprints. Provincial governments can also contribute data. One example is the government of New Brunswick (NB) which donated high-resolution imagery to ESRI World Imagery which can be used a base to create features (OSM Tasking Manager). This has likely encouraged many contributions in NB cities, leading to a confirmed high regional level of completeness. This can be seen in this study as Saint John had a very high level of completeness. Federal governments can also get involved. As stated above, Statistics Canada's crowdsourcing project is a federal initiative that has the goal to map all buildings in Canada. This has surely encouraged many contributions. In November 2018, Statistics Canada also launched the Open Database of Buildings. This database includes data from 61 datasets from various organizations consisting of 4.3 million records. Buildings can be downloaded on a provincial or territorial level. While all of these buildings may not yet be present in OSM, the fact that these open source building footprints are now available, combined with Statistics Canada's 2020 goal, likely means that they will make their way to OSM.

Many authors such as Goodchild and Li (2012) have suggested that having more contributors leads to increased quality of data. Based on this study, this suggestion seems valid. Kamloops and Saint John had the highest number of contributors as well as the highest overall levels of spatial data quality. Chilliwack had the lowest number of contributors while scoring near the bottom on spatial data quality. Based on this common assumption, gathering more editors for an area, rather than having only one editor per area should help to increase accuracy. It is worth noting that other factors such as the level of contributions, contributor trust and the general fact that cities with more buildings are likely to have more contributors remain important but were not studied in this thesis. These types of studies present a great opportunity for future research.

5.3: OSM for Planning

- 5.31 Building Footprints and Quality for Planning

The results from this study have shown that the quality of building footprints from OSM varies greatly by city and region. That being said, the use of OSM building footprints can be valuable for planners. Ensuring certain levels of spatial data quality, by performing tests similar to those done in this thesis, is necessary to ensure proper fitness for use for planners.

The use of building footprints can aid in a variety of planning scenarios. Like any type of dataset, there will be specific times when it is valuable for use in an analysis. One example encountered was when the Town of Caledon, Ontario received newly designated wetlands from the Ontario Ministry of Natural Resources and wanted to determine which/how many buildings were within 30 and 120 m of a wetland. Having an accurate and complete building footprint dataset was important for this purpose in order to determine potential risks to wetlands or nearby buildings. A similar need for accurate building footprints is to determine buildings impacted by a potential flood or other event. The insurance industry uses this data to determine if a house would be impacted by such an event. Previously, only parcel data was used for this purpose, however even if a parcel is affected, the house or building may not be. Thus, the use of building footprints for natural disaster management mapping is far more accurate than simply using parcel data. In other scenarios, a municipality/organisation may want to know the location and number of buildings for things such as proximity to sensitive areas, in determining catchment areas, determining demographics/population counts for an area, determining business/market areas etc.

For any type of analysis involving the location and number of building footprints in an area, assuring the fitness for use (FFU) of the data is important. For planners to ensure the FFU of data for these purposes, it would be important to ensure completeness as well as positional accuracy in the study area. This would ensure that appropriate counts of buildings are present in the correct area. This is especially important for analysis in a small area or if counting the number of buildings in a zoning category. For qualitative analysis, attributes are important to ensure the FFU. It is useful to know the type of building, name, address etc. For example, businesses would want to know if there are other businesses of the same type nearby. Planners may want to know how many schools or apartment buildings are in a certain area etc. It will ultimately be important to ensure a certain level of uniform attributes for such analysis.

As stated above, a certain level of attributes will be required to meet the FFU for certain types of analysis. As for completeness, in areas where there is no municipal building footprint layer, OSM can be useful. This was noted by Barrington-Leigh and Millard Ball (2017), especially in developing countries. Planners must know the general level of completeness to determine FFU. OSM Wiki notes that this is usually easy to determine by noticing missing buildings in certain parts of the city/town. As residential buildings are often missing (based on this study's results), planners and consultants may only be able to perform analyses related to commercial or institutional buildings as these are often more complete. The level of completeness of both buildings and attributes will play a key role in the usefulness of the data. As OSM data becomes more complete, planners and citizens can perform more types of analyses. For now, in some areas, citizens may be able to gain valuable information about the number and type of buildings in certain complete areas where open data is not available. Authors such as Mooney and Corcoran (2012), and Fan et al. (2014), have noted that completeness is very high in the UK and Germany, but less so in North America. Barrington Leigh and Millard Ball (2017) have noted that OSM completeness is greater than government data in many developing countries. As such, OSM data may be more valuable to planners in these areas.

Additional measures of spatial data quality are important to understand if the data is to be used for planning purposes. Ensuring a certain level of positional accuracy is important especially for studies where knowing the exact building location is essential, such as knowing how many buildings are contained on a large lot or calculating the distance between buildings. Having buildings that were not within parcel boundaries or that had inaccurate placement would not meet the FFU for this purpose. For certain analyses done over a large area, measures such as positional accuracy are not as important (such as numbers for an entire city). Having a dataset without topological errors is also important. If buildings overlap, errors can occur, such as buildings being counted twice. Additionally, commissions should be considered. As noted in this study, the user of the data will need to determine if they are new buildings or ones that do not belong. By performing the tests done in this thesis, planners can determine the level of spatial data quality of buildings and better assess the FFU for their purpose.

- 5.32 Opportunities for Planning using OSM Data

The vast availability of VGI and OSM data can prove extremely useful for planners. Stafford (2014) notes that GIS is essential for urban planning. As the management of land is complex, planners need specialized GIS software and skills. GIS serves as an analytical and modelling tool for planning. GIS is useful for helping with a variety of issues including location feasibility studies for large projects or even small buildings as well as environmental suitability of land. Data including the biological, physical and chemical properties of land prove useful for this analysis as well as determining the impact on surrounding habitats including wetlands. GIS is also used as a means of forecasting and monitoring. Analysis can be performed to monitor or predict change over time which can help determine the development and build-out areas for a city and allow for better allocation of resources (Stafford, 2014).

The use of GIS by planners requires spatial data. While many municipalities and governments produce spatial data, VGI and OSM present an opportunity to gather even more useful spatial data. OSM can help solve some of the issues pertaining to municipal spatial data. Firstly, open government data is static whereas OSM is constantly updated, thus this gives users more up-to-date data that tracks changes in a city such as new development. Secondly, as discussed earlier, areas in developing countries, as well as smaller municipalities worldwide, often do have the resources to create their own spatial data. OSM data has been shown to be more complete and accurate in these areas. This allows an opportunity for planners in these areas to use OSM data to conduct analysis of their local areas. Overall, VGI and OSM data present a great solution to help fill the spatial data gap for planners worldwide. Examples of VGI and OSM used for planning are noted in section 2.42.

While GIS and spatial data are necessary to properly manage a city and prepare for the future, a lack of resources in many cities in developing countries and rural areas prevents them from using expensive GIS software or creating local spatial data. Luckily, there exists a solution that can help planners in these areas perform vital GIS analysis. As previously discussed, OSM can provide useful spatial data including buildings, roads, trails and attributes all over the world. Furthermore, there exists a variety of open-source GIS freeware which planners can use for analysis. By downloading OSM shapefiles through services such as HOT OSM exporter, planners can then perform analysis in a variety of free programs such as QGIS and GeoDa. Creating awareness of OSM data and GIS freeware in developing and rural areas, as well as proving basic GIS training can help to better manage cities and towns all over the world.

One way to help improve the completeness and accuracy of OSM is to have planners contribute to it. As OSM can provide useful spatial data for planning purposes, it makes sense that planners with knowledge of the local area should contribute. This is especially useful in areas where government data does not exist or is not up-to-date. Local planners can contribute to OSM and then be able to use the updated data to perform analysis in their area. In areas where government data is available, planners should have a desire to make sure that the OSM in the area is complete, accurate and up-to-date so that local citizens and interested parties can make use of the data without worrying about quality. Furthermore, by helping to keep local area OSM data up-to-date with new buildings and land use changes, planners and the public alike will have the most complete data, which governments often take a long time to update. In general, municipal, provincial/state and national government agencies should encourage planners to contribute to and use OSM data. This is especially important in areas where government data is not available. Making planners aware of the OSM platform can help them contribute local data and be able to use it for their local planning purposes.

Chapter 6: Conclusion

6.1: Summary

Since its introduction in 2004, OpenStreetMap has developed into an expansive resource of data created through crowdsourcing as a form of volunteered geographic information. Contributors from around the world with varying levels of knowledge can contribute to this free, editable world map. While OSM has proven successful as a source of VGI, there still exists challenges with ensuring the spatial data quality of data from OSM. Evaluating spatial data quality for OSM data can be done in a variety of ways. The ISO has developed numerous measures of spatial data quality that can be applied to OSM data including positional accuracy, thematic accuracy and completeness. The research presented in this thesis had the objective to develop a replicable system to evaluate measures of spatial data quality for datasets taken from OSM. Another objective of the research was to evaluate the usefulness of OSM for planners. The third objective was to determine the level of quality of OSM datasets for ten cities in Canada and investigate factors for varying accuracy. The final objective of this research was to investigate ways to improve the completeness and accuracy of data in OSM. After conducting the study, this thesis has managed to complete these four objectives.

6.2: Meeting the Study's Objectives

The first objective was stated as: *develop a simple to use set of models to evaluate the spatial data quality of building footprints from OpenStreetMap. This approach is designed to be replicable and can be used for any area.* Throughout this study, this objective has been met. A system of models was developed to evaluate various measures of spatial data quality from OSM datasets. These measures include completeness, zonal completeness, commissions, positional accuracy, topological consistency and attribute accuracy. A model was developed to work with a reference dataset as well as without a reference dataset. These models were run on the ten study municipalities successfully and can be run on other datasets simply by changing the input data, thus proving this first objective to be a victory.

The second objective was stated as: *to understand how datasets from OpenStreetMap can be useful to planners.* Throughout this thesis, the topic of planning and OSM was touched upon in two areas. First, the literature review investigated studies of the uses of OSM data and how planners can take advantage of this data. Also, studies evaluating OSM quality around the world and its importance for planners were discussed. The discussion chapter of this thesis further investigated the usefulness of OSM for planning. Discussions on the use of building footprints for planning analysis as well as the level of spatial data quality required for such was touched upon. Furthermore, the vast opportunities for planning using OSM data were presented. These include filling in spatial data gaps, using OSM data for GIS analysis, having the most up-to-date data, having free data for use in GIS freeware in areas without financial resources and having planners contribute to OSM. Throughout this analysis, the second objective has been met.

The third objective was stated as: *to determine the level of quality of building footprints for OpenStreetMap for various cities in Canada and investigate reasons for variations.* By running the models, the overall level of spatial data quality was determined for ten Canadian cities. The spatial data quality measures include completeness, zonal completeness, commissions, positional accuracy, topological consistency and attribute accuracy. These measures were reported for the ten study municipalities and then compared. Furthermore, an analysis of completeness over time was done between January 2010 and July 2018 for all study municipalities. The number of building footprints over time and the locations in which they were created were noted, thus giving insight into the spatial data quality at different time periods. Also noted was the number of contributors in each city in July 2018. The reasons behind varying levels of quality across the study municipalities were presented in the discussion chapter. These include local knowledge and interest, the type of buildings and the perceived importance of cities and regions. By presenting the spatial data quality data for all study municipalities and investigating reasons for variations, the third objected has been completed.

6.3: Limitations

While this study performed numerous tests to evaluate the spatial data quality of building footprints from OpenStreetMap, there exist limitations with the analysis that was performed. It is worth noting that while many spatial data quality tests were performed, many additional tests could also be performed to measure additional aspects of spatial data quality. One notable exception in this study is the measure of shape accuracy. Shape accuracy refers to the accuracy of a polygon (building) in relation to its feature in real life. The number of sides, length of sides and interior angles make up factors that can be measured for determining shape accuracy. Shape accuracy was not measured in this study due to a lack of tools available to do so in modern versions of ArcGIS. Tools such as the *Shape Metrics Toolbox* and *V-LATE* were investigated for this study. The *Shape Metrics Toolbox* was designed for older versions of ArcGIS and was not compatible with the version used in this study. The *V-LATE* package was designed for landscape analysis and thus was not suitable for evaluating buildings. A trajectories package for R can be used to generalize buildings and improve shape accuracy. The R program however, is difficult for some to understand and the usefulness for measuring shape accuracy (rather than improving it) was not determined. This package could potentially be useful for further studies on building shape.

The ArcGIS data comparison toolset includes a feature compare tool. This tool was not suitable for this study as both feature classes require the same objectid (or other field) in order to be properly compared. The results only indicate if the feature types and shape types are the same. The other tools in this toolset are not suitable for use on polygons from different sources. Rather, the tools focus on measuring changes when a dataset is updated, or for comparing rasters or tables. The union tool in ArcGIS combines both datasets into one. Areas of buildings that do not overlap are created as new polygons. As such, the union dataset contains many more polygons than either the OSM or municipal ones, as these new polygons are added to the existing number of overlapping ones. One could measure the excess (non-overlapping) area. Higher numbers could indicate a lower shape accuracy but would more likely indicate a lower positional accuracy as a result of the offset of buildings between feature classes. The shape of the buildings themselves is not measured and thus this tool was excluded from the study. Further tools were explored, but they were designed for use on raster datasets, specifically for landscape analysis. While shape accuracy is important, it is worth noting that most buildings are comprised of simple rectangular shapes. If a building appears as an odd shape, then a user can verify it by checking a reference dataset (if available), air photo or by using street view/Google Maps.

Additional limitations include the number and choosing of cities for this study. While it would be beneficial to study more cities, in the interest of time and available data, only a select number could be studied. Cities in certain provinces had to be excluded, as the required datasets were not available. While this study aimed to see regional variation in terms of spatial data quality, it determined that no assumptions can be made about regional spatial data quality in Canada as each city is different, regardless of the region or province where it’s located.

The models themselves have certain limitations. They perform a select grouping of tests and provide results. It is important that the user understands what the models are doing and the results that are provided. Certain results, such as the mean near distance between centroids can be skewed by outliers or buildings without a 1:1 relation. Commissions must be evaluated further to determine if they are new buildings or errors. Roads that touch buildings must be evaluated to see if they are actual roads or rather footpaths/service roads (this was done in this study as the attribute table noted the type of road). When determining zonal completeness, buildings that are in two zones are counted in both, meaning the total number of building footprints reported is slightly higher than the actual number of building footprints in the dataset. The alternative is to use a "completely within" clause in which those buildings would not be counted at all in any zone. While these limitations exist, by identifying them and ensuring consistency among the evaluation of all cities, an accurate comparison of spatial data quality can be determined. It is worth noting that any study of spatial data quality will ultimately have some limitations.

6.4: Future Research

While this study was unique in its evaluation of spatial data quality for building footprints from OpenStreetMap, many additional avenues of research can still be pursued. First of all, this study was not able to measure shape accuracy of building footprints. It is recommended that future studies investigate this measure of spatial data quality for building footprints and other datasets taken from OSM. In order to do so, a different analysis program or plug-in would have to be used. It is also recommended that such a program or program plug-in be developed to measure shape accuracy, such as an updated version of the *Shape Metrics Toolbox for ArcGIS 9*. Developing additional programs, tools or plug-ins to measure additional facets of spatial data quality not presented in this study would also be useful. These can include other measures of positional or attribute accuracy as well as additional measures of topological consistency or temporal quality. While evaluating these additional spatial data quality measures for OSM building footprints would be useful, performing an evaluation on other VGI datasets would be beneficial. As VGI exists in multiple areas, ensuring quality is important. In addition to measuring spatial data quality, a study on improving contributions should be done. This study has made suggestions on ways to improve contributions to OSM. A study should implement these suggestions and monitor contributions. For example, a study could select an area to complete in OSM. It could then promote this project on social media, create mapping parties, present errors and work with university GIS programs to see which measures have the most impact on improving completeness and accuracy. In conclusion, the study performed in this thesis will help people to understand the importance and relevance of OSM building footprints and ensuring their spatial data quality; however additional studies can further contribute to the understanding and usefulness of OSM and VGI.

References

ASPRS Positional Accuracy Standards for Digital Geospatial Data. (2015). *Photogrammetric Engineering and Remote Sensing*, *81*(3), 1–26. https://doi.org/10.14358/PERS.81.3.A1-A26

Atzmanstorfer, K., Resl, R., Eitzinger, A., and Izurieta, X. (2014). The GeoCitizen-approach: community-based spatial planning – an Ecuadorian case study. *Cartography and Geographic Information Science*, *41*(3), 248–259. https://doi.org/10.1080/15230406.2014.890546

Ballatore, A., and Mooney, P. (2015). Conceptualising the geographic world: the dimensions of negotiation in crowdsourced cartography. *International Journal of Geographical Information Science* 29, 2310–2327. https://doi.org/10.1080/13658816.2015.1076825

Barrington-Leigh, C., and Millard-Ball, A. (2017). The world's user-generated road map is more than 80% complete. *PLOS ONE*, *12*(8), e0180698. https://doi.org/10.1371/journal.pone.0180698

Beginner's Guide 1.3. OpenStreetMapWiki. https://wiki.openstreetmap.org/wiki/Beginners_Guide_1.3. Accessed May 14, 2019

Boeing, G. (2017). OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems*, *65*, 126–139. https://doi.org/10.1016/j.compenvurbsys.2017.05.004

Brinkhoff, T. (2016). Open Street Map Data As Source For Built-Up And Urban Areas On Global Scale. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLI-B4*, 557–564.

Budhathoki, N., and Haythornthwaite, C. (2013). Motivation for Open Collaboration: Crowd and Community Models and the Case of OpenStreetMap. *American Behavioral Scientist*, *57*(5), 548–575. https://doi.org/10.1177/0002764212469364

Buildings. OpenStreetMap Wiki. https://wiki.openstreetmap.org/wiki/Buildings. Accessed May 14, 2019

Category:Features. OpenStreetMap Wiki. https://wiki.openstreetmap.org/wiki/Category:Features. Accessed May 14, 2019

Charbonneau, L., Donner, J., Filion, G., and Pierce, N. (2014). *CanTopo Map Standards and Specifications 1:50 000* (No. 2). https://doi.org/10.4095/293993

Chen, C. (2017). *Gamification in a Volunteered Geographic Information context with regard to contributors' motivations: A case study of OpenStreetMap*. (Master's Thesis). University of Waterloo.

City of Edmonton. Edmonton 311 App Download. https://www.edmonton.ca/programs_services/edmonton-311-app.aspx. Accessed May 1, 2019

City of Saint John. Central Peninsula Neighbourhood Plan. http://www.saintjohn.ca/en/home/cityhall/developmentgrowth/neighbourhood-plan.aspx. Accessed May 1, 2019

Coleman, D., Georgiadou, Y., and Labonte, J. (2009). Volunteered Geographic Information: the nature and motivation of produsers. *International Journal of Spatial Data Infrastructures Research*, *4*(4), 332–358. https://doi.org/10.2902/

Cowan, T. (2013). *A framework for investigating volunteered geographic information relevance in planning* (Master's Thesis). University of Waterloo.

Devillers, R., Bédard, Y., Jeansoulin, R., and Moulin, B. (2007). Towards spatial data quality information analysis tools for experts assessing the fitness for use of spatial data. *International Journal of Geographical Information Science*, *21*(3), 261–282. https://doi.org/10.1080/13658810600911879

Devillers, R., Stein, A., Bédard, Y., Chrisman, N., Fisher, P., and Shi, W. (2010). Thirty Years of Research on Spatial Data Quality: Achievements, Failures, and Opportunities: Thirty Years of Research on Spatial Data Quality. *Transactions in GIS*, *14*(4), 387–400. https://doi.org/10.1111/j.1467-9671.2010.01212.x

Dittus, M. (2016). Building Large-scale Crowdsourcing Communities with the Humanitarian OpenStreetMap Team. *State of the Map 2016, Brussels*

Drummond, W. J., and French, S. P. (2008). The Future of GIS in Planning: Converging Technologies and Diverging Interests. *Journal of the American Planning Association*, *74*(2), 161–174. https://doi.org/10.1080/01944360801982146

Fan, H., Zipf, A., Fu, Q., and Neis, P. (2014). Quality assessment for building footprints data on OpenStreetMap. *International Journal of Geographical Information Science*, *28*(4), 700–719. https://doi.org/10.1080/13658816.2013.867495

Fast, V., and Rinner, C. (2014). A Systems Perspective on Volunteered Geographic Information. *ISPRS International Journal of Geo-Information*, *3*(4), 1278–1292. https://doi.org/10.3390/ijgi3041278

Feick, R., and Roche, S. (2013). Understanding the Value of VGI. In D. Sui, S. Elwood, and M. Goodchild (Eds.), *Crowdsourcing Geographic Knowledge* (pp. 15–29). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-4587-2_2

Filippovskaa, Y., and Kadab, M. (2010). Shape assessment of generalized building footprints. In *Proceedings of ASPRS, CaGIS, ISPRS, and AutoCarto Fall Conference, Orlando, Florida, USA*.

Fonte, C. C., Patriarca, J. A., Minghini, M., Antoniou, V., See, L., and Brovelli, M. A. (2017). Using OpenStreetMap to Create Land Use and Land Cover Maps: Development of an Application. *Volunteered Geographic Information and the Future of Geospatial Data*, 113–137. https://doi.org/10.4018/978-1-5225-2446-5.ch007

Foody, G. M., See, L., Fritz, S., van der Velde, M., Perger, C., Schill, C., … Comber, A. (2015). Accurate Attribute Mapping from Volunteered Geographic Information: Issues of Volunteer Quantity and Quality. *The Cartographic Journal*, *52*(4), 336–344. https://doi.org/10.1179/1743277413Y.0000000070

Gallagher, J., Orcutt, J., Simpson, P., Wright, D., Pearlman, J., and Raymond, L. (2015). Facilitating open exchange of data and information. *Earth Science Informatics*, *8*(4), 721–739. https://doi.org/10.1007/s12145-014-0202-2

Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal*, *69*(4), 211–221. https://doi.org/10.1007/s10708-007-9111-y

Goodchild, M. F., and Li, L. (2012). Assuring the quality of volunteered geographic information. *Spatial Statistics*, *1*, 110–120. https://doi.org/10.1016/j.spasta.2012.03.002

Grira, J., Bédard, Y., and Roche, S. (2009). Spatial Data Uncertainty in the VGI World: Going from Consumer to Producer. *Geomatica*, 64(1), 61-71. Retrieved from http://yvanbedard.scg.ulaval.ca/wp-content/documents/publications/547.pdf

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. https://doi.org/10.1068/b35097

Haklay, M., Antoniou, V., Basiouka, S., Soden, R., and Mooney, P. (2014). Crowdsourced geographic information use in government. *Report to GFDRR (World Bank).* London

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. https://doi.org/10.1179/000870410X12911304958827

Hamari, J., and Koivisto, J. (2015). "Working out for likes": An empirical study on social influence in exercise gamification. *Computers in Human Behavior*, *50*(Complete), 333–347. https://doi.org/10.1016/j.chb.2015.04.018

Hecht, R., Kunze, C., and Hahmann, S. (2013). Measuring Completeness of Building Footprints in OpenStreetMap over Space and Time. *ISPRS International Journal of Geo-Information*, *2*(4), 1066–1091. https://doi.org/10.3390/ijgi2041066

Horita, F., Degrossi, L., Assis, L. F., Zipf, A., and De Albuquerque, J. (2013). The use of Volunteered Geographic Information and Crowdsourcing in Disaster Management: a Systematic Literature Review. *Proceedings of the Nineteenth Americas Conference on Information Systems, Chicago, Illinois, August 15-17, 2013.*

HOT Tasking Manager. tasks.hotosm.org. Accessed May 1, 2019

How We Map. OpenStreetMap Wiki. https://wiki.openstreetmap.org/wiki/How_We_Map. Accessed May 14, 2019

ISO-TC211 N3521 Text for FDIS 19157 Data quality. (2013). *International Organization for Standardization*

Kalantari, M., Rajabifard, A., Olfat, H., and Williamson, I. (2014). Geospatial Metadata 2.0 – An approach for Volunteered Geographic Information. *Computers, Environment and Urban Systems*, *48*, 35–48. https://doi.org/10.1016/j.compenvurbsys.2014.06.005

Kelly, P. (2007). Role of Spatial Data Infrastructures in Managing Our Cities, *Keynote address at the FIG Commission 3 2007 Annual Meeting and Workshop*, Athens, Greece.

KPMG. (2016). *2015 assessment of the Canadian Geospatial Data Infrastructure* (No. 49e). https://doi.org/10.4095/297880

Kuznetsov, S. (2006). Motivations of Contributors to Wikipedia. *SIGCAS Comput. Soc.*, *36*(2). https://doi.org/10.1145/1215942.1215943

Lokhat, I., and Touya, G. (2016). Enhancing building footprints with squaring operations. *Journal of Spatial Information Science*, (13). https://doi.org/10.5311/JOSIS.2016.13.276

Min, D., Zhilin, L., and Xiaoyong, C. (2007). Extended Hausdorff distance for spatial objects in GIS. *International Journal of Geographical Information Science*, *21*(4), 459–475. https://doi.org/10.1080/13658810601073315

Missing Maps. https://www.missingmaps.org/. Accessed May 1, 2019

Mooney, P., and Corcoran, P. (2012). Characteristics of Heavily Edited Objects in OpenStreetMap. *Future Internet*, *4*(4), 285–305. https://doi.org/10.3390/fi4010285

Mooney P., and Corcoran P. (2013). Analysis of interaction and co-editing patterns amongst OpenStreetMap contributors. *Trans GIS*

Mooney, P. and Minghini, M. (2017). A Review of OpenStreetMap Data. In V. Antoniou and Geographic Directorate, PAPAGOU Camp, GR, *Mapping and the Citizen Sensor* (pp. 37–59). Ubiquity Press. https://doi.org/10.5334/bbf.c

Natural Resources Canada, and Public Safety Canada. (2017). *Federal airborne LiDAR data acquisition guideline* (No. 117e). https://doi.org/10.4095/304669

Natural Resources Canada Surveyor General Branch. (2014). *National Standards for the Survey of Canada Lands*. (Version 1.0)

Nikšič, M, Campagna, M, Massa, P, Caglioni, M, Nielsen, T T. (2017). Opportunities for Volunteered Geographic Information Use in Spatial Planning. In: Foody, G, See, L, Fritz, S, Mooney, P, Olteanu-Raimond, A-M, Fonte, C C and Antoniou, V. (eds.) *Mapping and the Citizen Sensor*. Pp. 327–349. London: Ubiquity Press. DOI: https://doi.org/10.5334/bbf.n. License: CC-BY 4.0

OpenStreetMap. https://www.openstreetmap.org. Accessed April 17, 2019.

OpenStreetMap. OpenStreetMap Changesets. https://www.openstreetmap.org/history. Accessed April 17, 2019.

OpenStreetMap Wiki. https://wiki.openstreetmap.org/wiki/Main_Page. Accessed April 17, 2019.

Osei, A., Merem, E.C., and Twumasi, Y.A. (2006). The applications of geospatial information technology in land management: a case study of Lagos, Nigeria. In *Proceedings of the 9th International Conference of the Global Spatial Data Infrastructure Association (GSDI-9). Spatial Information: Tool for Reducing Poverty*. 6-10 November. Santiago, Chile.

OSM Tasking Manager. http://tasks.osmcanada.ca/. Accessed April 17, 2019.

Pászto, V., Brychtová, A., and Marek, L. (2015). On Shape Metrics in Cartographic Generalization: A Case Study of the Building Footprint Geometry. In J. Brus, A. Vondrakova, and V. Vozenilek (Eds.), *Modern Trends in Cartography* (pp. 397–407). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-07926-4_30

Pászto, V., Marek, L., and Tucek, P. (2013). Are shape metrics useful for a geocomputation? CORINE land-cover analysis case study. *CEUR Workshop Proceedings*, *971*, 26–35.

Roche, S. (2011). De la cartographie participative aux WikiSIG. In O. Walser, L. Thévoz, F. Joerin, M. Schuler, S. Joost, B. Debarbieux, and H. Dao (Eds.), *Les SIG au service du développement territorial* (pp. 117–129). Lausanne: Presses polytechniques et universitaires romandes.

Stafford, F. (2014). GIS in Urban Planning. *GIS Lounge*. https://www.gislounge.com/gis-urban-planning/

Statistics Canada. (2018). Census Profile, 2016 Census. https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/index.cfm?Lang=E

Sui, D., Goodchild, M., and Elwood, S. (2013). Volunteered Geographic Information, the Exaflood, and the Growing Digital Divide. In D. Sui, S. Elwood, and M. Goodchild (Eds.), *Crowdsourcing Geographic Knowledge* (pp. 1–12). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-4587-2_1

taginfo. https://taginfo.openstreetmap.org/. Accessed April 17, 2019

Teymurian, F., Alesheikh, A., Alimohammadi, A., and Sadeghi-Niaraki, A. (2013). VGI based urban public transport. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XL-1/W3*, 425–430. https://doi.org/10.5194/isprsarchives-XL-1-W3-425-2013

Touya, G., Antoniou, V., Olteanu-Raimond, A.-M., & Van Damme, M.-D. (2017). Assessing Crowdsourced POI Quality: Combining Methods Based on Reference Data, History, and Spatial Relations. *ISPRS International Journal of Geo-Information*, *6*(3), 80. https://doi.org/10.3390/ijgi6030080

WikiProject Canada/Building Canada 2020 - OpenStreetMap Wiki. https://wiki.openstreetmap.org/wiki/WikiProject_Canada/Building_Canada_2020. Accessed April 17, 2019

Appendix 1: Mid-Sized Cities in Canada

Name	CSD Type	Province	Population 2016	Area (km2)
Strathcona County	Specialized municipality (SM)	Alberta	98044	1182.78
Brantford	**City (CY)**	**Ontario**	**97496**	**72.44**
Saint-Jean-sur-Richelieu	Ville (V)	Quebec	95114	226.63
Cape Breton	Regional municipality (RGM)	Nova Scotia	94285	2430.06
Lethbridge	**City (CY)**	**Alberta**	**92729**	**122.09**
Clarington	Municipality (MU)	Ontario	92013	611.4
Pickering	City (CY)	Ontario	91771	231.55
Nanaimo	City (CY)	British Columbia	90504	90.76
Kamloops	**City (CY)**	**British Columbia**	**90280**	**299.25**
Niagara Falls	**City (CY)**	**Ontario**	**88071**	**209.73**
North Vancouver	District municipality (DM)	British Columbia	85935	160.76
Victoria	City (CY)	British Columbia	85792	19.47
Brossard	Ville (V)	Quebec	85721	45.23
Repentigny	Ville (V)	Quebec	84285	61.23
Newmarket	Town (T)	Ontario	84224	38.45
Chilliwack	**City (CY)**	**British Columbia**	**83788**	**261.65**
Maple Ridge	City (CY)	British Columbia	82256	266.78
Peterborough	City (CY)	Ontario	81032	64.25
Drummondville	Ville (V)	Quebec	75423	247.15
Kawartha Lakes	City (CY)	Ontario	75423	3084.38
Saint-Jérôme	Ville (V)	Quebec	74346	90.44
Prince George	**City (CY)**	**British Columbia**	**74003**	**318.26**
Sault Ste. Marie	City (CY)	Ontario	73368	223.24
Moncton	City (C)	New Brunswick	71889	141.92
Sarnia	City (CY)	Ontario	71594	164.85
Wood Buffalo	Specialized municipality (SM)	Alberta	71589	61777.65
New Westminster	City (CY)	British Columbia	70996	15.63
Saint John	**City (C)**	**New Brunswick**	**67575**	**315.96**
Caledon	Town (T)	Ontario	66502	688.16
Granby	Ville (V)	Quebec	66222	152.79
St. Albert	City (CY)	Alberta	65589	48.45
Norfolk County	City (CY)	Ontario	64044	1607.55
Medicine Hat	City (CY)	Alberta	63260	112.04
Grande Prairie	**City (CY)**	**Alberta**	**63166**	**132.73**
Airdrie	City (CY)	Alberta	61581	84.57

Halton Hills	**Town (T)**	**Ontario**	**61161**	**276.27**
Port Coquitlam	City (CY)	British Columbia	58612	29.17
Fredericton	City (C)	New Brunswick	58220	132.57
Blainville	Ville (V)	Quebec	56863	55.16
Saint-Hyacinthe	Ville (V)	Quebec	55648	188.97
Aurora	Town (T)	Ontario	55445	49.85
North Vancouver	City (CY)	British Columbia	52898	11.85
Welland	City (CY)	Ontario	52293	81.04
North Bay	City (CY)	Ontario	51553	319.11
Belleville	City (CY)	Ontario	50716	247.25
Mirabel	Ville (V)	Quebec	50513	485.07
Shawinigan	Ville (V)	Quebec	49349	734.84
Dollard-Des Ormeaux	Ville (V)	Quebec	48899	14.97
Brandon	City (CY)	Manitoba	48859	77.41
Rimouski	Ville (V)	Quebec	48664	339.64
Châteauguay	Ville (V)	Quebec	47906	35.95
Mascouche	Ville (V)	Quebec	46692	107
Cornwall	City (CY)	Ontario	46589	61.56
Victoriaville	Ville (V)	Quebec	46130	84.23
Whitchurch-Stouffville	Town (T)	Ontario	45837	206.22
Haldimand County	City (CY)	Ontario	45608	1251.54
Georgina	Town (T)	Ontario	45418	287.75
Saint-Eustache	Ville (V)	Quebec	44008	70.51
Quinte West	City (CY)	Ontario	43577	494.02
West Vancouver	District municipality (DM)	British Columbia	42473	87.26
Rouyn-Noranda	Ville (V)	Quebec	42334	6009.86
Timmins	City (CY)	Ontario	41788	2978.83
Boucherville	Ville (V)	Quebec	41671	70.5
Woodstock	City (CY)	Ontario	40902	48.97
Salaberry-de-Valleyfield	Ville (V)	Quebec	40745	107.13
Vernon	City (CY)	British Columbia	40116	96.05
Rocky View County	Municipal district (MD)	Alberta	39407	3836.33
St. Thomas	City (CY)	Ontario	38909	35.63
Mission	District municipality (DM)	British Columbia	38833	227.65
Vaudreuil-Dorion	Ville (V)	Quebec	38117	72.73
Brant	City (CY)	Ontario	36707	843.25
Lakeshore	Town (T)	Ontario	36611	530.33
Innisfil	Town (T)	Ontario	36566	262.71
Charlottetown	City (CY)	Prince Edward Island	36094	44.34

Prince Albert	City (CY)	Saskatchewan	35926	67.29
Langford	City (CY)	British Columbia	35342	39.94
Bradford West Gwillimbury	Town (T)	Ontario	35325	201.04
Sorel-Tracy	Ville (V)	Quebec	34755	57.46
New Tecumseth	Town (T)	Ontario	34242	274.21
Spruce Grove	City (CY)	Alberta	34066	32.2
Moose Jaw	City (CY)	Saskatchewan	33890	50.68
Penticton	City (CY)	British Columbia	33761	42.1
Port Moody	City (CY)	British Columbia	33551	25.89
West Kelowna	District municipality (DM)	British Columbia	32655	123.53
Campbell River	City (CY)	British Columbia	32588	144.36
Saint-Georges	Ville (V)	Quebec	32513	199.27
Val-d'Or	Ville (V)	Quebec	32491	3550.7
Côte-Saint-Luc	Ville (V)	Quebec	32448	6.96
Parkland County	Municipal district (MD)	Alberta	32097	2390.23
Stratford	**City (CY)**	**Ontario**	**31465**	**28.28**
Pointe-Claire	Ville (V)	Quebec	31380	18.9
Orillia	City (CY)	Ontario	31166	28.58
Alma	Ville (V)	Quebec	30776	196.54
Fort Erie	Town (T)	Ontario	30710	166.27
LaSalle	Town (T)	Ontario	30180	65.35
			5273968	

Appendix 2: Zonal Completeness Tables by Municipality

For the following tables, all numbers have been rounded to the nearest digit. For Halton Hills and Grande Prairie, a “not specified” category was removed with an insignificant land area. For Stratford, the “not specified” category includes 3 large parcels that consist of fields and trees. It was not removed due to the total land area.

Halton Hills

Zoning	Municipal BFs	OSM BFs	% Completeness	Area of Mun. BFs (m2)	Area of OSM BFs (m2)	% Completeness
Agricultural	1045	43	4	351103	36555	10
Commercial	541	203	38	185326	141209	76
Community Node	68	63	93	89655	89889	100
Conservation	2	0	0	418	0	0
Country Res	203	0	0	60322	0	0
Development	101	21	21	47242	31601	67
Employment	157	117	75	338123	310590	92
EPA	101	8	8	20365	4913	24
Floodline	40	7	18	8186	3551	43
Floodplain	18	2	11	3577	718	20
Gateway	26	5	19	56148	5911	11
Hamlet Res	825	3	0	159888	713	0
High Density Res	22	17	77	23273	18996	82
Industrial	74	1	1	146510	143	0
Institutional	89	63	71	178450	168248	94
Low Density Res	12989	90	1	1926365	22251	1
Medium Density Res	1528	38	2	218343	34494	16
Mineral	1	0	0	119	0	0
NEC Area	885	42	5	230492	32425	14
OMB	1	0	0	195	0	0
Open Space	52	24	46	44090	34571	78
Prot Countryside	1271	39	3	304437	27059	9
Res/Com	12	0	0	1602	0	0
Rural Res	520	4	1	87336	558	1
Special Study Area	0	0	#DIV/0!	0	0	#DIV/0!
Transportation	1	4	400	301	7861	2613
ZBA	3	2	67	6123	6305	103

Niagara Falls

Zoning	# Of Mun. BFs	# Of OSM BFs	% Completeness	Area of Mun. BFs	Area of OSM BFs	% Completeness
Environmental conservation area	442	90	20	208103	152070	73
Environmental protection area	109	11	10	35273	15216	43
Extractive industrial	3	1	33	830	326	39
Good general agriculture	2002	37	2	401668	27288	7
Industrial	845	183	22	672337	381777	57
Major commercial	512	175	34	315055	277640	88
Minor commercial	532	169	32	146756	90296	62
Niagara escarpment plan area	230	4	2	41424	4178	10
Open space	203	34	17	139728	86903	62
Parkway residential	152	0	0	36382	0	0
Refer to schedule a-3	56	12	21	24564	31411	128
Residential	33509	945	3	4549767	464777	10
Resort commercial	27	4	15	15773	16794	106
Theme park marineland	49	5	10	20336	5187	26
Tourist commercial	1432	549	38	661748	507927	77
Total	**40103**	**2219**	**6**	**7269743**	**2061790**	**28**

Kamloops

Zoning	# Of Mun. BFs	# Of OSM BFs	% Completeness	Area of Mun. BFs (m2)	Area of OSM BFs (m2)	% Completeness
Agricultural	931	332	35.66	161948.22	69132.34	42.69
Airport	86	47	54.65	26084.84	22523.78	86.35
CBD	148	144	97.30	155802.88	155059.90	99.52
Churches	65	56	86.15	36938.19	36136.61	97.83
Commercial	895	890	99.44	650102.23	646346.00	99.42
Comprehensive Res	633	439	69.35	105202.96	84936.49	80.74
Country Residential	997	820	82.25	127827.94	98065.01	76.72
Development	414	393	94.93	150111.76	143518.74	95.61
Funeral Homes	1	1	100.00	752.37	752.47	100.01
Future Development	94	21	22.34	32855.22	5430.87	16.53
Hotel	0	0	#DIV/0!	0.00	0.00	#DIV/0!
Industrial	769	700	91.03	521607.10	487162.13	93.40
Mobile Homes	2811	2743	97.58	229294.74	221224.97	96.48
Multi Family Low Dens	1205	1052	87.30	357198.91	355000.26	99.38
Multi Family Med Dens	828	824	99.52	267829.51	269701.05	100.70
Multiple Family High Dens	80	83	103.75	26076.18	27381.34	105.01
Open Space	5	18	360.00	986.57	4441.13	450.16
Parks	263	266	101.14	78627.60	85527.79	108.78
Private Recreational	0	0	#DIV/0!	0.00	0.00	#DIV/0!
Pub	14	14	100.00	6054.66	5837.10	96.41
Public	0	1	#DIV/0!	0.00	14569.40	#DIV/0!
Public Use	260	221	85.00	178151.31	177772.44	99.79
Railway	46	35	76.09	19325.84	17023.58	88.09
Resource Extraction	14	4	28.57	2841.09	477.68	16.81
Schools	92	103	111.96	173905.38	174626.20	100.41
Single Res	17133	13558	79.13	2962894.53	2239170.02	75.57
Two Family Res	8253	5595	67.79	1142161.95	816037.29	71.45
Total	**36037**	**28360**	**78.70**	**7414581.98**	**6157854.59**	**83.05**

Prince George

Land Use	# of Mun. BFs	# of OSM BFs	% Completeness	Area of Mun. BFs	Area of OSM BFs	% Completeness
Business and Industrial	1576	467	30	1004820	661602	66
Commercial	713	413	58	510308	423003	83
Recreation and Institutional	481	154	32	424504	355212	84
Residential	21549	303	1	3693795	87641	2
Rural	3381	11	0	639096	18392	3
Site Specific	230	65	28	196768	168142	85
Utility	123	22	18	42763	27202	64
Total	**28053**	**1435**	**5**	**6512055**	**1741194**	**27**

Chilliwack

Zoning	# of Mun. BFs	# of OSM BFs	% Completeness	Area of Mun. BFs (m2)	Area of OSM BFs (m2)	% Completeness
Agricultural	11112	157	1	3621467	202203	6
Airport	34	16	47	22618	13212	58
Commercial	825	355	43	560174	337341	60
Comprehensive Development	2276	61	3	545711	107387	20
Ecovillage	19	0	0	4256	0	0
Industrial	709	246	35	444135	264647	60
Mobile Home Park	528	0	0	48271	0	0
Multi-Family Residential	2261	41	2	816215	45826	6
Outdoor Recreation	209	4	2	55987	1827	3
Public	400	60	15	335340	218659	65
Reserve	3229	54	2	600184	78214	13
Residential	20155	37	0	2858335	14415	1
Rural	1841	41	2	272147	9847	4
University Village	55	9	16	59975	29806	50
Total	**43653**	**1081**	**2**	**10244815**	**1323385**	**13**

Saint John

Zoning	# of Mun. BFs	# of OSM BFs	% Completeness	Area of Mun. BFs (m2)	Area of OSM BFs (m2)	% Completeness
Commercial	1413	1417	100	992614	993071	100
Community Facility	457	320	70	387426	349604	90
EPA	25	3	12	1337	363	27
Future Dev	77	32	42	6165	8331	135
Industrial	861	479	56	738136	686346	93
Integrated Development	2	3	150	3682	2453	67
Low Density Residential	10883	7513	69	1105204	941860	85
Multi-Unit Residential	14045	9035	64	1372433	1191473	87
Park	257	113	44	34382	32340	94
Pit and Quarry	22	2	9	1681	1239	74
Rural	8825	2339	27	689180	262441	38
Special Zone	3	1	33	181	188	104
Transportation	110	59	54	151396	141377	93
Utility Service	113	52	46	44829	36821	82
Total	**37093**	**21368**	**58**	**5528647**	**4647906**	**84**

Lethbridge

Zoning	# of Mun. BFs	# of OSM BFs	% Completeness	Area of Mun. BFs (m2)	Area of OSM BFs (m2)	% Completeness
Commercial	909	317	35	856689	437684	51
Direct Control District	1210	48	4	385913	145357	38
Future Urban Development District	678	10	1	59160	3499	6
High Density Residential District	178	13	7	60445	9907	16
Industrial	1384	1	0	769643	873	0
Industrial Business District	266	7	3	193729	9718	5
Low Density Residential	47717	269	1	5060050	51820	1
Medium Density Residential	3264	84	3	578784	36098	6
Mixed Density Residential District	393	0	0	53511	0	0
Mobile Home District	1190	0	0	117461	0	0
Parks and Recreation District	231	11	5	32556	13457	41
Public Building District	607	156	26	532300	352605	66
Public Transportation District	77	4	5	7375	7091	96
Specialist Office District	8	0	0	781	0	0
Urban Innovation District	10	0	0	4822	0	0
Valley District	386	22	6	98437	45365	46
Total	**58508**	**942**	**2**	**8811658**	**1113473**	**13**

Grande Prairie

Zoning	# of Mun. BFs	# of OSM BFs	% Completeness	Area of Mun. BFs (m2)	Area of OSM BFs (m2)	% Completeness
AP Airport District	64	61	95	23897	25045	105
Commercial	874	558	64	708659	604044	85
Direct Control	74	53	72	57281	60625	106
Industrial	769	592	77	541071	516605	95
MHC Man Home Community	1244	339	27	84789	20722	24
MHS Man Home Subdivision	608	0	0	37988	0	0
MP Muskoseepi Park	84	91	108	9820	10227	104
PS Public Service	221	159	72	282516	253245	90
RC Combined Density Residential	328	224	68	41048	31299	76
RG General Residential	11796	6111	52	1209824	613652	51
RH High Density Residential	1	1	100	1443	1422	99
RM Medium Density Res	497	248	50	196020	105801	54
RR Restricted Residential	3665	1586	43	405685	185003	46
RS Small Lot Residential	8154	2987	37	871863	342286	39
RSA Rural Service Area	847	15	2	134105	4987	4
RSR Restricted Small Lot Residential	4	0	0	237	0	0
RT Residential Transition	1201	76	6	101190	11619	11
UR Urban Reserve	140	53	38	14974	5311	35
Total	**30571**	**13154**	**43**	**4722410**	**2791893**	**59**

Stratford

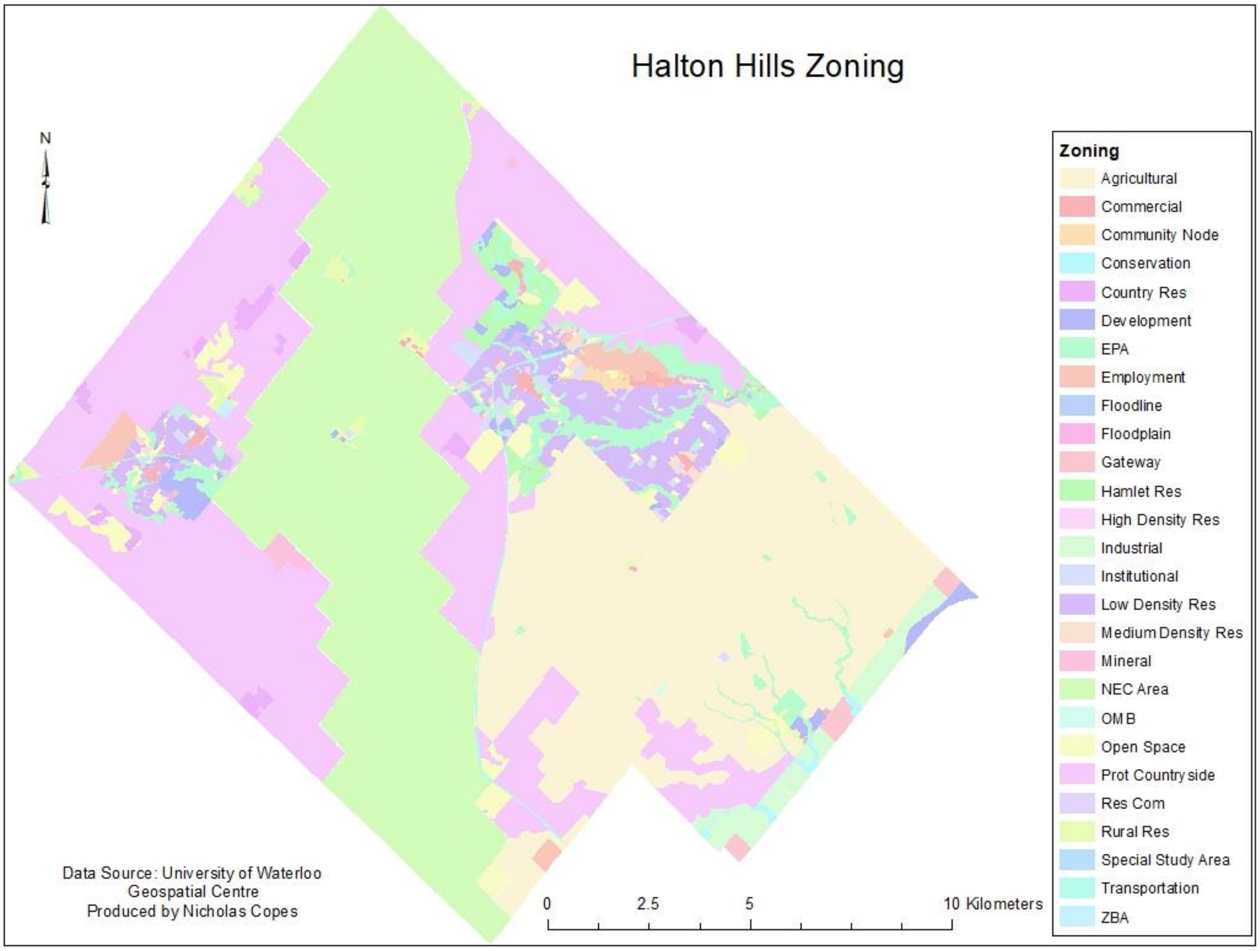
Zoning	# of Mun. BFs	# of OSM BFs	% Completeness	Area of Mun. BFs (m2)	Area of OSM BFs (m2)	% Completeness
Not Specified	59	15	25	4951	9612	194
Agricultural	139	79	57	17884	17320	97
Commercial	456	304	67	123804	257584	208
Future Residential	56	32	57	8087	8075	100
Institutional	468	391	84	247353	839481	339
Mixed Use Residential	195	86	44	20630	16116	78
Park	66	43	65	11713	46165	394
Residential	11340	5388	48	1129079	1061169	94
Total	**12779**	**6338**	**50**	**1563502**	**2255523**	**144**

Brantford

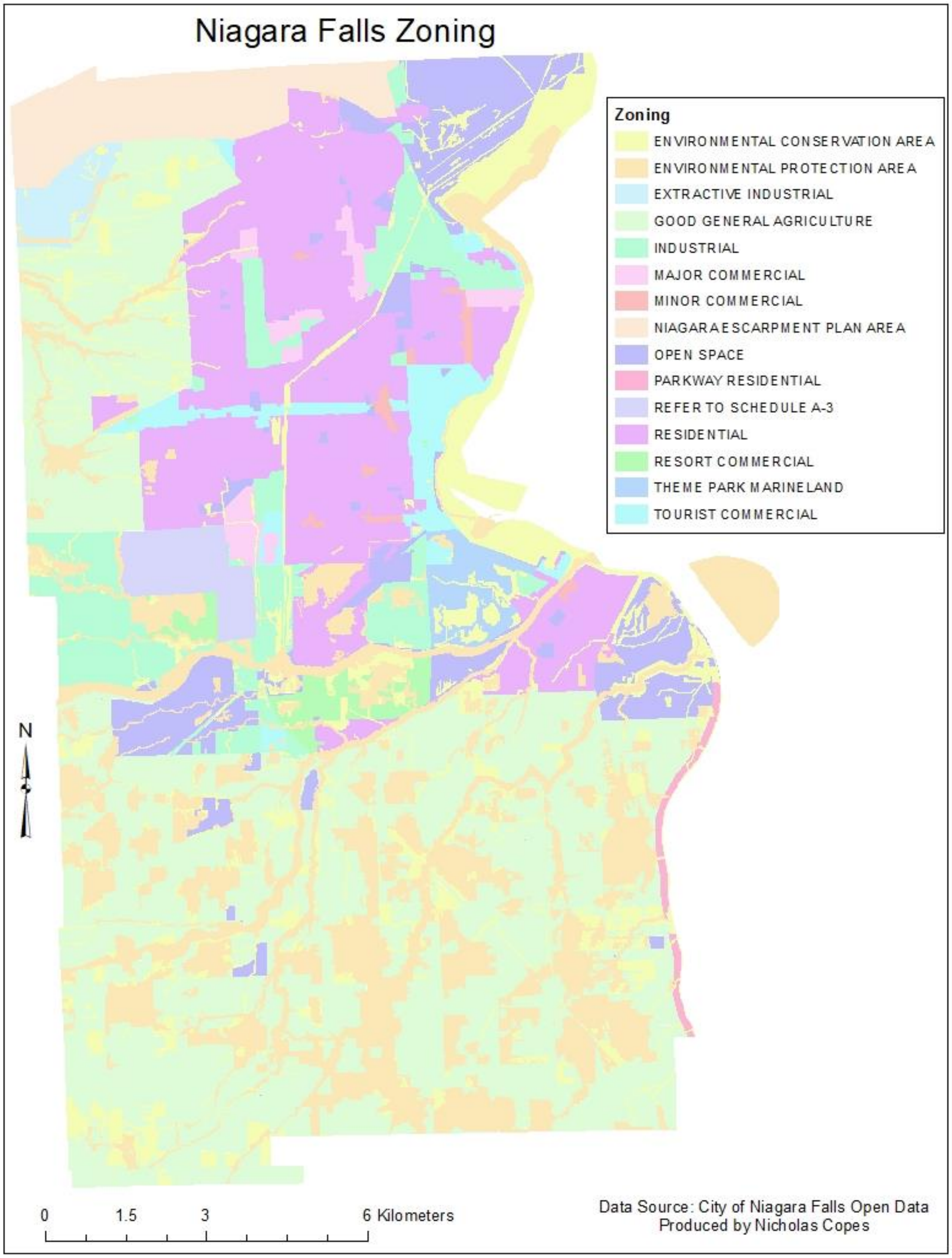
Zoning	# of Mun. BFs	# of OSM BFs	% Completeness	Area of Mun. BFs (m2)	Area of OSM BFs (m2)	% Completeness
Agricultural	414	28	7	78855	14836	19
Commercial	1032	555	54	729240	621548	85
Development Constraint Zone	13	0	0	1216	0	0
Industrial	771	469	61	1923952	1827154	95
Institutional	276	179	65	316051	335463	106
Mixed Use	422	54	13	81156	33328	41
Natural Heritage Zone (County)	41	2	5	6370	489	8
Open Space	219	66	30	103982	103042	99
Planned Unit Development Type One Zone	1	0	0	197	0	0
Recreational Facilities Zone (County)	14	0	0	1539	0	0
Residential	31951	879	3	4249136	221970	5
Residential High Density Zone	137	79	58	109832	111627	102
Residential Medium Density	1181	360	30	445824	205615	46
Six Nations of the Grand River Territory	10	5	50	6436	6388	99
Temporary Zone (County)	4	0	0	352	0	0
Total	**36486**	**2676**	**7**	**8054138**	**3481459**	**43**

Appendix 3: Zoning Maps

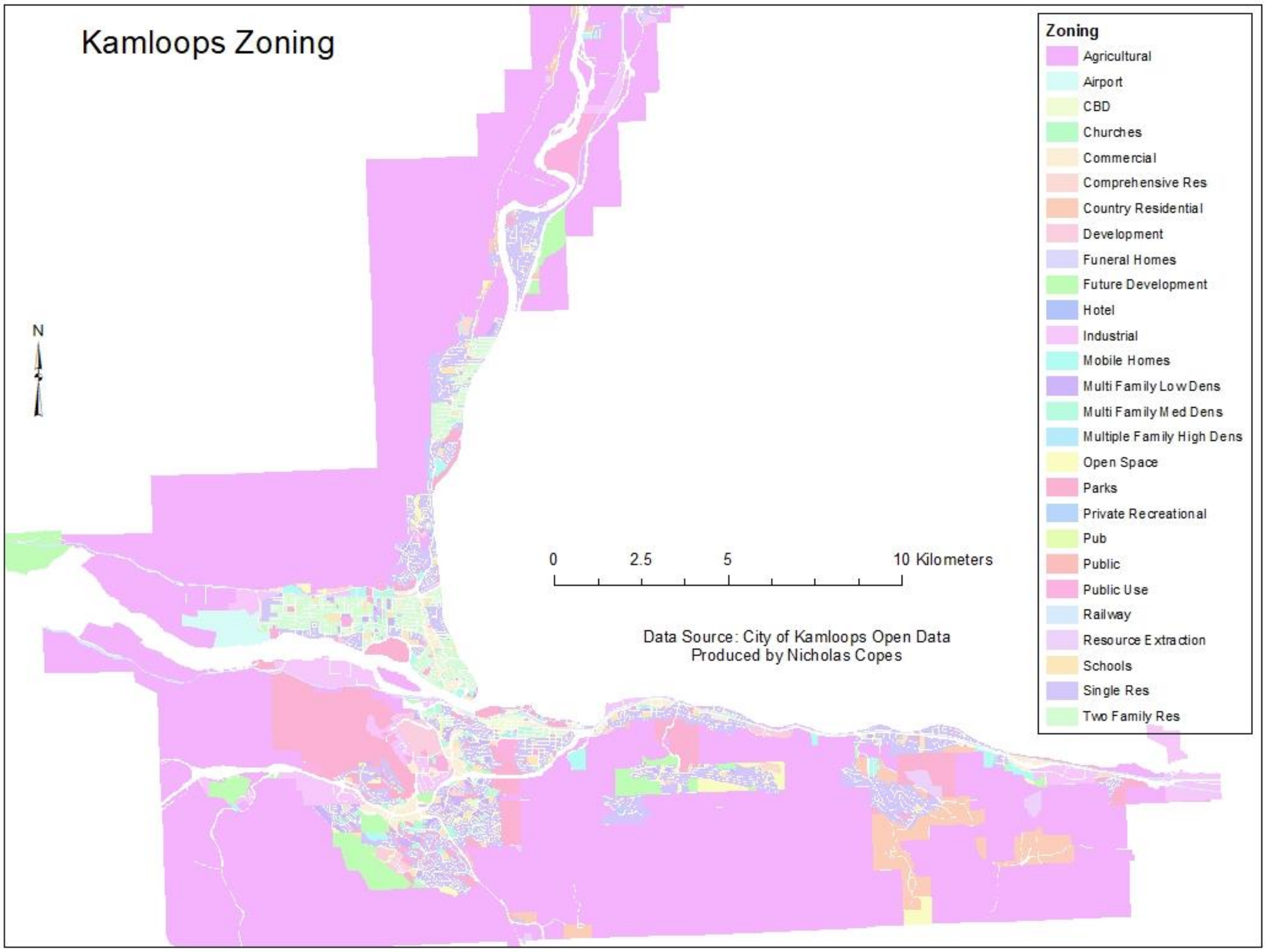
Halton Hills



Niagara Falls

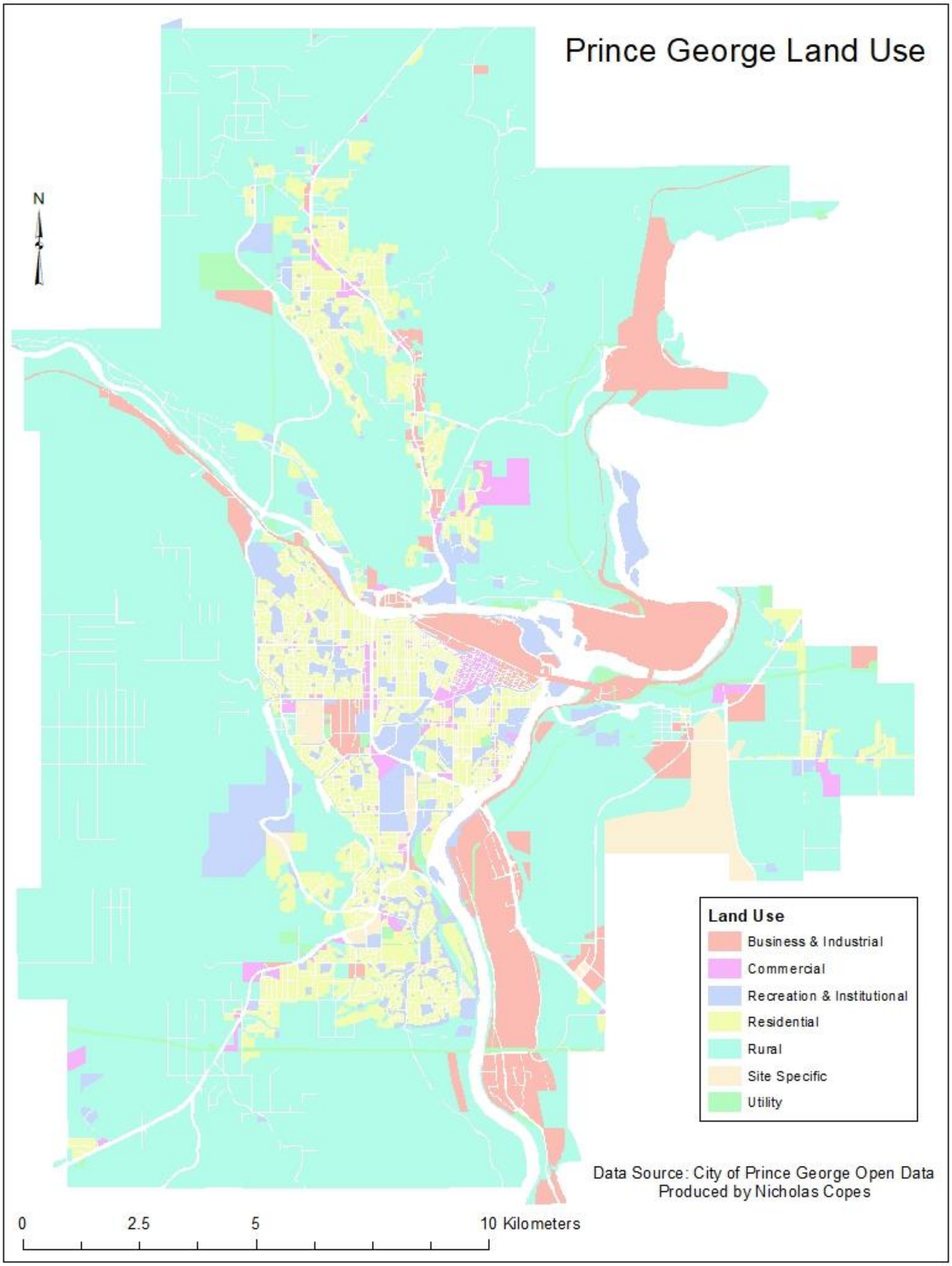


Kamloops

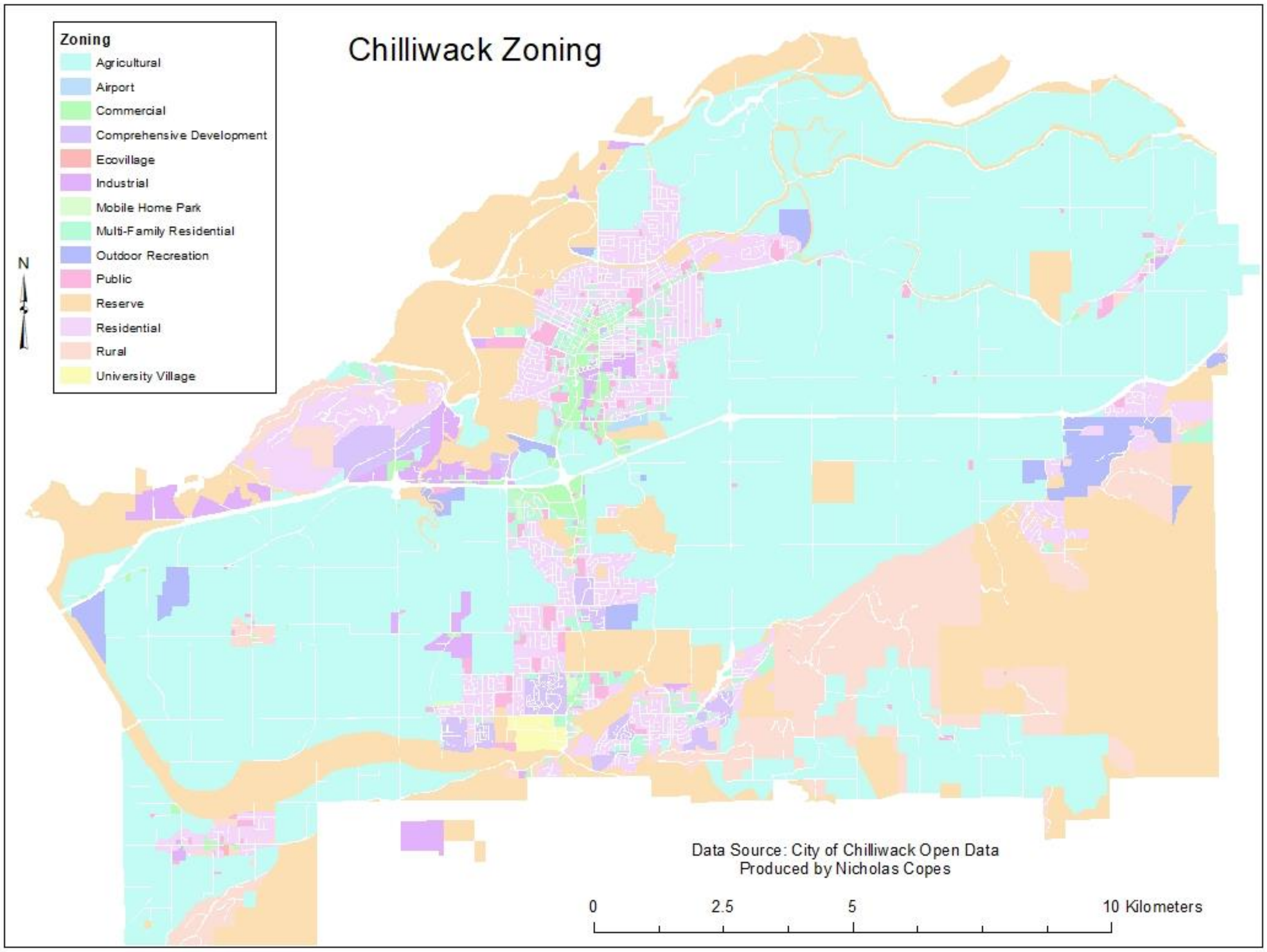


Kamloops Zoning
N
0
2.5
5
10 Kilometers
Data Source: City of Kamloops Open Data
Produced by Nicholas Copes
Zoning
Agricultural
Airport
CBD
Churches
Commercial
Comprehensive Res
Country Residential
Development
Funeral Homes
Future Development
Hotel
Industrial
Mobile Homes
Multi Family Low Dens
Multi Family Med Dens
Multiple Family High Dens
Open Space
Parks
Private Recreational
Pub
Public
Public Use
Railway
Resource Extraction
Schools
Single Res
Two Family Res

Prince George

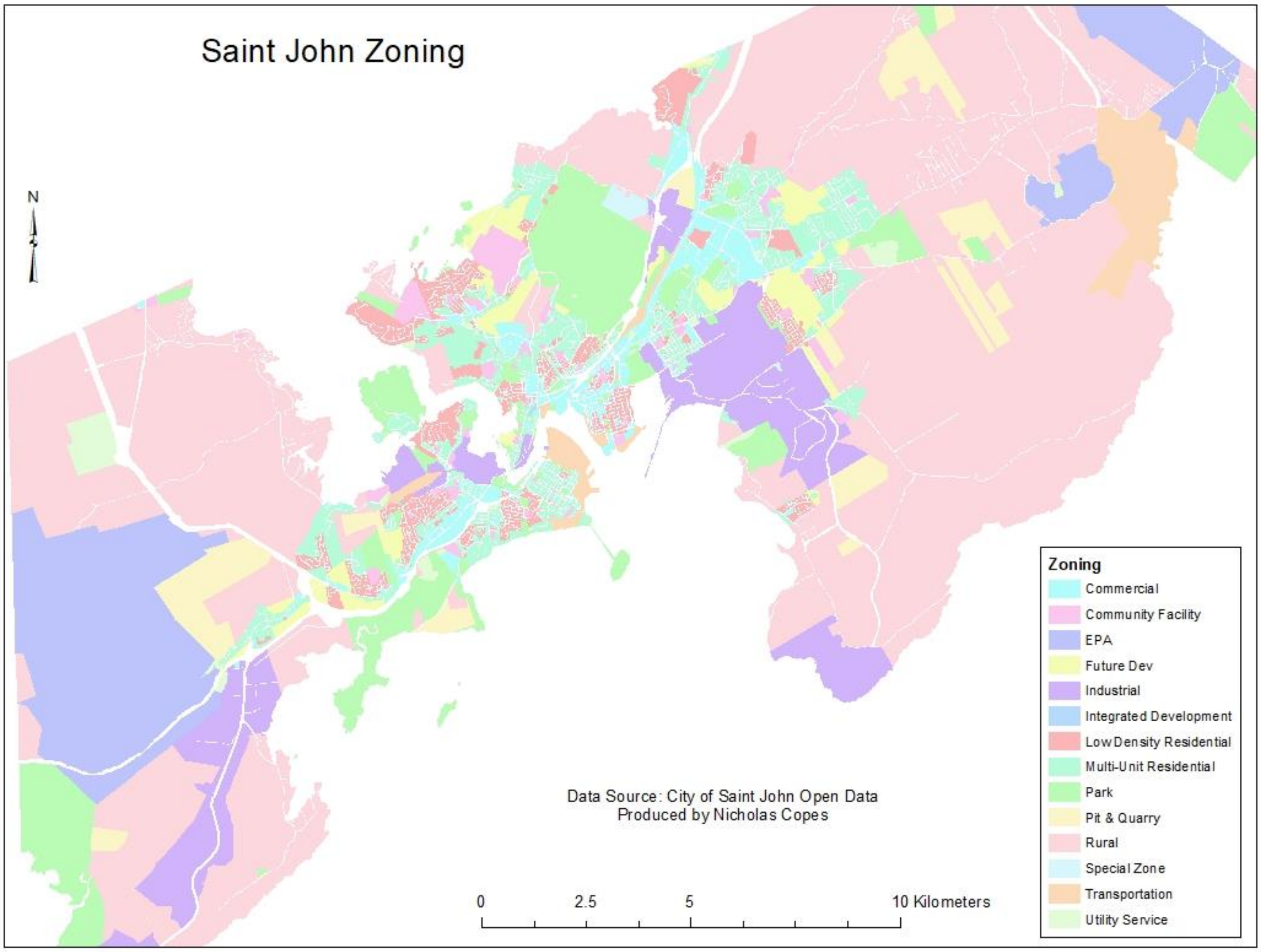


Chilliwack

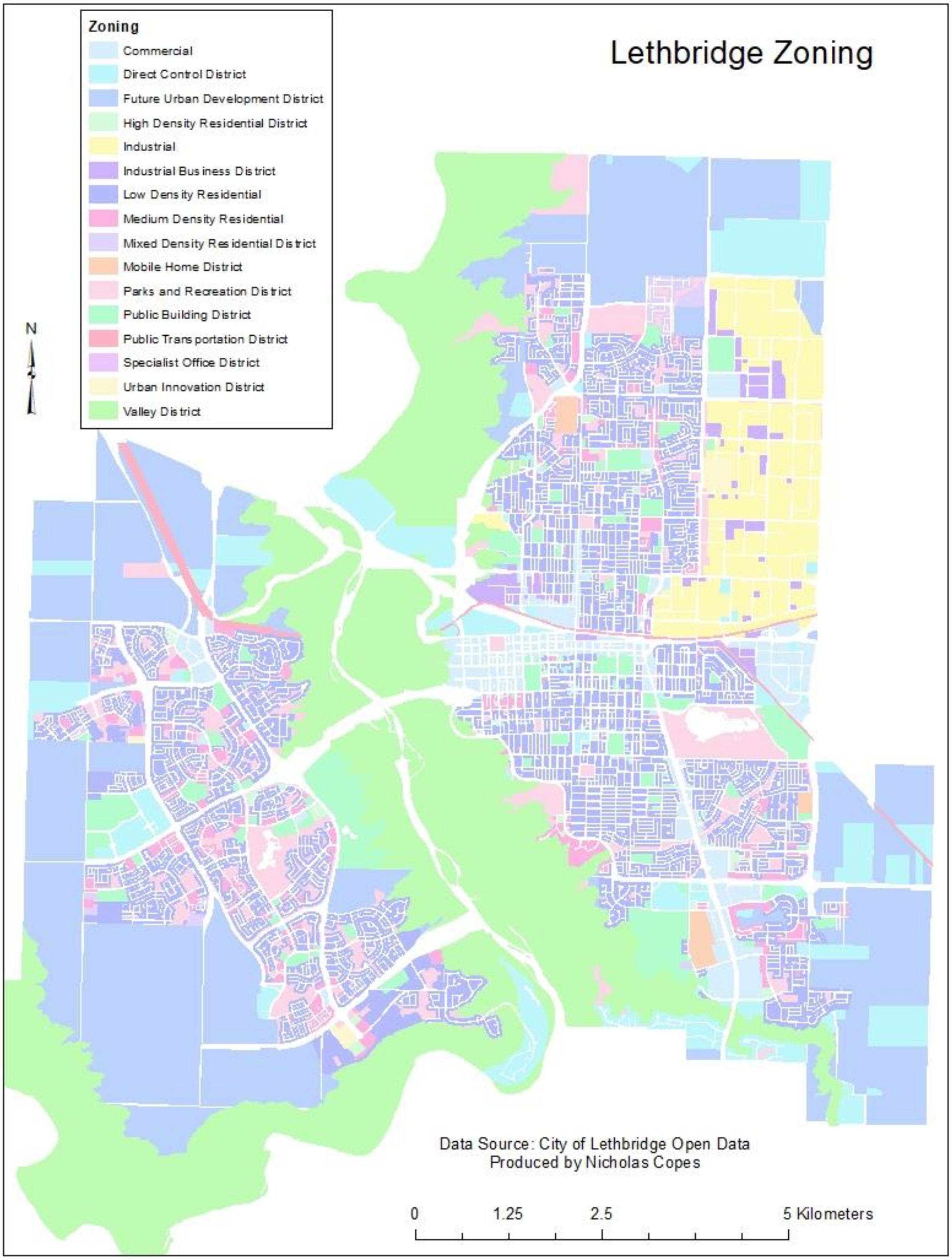


Chilliwack Zoning
Zoning
Agricultural
Airport
Commercial
Comprehensive Development
Ecovillage
Industrial
Mobile Home Park
Multi-Family Residential
Outdoor Recreation
Public
Reserve
Residential
Rural
University Village
N
Data Source: City of Chilliwack Open Data
Produced by Nicholas Copes
0
2.5
5
10 Kilometers

Saint John

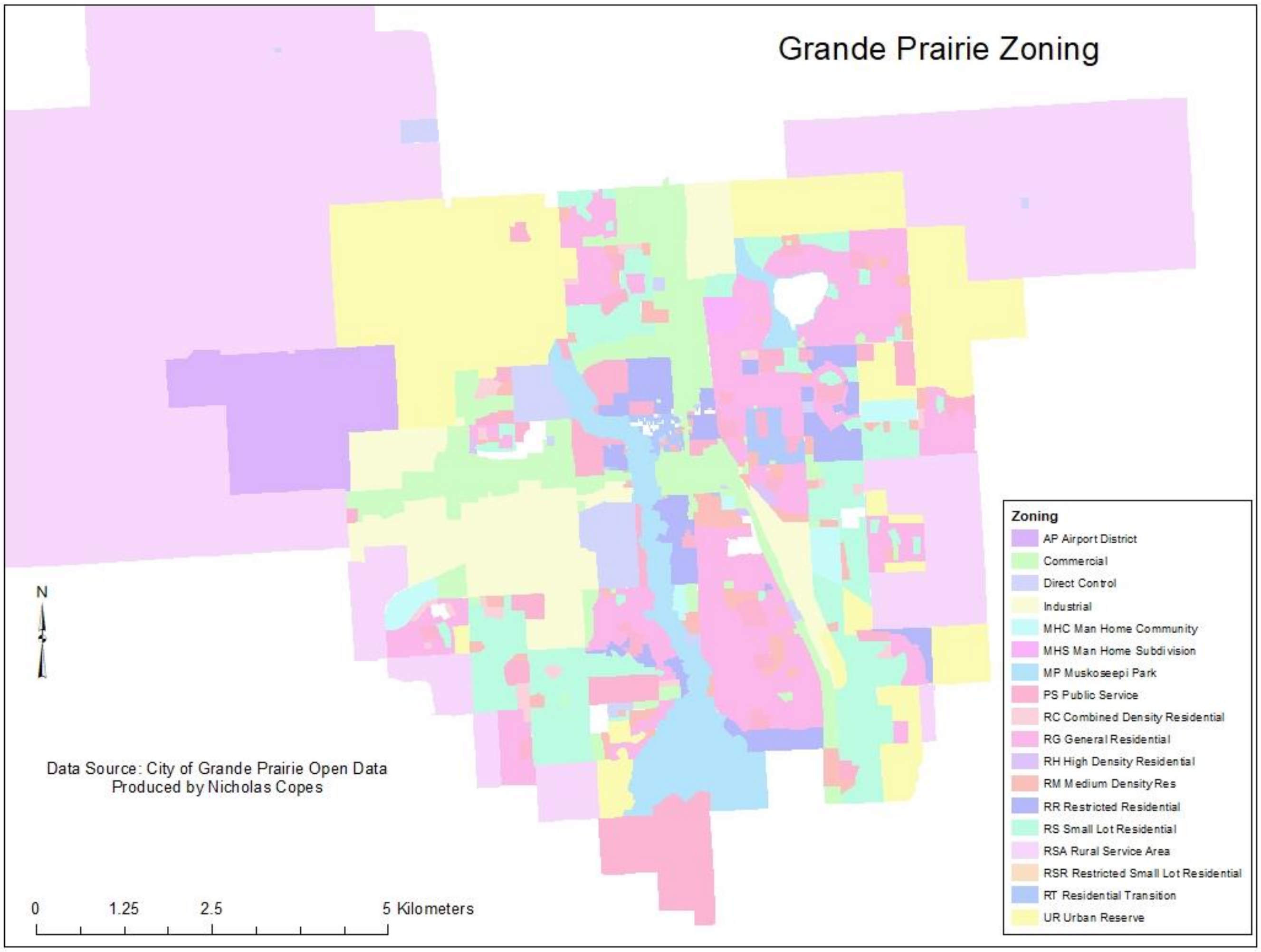

Saint John Zoning
N
Data Source: City of Saint John Open Data
Produced by Nicholas Copes
0
2.5
5
10 Kilometers
Zoning
Commercial
Community Facility
EPA
Future Dev
Industrial
Integrated Development
Low Density Residential
Multi-Unit Residential
Park
Pit & Quarry
Rural
Special Zone
Transportation
Utility Service

Lethbridge

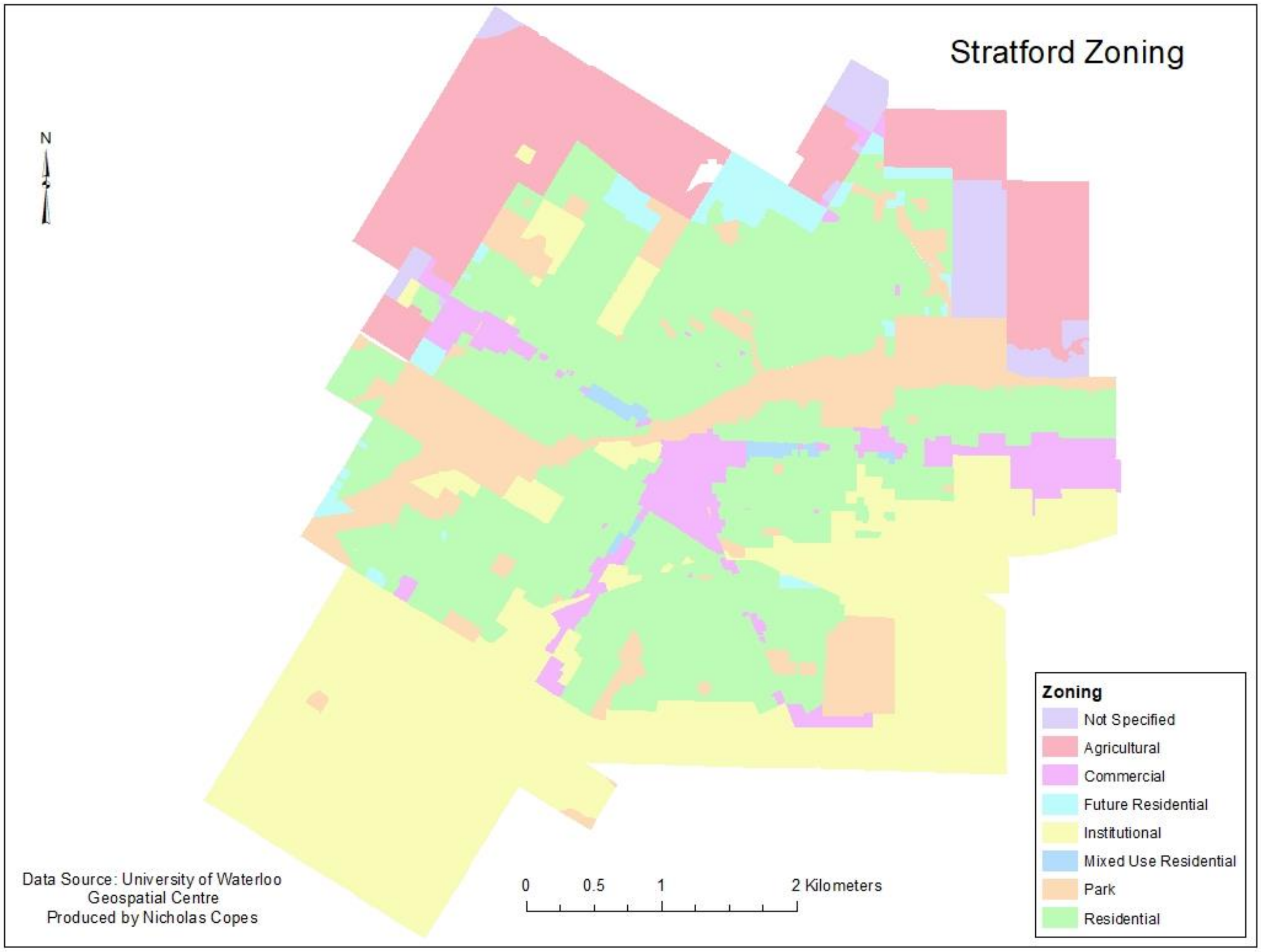


Lethbridge Zoning
Zoning
Commercial
Direct Control District
Future Urban Development District
High Density Residential District
Industrial
Industrial Business District
Low Density Residential
Medium Density Residential
Mixed Density Residential District
Mobile Home District
Parks and Recreation District
Public Building District
Public Transportation District
Specialist Office District
Urban Innovation District
Valley District
N
Data Source: City of Lethbridge Open Data
Produced by Nicholas Copes
0
1.25
2.5
5 Kilometers

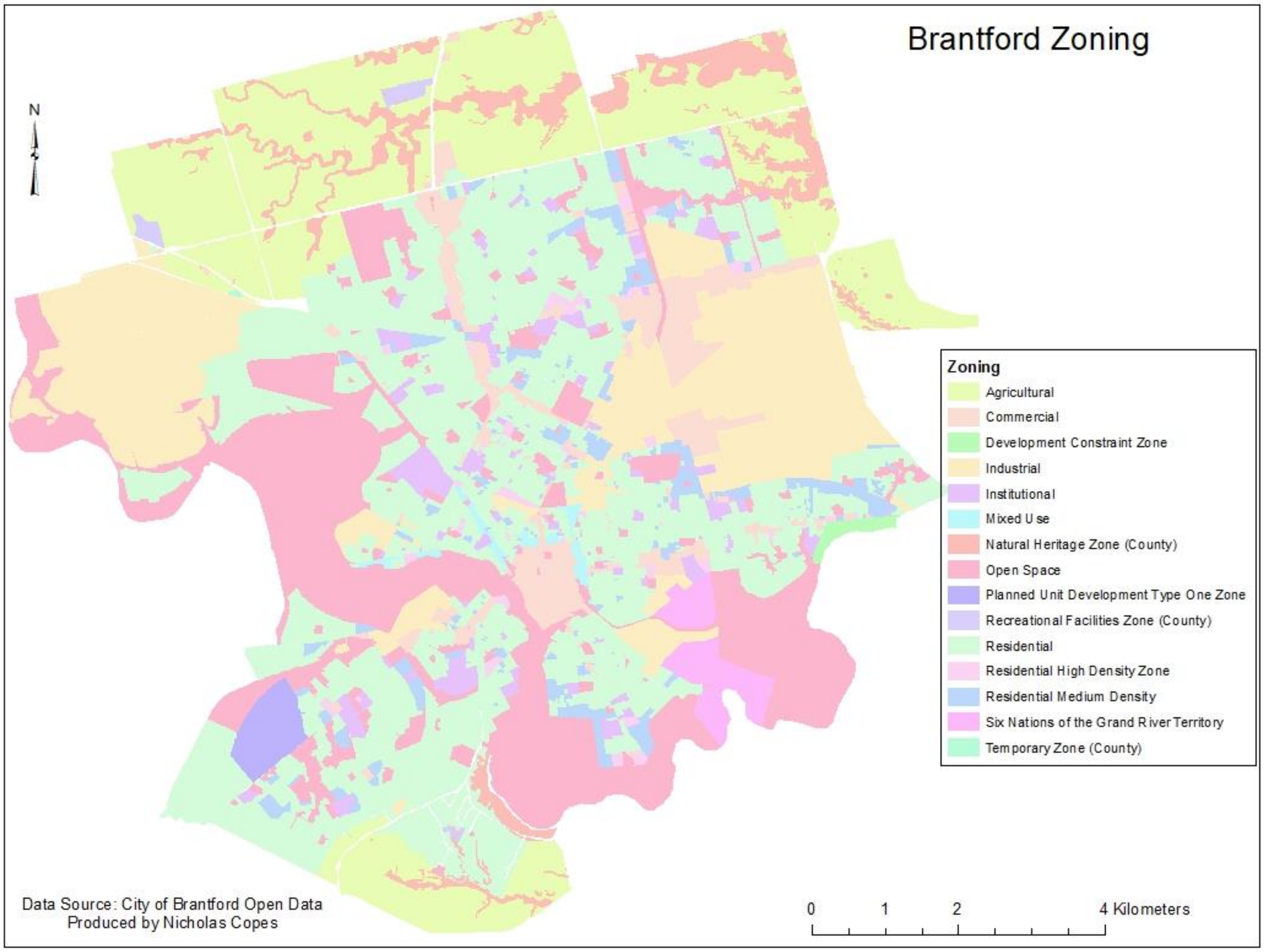
Grande Prairie



Stratford


Stratford Zoning
N
Data Source: University of Waterloo
Geospatial Centre
Produced by Nicholas Copes
0 0.5 1 2 Kilometers
Zoning
Not Specified
Agricultural
Commercial
Future Residential
Institutional
Mixed Use Residential
Park
Residential

Brantford



Brantford Zoning
N
Zoning
Agricultural
Commercial
Development Constraint Zone
Industrial
Institutional
Mixed Use
Natural Heritage Zone (County)
Open Space
Planned Unit Development Type One Zone
Recreational Facilities Zone (County)
Residential
Residential High Density Zone
Residential Medium Density
Six Nations of the Grand River Territory
Temporary Zone (County)
Data Source: City of Brantford Open Data
Produced by Nicholas Copes
0 1 2 4 Kilometers

Appendix 4: Metadata for Municipal Datasets

Municipality	Building Footprints	Zoning	Parcels
Halton Hills	Data provided under license to UW, Jan 9,2017	2010 Zoning	Info not available
Niagara Falls	Derivative product from ortho imagery flown region wide in 2010. Last edited May 31, 2012. No ongoing maintenance.	Official land use dataset has ongoing maintenance, last edited Feb 15, 2018. Source: Official Plan Document Planning, Building and Development - City of Niagara Falls.	Internally updated using PIN Searches - Land Registry Office, Registered Plans, Reference Plans, Property Merges and MPAC Sales reports. Original layer was created using Ontario Base Mapping as hard copy base, Ortho Imagery and Regional hybrid property data used to Quality Control layer.
Kamloops	Created May 19, 2016, periodic updates	Created April 29, 2016, periodic updates	Created Feb 26, 2018, periodic updates
Prince George	Created Feb 11, 2016, updated Jan 4, 2018	Created Feb 8, 2016, updated Jan 3, 2018. Zoning by-law classes adopted April 30, 2007	Created Feb 11, 2016, updated Jan 4, 2018. Parcel boundaries as defined by the BC Land Title System.
Chilliwack	Derived from 2016 LIDAR data. This data is uploaded to the Open Data web page weekly and is as current as the City of Chilliwack database.	This data is uploaded to the Open Data web page weekly and is as current as the City of Chilliwack database.	This data is uploaded to the Open Data web page weekly and is as current as the City of Chilliwack database.
Saint John	Created Sept 6, 2017, periodic updates	Created Nov 21, 2017, periodic updates	Created Jan 22, 2018, periodic updates
Lethbridge	BFs as of April 2015. Refresh Frequency: 2 Years. Updated Dec 17, 2015.	Refresh Frequency: As Available. Updated Dec 17, 2015.	Refresh Frequency: As Available. Updated Dec 17, 2015.
Grande Prairie	Created Nov 8, 2016. Outline of each building within the city based on either the real property reports (RPRs), or the roof top from an aerial photograph. As RPR's and new aerial imagery are acquired this dataset is updated.	Created Nov 9, 2016. Update frequency: as required.	Created Nov 14, 2016. City of Grande Prairie property parcels, this does not represent legal cadastre. Update frequency: monthly.
Stratford	Info not available. Data retrieved from UW Geospatial Centre.	Info not available	Info not available
Brantford	ISO 19139 Metadata Standard used. Last update Aug 9, 2018.	ISO 19139 Metadata Standard used. Last update Sept 5, 2018.	ISO 19139 Metadata Standard used. Last update Nov 29, 2018.