An Examination of Processes based on Open Standards in Support of Service Location

by

Mark Cheresna

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Master of Environmental Studies
in
Geography

Waterloo, Ontario, Canada, 2013

Authors 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 examiners.

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

ABSTRACT

A private telecom carrier partnered with the University of Waterloo to examine opportunities to improve their asset management processes. A reliance on traditional CAD technology made it difficult to generate an enterprise view of operational assets, such as poles and cables, since CAD documents were limited to neighbourhood scale coverage. The CAD documents had to communicate logical and locational properties of these assets. These requirements were often at odds since the elements in these CAD documents were occasionally moved to clarify logical aspects, the most common being connectivity with other telecommunications hardware. Elements within the drawings were also restricted to two dimensions, a legacy of early adoption of CAD technology within the telecom carrier.

Developments in GIS and architectural technology that have occurred since the introduction of CAD offer opportunities to manage assets using enterprise geospatial systems with three dimensional content. Prominent technologies and standards, such as CityGML and Oracle, will be examined to develop a model to support requirements related to service location. A service location for this paper is a site that requires the deployment of specific resources to meet the needs of a service request. Additionally, as location displacement is an issue that needs to be addressed, an evaluation of data quality processes related to location will be presented. The results from this evaluation will then be used to construct a final standards based 3D geospatial service location model, one that should address the needs of the partner carrier.

Acknowledgements

I would like to thank Dr. Doug Dudycha and Dr. Peter Deadman for their guidance, assistance, and positive and enthusiastic support. I would also like to thank Dr. Derek Robinson and Dr. Robert Feick for their roles as readers. Steve Xu and Jonathan Morgan from the University of Waterloo Map Library were extremely helpful in my data collection requirements and their responsiveness was impressive. My submission would not have been possible without the valuable administrative support provided by Susie Castella in Geography Division at the University of Waterloo. I am also grateful for the assistance and insight provided by two research team members. Peter Glenday, a PhD candidate with over twenty years of experience with GIS in the telecommunications industry, and Yunwei Dong, a PhD candidate, were both remarkable in their understanding of GIS telecommunications issues.

Individuals from outside the University of Waterloo were also helpful, including Stephen Perkins from the City of Ottawa, Chris Butler from Teranet, Dan Shannon from Telus, and Ryan Griffin from Cansel. I would also like to thank Jurgen Mack, John Scott, and Eyasu Ambaye for their insight and expertise with open source software and overall enthusiasm for GIS.

Dedication

To Debbie, Lauren, and Simon for their support, patience, and understanding.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

2D	2 Dimensional
3D	3 Dimensional
A-GPS	Assisted GPS
ALI	Automatic Location Identification
ANI	Automatic Number Identification
AOA	Angle of Arrival
ARP	Address Resolution Protocol
BBC	British Broadcasting Corporation
BIM	Building Information Model
BISDM	Building Interior Space Data Model
CAD	Computer Aided Design
CMB	Community Mail Boxes
CO	Central Office
COTS	Commercial Off The Shelf
CRTC	Canadian Radio-television and Telecommunications Commission
CSS	Cascading Style Sheets
CWTA	Canadian Wireless Telecommunications Association
DDL	Data Definition Language
DEM	Digital Elevation Model
DMTI	Desktop Mapping Technologies Inc.
DTD	Document Type Definition
DWG	AutoCAD Drawing File
E-911	enhanced 911
EAR	Enterprise Archive
EOTD	Enhanced Observed Time Difference
EPSG	European Petroleum Standards Group
ESRI	Environmental Systems Research Institute
ETL	Extract, Transform, and Load
FME	Feature Manipulation Engine
FSA	Forward Sortation Unit
FTTH	Fiber to the Home
GD	General Delivery
GDI NRW	Geodata Infrastructure North-Rhine Westphalia
GIS	Geographic Information System
GML	Geography Markup Language
GNSS	Global Navigation Satellite Systems
GPS	Global Positioning Satellite
HTML	Hypertext Markup Language
HVAC	Heating Ventilation and Air Conditioning
IDE	Integrated Development Environment
IFC	Industry Foundation Classes
IP	Intellectual Property
IP	Internet Protocol

ISO	International Organization for Standardization
ITU	International Telecommunications Union
JSON	JavaScript Object Notation
KML	Keyhole Markup Language
KMZ	KML compressed format (zipped)
LDU	Local Delivery Unit
LL-CSRS	Latitude/Longitude – Canadian Spatial Reference System
LL-WGS84	Latitude/Longitude WGS 1984
LOD	Level of Detail
LVR	Large Volume Receiver
MAC	Media Access Control
MSI	Mobile Spatial Interaction
OASIS	Organization for the Advancement of Structured Information Standards
OECD	Organization for Economic Cooperation and Development
OGC	Open Geospatial Consortium
POTS	Plain Old Telephone Service
PPGIS	Public Participatory GIS
PSTN	Public Switched Telephone Network
QGIS	Quantum GIS
RAM	Random Access Memory
RDBMS	Relational Database Management System
RSS	Received Signal Strength
SIG	Special Interest Group
SQL	Structured Query Language
TDOA	Time Difference of Arrival
TIC	Terrain Intersection Curve
TIGER	Topologically Integrated Geographic Encoding and Referencing
TIN	Triangulated Irregular Network
TOA	Time of Arrival
UGC	User Generated Content
UML	Unified Modelling Language
VoIP	Voice over Internet Protocol
VGI	Volunteered Geographic Information
W3C	World Wide Web Consortium
WCS	Web Coverage Service
WFS	Web Feature Service
WGS	World Geodetic System
WMS	Web Mapping Service
xAL	eXtensible Address Language
XML	eXtensible Markup Language
XSD	XML Schema Definition
XSLT	Extensible Stylesheet Language Transformations
ZIP	Zone Improvement Plan

1 Introduction

The telecommunications industry has defined and shaped the information age. It is no coincidence then that information management in support of services should be considered as a critical enterprise function that enables decision making and contributes to innovative thinking. Information management is necessary to systematically and predictably maintain services and assets, and accommodate client needs. If recognized as a core component of that information, spatial data can serve as a common element that can bind what might otherwise be incongruous data.

Reliance on spatial data could, however, introduce vulnerabilities into application databases if spatial features are duplicated, ambiguous, incomplete, or incorrect. These negative quantitative and qualitative aspects may be embedded in either the attributes of geographic features, which describe the characteristics of the representative object, the coordinates of the features - or both. Qualitative and quantitative issues could exist at a more abstract level, for example, in geographic layer metadata. As such, processes need to be put into place to identify and rectify data issues that could undermine confidence in spatial databases.

These aspects should not be examined in isolation since by doing so they could deter an enterprise from integrating spatial data into core information systems. On a more positive note, the use of spatial data in a core business application is implicit in such businesses as Call Before You Dig (One Call) and E-911. Applications in support of these operations provide information that is specific to a site requiring servicing, or in other words to a service location. These examples demonstrate that spatial content can extend functionality that would otherwise be difficult to implement using non-spatial databases.

Detailed information on resource deployment is essential for maintaining assets, identifying and repairing underperforming infrastructure, and delivering emergency and support resources. Although these sample activities benefit from geospatial information, a more comprehensive approach would ensure that spatial information provides more depth to broader enterprise functions such as marketing, infrastructure planning, and strategic planning. These benefits cannot be realized without tighter synchronization and interoperability in business and data management processes.

1.1 Technology Trends in the Telecommunications Industry

Telecommunications companies have to operate in a very competitive market that, to survive, must take advantage of economies of scale in order to remain competitive and maintain research and development activities (Houseman, 1991). Economies of scale promote integration, interoperability, and convergence as business strategies (Chan-Olmsted & Jamison, 2001) and a successful execution of such strategies should give an enterprise an advantage when it has to cope with market consolidation pressures. To clarify, interoperability is “the ability of a system or components of a system to provide information sharing and inter-application process control, through a mutual understanding of request and response mechanisms” (A.T. Kralidis in Hall and Leahy, 2008 referencing Groot and McLaughlin, 2000, p. 4), and provides a net gain or enhancement resulting from improvements in the interaction of the component parts of a system (Gridwise.org, 2009; State of New York, 2011; De Sousa & Stuckmann, nd). Integration relies on “the combination of a variety of sources of information, from fundamental science to the details of the manufacturing environment" (Iansiti, 1995, p. 522) and requires a thorough analysis of technological options (Iansiti, 1995). One potential outcome of business integration is convergence, which is characterized as the delivery of increased functionality and services across complementary product lines (Plevyak & Sahin, 2010). Convergence is driven by “business-reduced cost, enhanced capability and the ability to enter new markets” (Jones & Cowan, 2007, p. 72) and is an established trend in the telecommunications industry (PWC, 2006) as it strives to deliver convenience to the consumer. Consequently, the convenience that has resulted specifically from the convergence of voice and data technologies has increased societal dependence on mobile devices (Jones & Cowan, 2007), with the underlying telecommunication networks “becoming the foundation on which people are, more and more, building their human and social networks” (Plevyak & Sahin, 2010, p. 15). If one accepts that a technology dependence is well established and that demands for new services will continue to push for convergence in the marketplace, then model developers should anticipate any impacts that might arise from convergence.

Technological convergence can result from different business tactics: organic (internal) growth, strategic alliances and partnerships (external), or through mergers and acquisitions (external), with each bringing benefits and risks (PWC, 2006). For companies that support both infrastructure and services, convergence of various business lines may not result in expected improvements in efficiency and levels of service (Grove & Baumann, 2012) without carrying some negative risk. These risks are amplified through cultural and technological differences, and are more likely to result from mergers and acquisitions (PWC, 2006). The resulting modifications to systems and related workflows may make it difficult to adopt changes at a large scale (Morgan, 2009). As system infrastructure changes, so does the need to manage data, regardless of business strategy.

Technological advancements are not isolated to the telecommunications industry. Advances in Geographic Information Systems (GIS) and mapping have moved the technology from relative public obscurity to a mandatory requirement in mobile applications, as demonstrated by the market reaction to Apple Inc. when it dropped support for Google Maps in favour of their own map service (Forbes, 2012). Like the telecommunications sector, new products and services based on research and development in GIS will continue to extend the technology into new markets, and offer opportunities for integration and convergence in business areas that were once on the periphery of the technology. Building Information Models (BIMS) and 3D applications and data resulting from improvements in computer processing power have emerged as candidate technologies that can be used to consolidate business processes in sectors of the economy that are reliant on location based information. The telecommunications sector, which typically operates over a large spatial extent and with asset management that can benefit from 3D modelling, with horizontal (x,y) and vertical (z) dimensions, is ideally positioned to take advantage of these innovations. By investing in true 3D GIS data models and technology at the enterprise level, telecommunications companies should be able to improve asset management processes and data quality, thereby reducing operating costs by consolidating business processes, controlling inventory through an overall picture of assets, and reducing response times by improving the certainty of service location.

1.2 Goals and Objectives

This paper will demonstrate that the synchronization and integration of business processes, and in particular those related to service location, is more easily achieved with a consistent, accurate, unambiguous, and complete data framework that takes advantage of 3D GIS. For the purposes of this paper, integration is holistic, and is applied to business processes, systems, and data, the latter of which can be further subdivided into structure and content. Open standards, i.e. standards that are publically available and have been vetted by stakeholders in academia, industry, and government, contribute to a target framework by ensuring consistency in data models and interfacing technologies. Furthermore, industry standards that are open ensure continuity in supply chains and builds trust in collaborative business processes (ITU, 2005). Open data, i.e. data that are publically available and can be freely redistributed (Region of Waterloo, 2010), will also be examined as a method to augment enterprise data repositories, which in turn should extend business functionalities and capabilities.

The goal of this research is to answer the following question: can models and associated processes based on open geospatial standards that accommodate 3D modelling simplify and enhance the management of spatial data, specifically with regards to service location and related asset management? Objectives in support of this goal include activities that:

- develop an understanding of service location and identify data requirements and candidate data sources that would enhance current spatial data holdings in support of service location
- evaluate and, to the extent necessary, modify and extend standards-based data models to create an interim service location database that incorporates current trends that suggest that true 3D modelling is becoming an expectation in GIS;

- evaluate two different methods to collect service location information, one based on geocoding with another based on a detection process using orthoimagery, and assess the results using a statistical comparison to baseline GPS data collected in the field; and
- incorporate collected field service information into a revised final service location model.

A telecommunications carrier partnering with the University of Waterloo identified a number of research problems that were investigated by a telecommunications research team within the Department of Geography and Environmental Management, with an applied solution to service location issues being one of a handful of priority areas of research. The research activities related to the outlined objectives were identified in response to the requirements of the partner carrier. The results of these research activities will be incorporated into a prototype model, relying on secondary data from external sources for core content supplemented by primary data collected in the field. The final model will be focused on the dissemination of asset and client location, but the findings should be applicable to the broader utility sector.

1.3 Thesis Organization

Following this introduction, a review of the current issues in the telecommunications industry related to GIS is discussed in Chapter 2, which includes a literature review. A methodology to support the objectives of this paper is described in Chapter 3. The methodology describes the data and preparation activities, service location algorithms, and framework for model analysis. Chapter 4 contains an analysis and implementation of processes to enhance service location data. The chapter examines data models as they currently exist and describes proposed revisions. Chapter 4 also identifies challenges to integration. A summary of findings is made in Chapter 5, which also includes a sub-section on future work.

Throughout this paper, typographic conventions are used. Table names and entities are capitalized and italicized. In addition to acronyms, tabular column names are fully capitalized. Computer code is presented in Courier font and blocks of code are indented. Tabular content is contained within single quotation marks.

2 Processes Related to Location Identification

In this chapter, the current state of data holdings for the partner carrier will be described, thus highlighting the challenges for service location at the enterprise level. Various location discovery technologies as they may relate to telecommunications will also be outlined. User generated content (UGC) will be examined as an optional source of spatial data. A brief historical outline of changes in the telecommunications field will be followed by a discussion of future trends. The review of this information should provide the reader with an understanding of the current and future limitations (or opportunities) of many suitable technologies, sources of data, and data models, which will then serve as a benchmark to which the objectives of this paper will be measured. This review will also build a case for the chosen methodology presented in Section 3.

An examination of data modelling and integration processes as a framework to support service location for the utility and telecommunications industries was inspired by the requirements of a national telecommunications carrier. The carrier developed an extensive set of CAD products over a number of decades to support asset infrastructure that detailed physical objects, such as poles and cables. Data compiled in CAD to address the service needs for each local exchange, also known as a central office (CO) (Goleniewski, 2007) as it is central to the neighbourhood it services, resulted in large quantities of partitioned data that were difficult to assess in terms of quality and completeness. In addition, an evaluation of a sample of data supplied by the telecom partner by other participants in the telecommunications research team revealed that a consistent coordinate reference system was not applied, making it more difficult to compile aggregated data, such as asset inventories. Engineering Services that operate within the partner telecom carrier have historically been as much concerned with the logical relationship of assets rather than the accurate spatial representation of assets. The schematic representation of features, which stress the logical relationship at the expense of spatial accuracy, makes it difficult to confidently understand the current state of assets as physically implemented. The London Underground (i.e. subway) transit system diagram, technically a cartogram, is a well-known example of how abstraction can convey meaningful information. The same conceptual approach is used by transit authorities elsewhere, even if the presentation of such information does vary (Longley, Goodchild, Maguire, & Rhind, 2011). Spatially, the Underground is not accurately represented in the diagram but the important logical relationships between lines are communicated. The diagram cannot be used to manage maintenance on the lines, or even be used to accurately determine trip duration, without supplemental information. Similarly, the connectivity of poles, cables, and demarcation points are functionally an important part of infrastructure and asset management in the telecommunications industry. Schematic diagrams help to communicate these concepts, but they do not support all operational activities. A final limitation with the inventory of CAD files managed by the partner carrier is that many are 2D models that cannot account for overhead and underground features without introducing displacement.

Modifications to the underlying information managed by the partner carrier should improve the quality of data and reduce redundancy. This paper identifies ways to prepare and transform data obtained through various location detection strategies, which can be evaluated against GPS points collected in the field. This paper also identifies processes to share and consolidate data to enhance enterprise applications that are reliant on spatial data.

Established enterprises in the telecommunications sector are dependent on assets that are geographically dispersed and interconnected, often combining both infrastructure and service business lines. Infrastructure business lines concentrate on managing and maintaining physical assets, such as vaults, ducts, cables, towers, etc. Service business lines are less tangible and are focused on the delivery of content, whether that be voice communications, internet, video, etc.; however, these services are dependent on the underlying physical infrastructure. Partly as a result of locational dependencies, combined business lines result in complex systems. Spatial data as a common element can act as a catalyst for integration and interoperability, and can form a basis for system improvements related to service location.

An examination of opportunities to identify integration benefits in GIS should start with data and data flows. A spatial data baseline for service location can be established using traditional datasets from private and public suppliers that are derived from field and aerial surveys. A baseline developed using these sources is suitable for fixed assets that are easily verifiable; however, a static baseline is insufficient for assets that vary over space and time.

Diverse technologies have enabled service providers to locate assets and customers in real time, if necessary, using a number of approaches. Many technologies superimpose client location information on baseline data, and typically these technologies are not used to derive data. Regardless, an examination is necessary to determine if there are candidate technologies that can be used to augment baseline data.

2.1 An Evaluation of Sources of Location Information

Global Navigation Satellite Systems (GNSS), the most common being the US-based Global Positioning System (GPS), is now almost universally known. Data from wireless devices can be integrated with coordinates from GNSSs to establish absolute position. Indoor locations of devices can be determined using the decay of Received Signal Strength (RSS) (Lee & Lee, 2007; Fang & Lin, 2012) or through indoor beacon services, using for example Cricket technology (Kwok, Lau, & Lau, 2007). Alternatively, computer networks can be traced using Media Access Control (MAC) addresses through a router Address Resolution Protocol (ARP) table (Schulzrinne & Arabshian, 2002). Service providers and clients can incorporate direction vectors and bearings from accelerometers and magnetometers embedded in mobile devices, and these devices can be used underground or in buildings (Thiagarajan, Biagioni, Gerlich, & Eriksson, 2010). Client devices can be located based on the nearest cellular tower, or within a cell. Cell-ID (Sadoun & Al-Bayari, 2007; Trevisani & Vitaletti, 2004), also referred to as Cell Of Origin (Kwok, Lau, & Lau, 2007), is a simple, cost effective, low resolution solution to determine the location of wireless devices. An improvement on Cell-ID is trilateration (Kwok, Lau, & Lau, 2007) or triangulation, which can rely on different methods such as Time of Arrival (TOA), Enhanced Observed Time Difference (EOTD), Angle of Arrival (AOA) (Lee & Lee, 2007; Sadoun & Al-Bayari, 2007; Kwok, Lau, & Lau, 2007) and Time Difference of Arrival (TDOA) (Colman, 2001). TOA and EOTD take into account time delay while the less accurate AOA is based on angular measurements. These methods can also be enhanced using satellite navigation systems or Assisted GPS (A-GPS) (Sadoun & Al-Bayari, 2007; Kwok, Lau, & Lau, 2007). Although all these technological solutions provide locational information, each option serves different purposes and has different thresholds for range and accuracy. Before any can be dismissed as a data source solution, further evaluation is needed.

Studies on the locational accuracy of Cell ID vary. Trevisani and Vitaletti (2004) indicate that Cell ID may only be accurate from between 500 and 800 metres, while Kwok, Lau, and Lau (2007) indicate that location determined by Cell of Origin could be accurate from anywhere between 100 to 1000 metres. Signal triangulation in Canada is accurate up to 100 metres (Bell, 2011; Rogers, 2011; Telus, 2011), and similarly in the United States, E-911 Phase II can interpolate location as far as 300 feet (approximately 100 metres) from the actual location if GNSS is not available (FCC, n.d. a). The accuracy of locations established by GNSSs is affected by barriers along the line of sight (LOS) (Fang & Lin, 2012; Shaik, Das, Zhao, & Liao, 2009) and has limited function, if any, indoors or underground (Gong, Chen, Bialostozky, & Lawson, 2011; Lee & Lee, 2007). Depending on urban form, the accuracy of GNSS signals can be degraded (Thiagarajan, Biagioni, Gerlich, & Eriksson, 2010). RSS is only valid for up to 100 metres (Shaik, Das, Zhao, & Liao, 2009). Beacons can interfere with each other, their signals attenuate, and the signal range is even less than that of RSS (Kwok, Lau, & Lau, 2007). Table 1 contains a summary of accuracy and range for each location method. While the accuracy of some of these methods might be suitable for specific business requirements, none provide an accurate, continuous, contextual data framework.

Table 1 - Summary of Methods to Detect Location

Location Method	Accuracy	Range	Comments
Cell ID or Cell Of Origin (COO)	100 to 1000 metres (Kwok, Lau, & Lau, 2007)	Varies	Dependent on cell tower density. Functional indoors (Kwok, Lau, & Lau, 2007).

Location Method	**Accuracy**	**Range**	**Comments**
Triangulation or Trilateration	100 metres (Kwok, Lau, & Lau, 2007)	Varies	Dependent on cell tower density. Requires at least three cell towers to approximate a location. Needs LOS.
GPS	Dependent on quality of receiver. 10s of metres (Kwok, Lau, & Lau, 2007) or better	Global	Affected by urban canyons. Not functional indoors or underground. Needs LOS of at least 3 satellites.
RSS	Relative location	100 metres	Limited functionality.
Beacon	Relative location (Kwok, Lau, & Lau, 2007)	<100 metres	Identifies closest beacon to determine location. Limited functionality.
A-GPS	10s of metres (Kwok, Lau, & Lau, 2007) or better	Global	

From the perspective of consumer devices, the limitations and capabilities of the technologies highlighted so far suggest that “the future will be about hybrid positioning systems” (Wireless News, 2009), implying that a combination of technologies is necessary to support location based services. This observation is further solidified in a Space Daily article, which stated that “A-GPS, Wi-Fi and Cell-ID will be the winning combination offering accuracy, availability, interoperability and short fix times at low cost” (Space Daily, 2009). Situations in which any one or combination of these technologies can contribute to an improved spatial data infrastructure should be identified to ensure and improve services and to create a more complete data framework.

From a data perspective of service location, a model has to support the requirements of various client devices, and indirectly the methods to determine position derived from a number of fixed and mobile devices amongst many service providers. Interfacing devices would benefit from the integration of model representations of physical assets, such as cable infrastructure, cell towers, or sensor devices, which functionally may not be directly related. In the telecommunications industry, the representation of assets requires a model that retains a level of detail and precision necessary to support engineering and architectural processes. An adopted model should provide an intricate, micro-level view of telecommunications infrastructure components. A model must also provide a macro-level view of the service area to support such functions as optimized routing or cell phone coverage. Furthermore, a model must be able to handle data that are not homogeneous in scale, vintage, and collection methods (Gröger & Plümer, 2012). Representation of features alone cannot form the basis of modelling, as value is driven by relationships between representative features. Spatial information that supports a variety of services in the telecommunications sector is dependent on connectivity, adjacency, containment, and other topological relationships. Data models and databases must define rules that both constrain and enhance the topologies, relationships, and properties of representative objects. The relationships between geometric street networks to address ranges and more specifically textual civic addresses provide a framework through which location information can be derived. Processes to critically evaluate the accuracy of the content in these models and frameworks, if not in existence, need to be put into place to ensure that new content gathered through increasingly accurate means can be imported into a database without inheriting any significant distortion from the target system. Alternatively, sub-standard data from external sources could more easily be identified when a target framework expects that data meet or exceed an established threshold of quality. With these points in mind, GIS offers an ideal a solution to address these challenges in enterprise systems, since the technology can offer seamless coverage of assets at various scales.

A model that abstractly represents a physical object is incomplete without explicitly accounting for three dimensional (3D) geometric space. Thill, Dao, & Zhou (2011) observed that as urban populations increase and development becomes more vertical, a need for 3D modelling becomes more apparent. Operational processes also need to model information in 3D to manage assets, and this need is highlighted with recent concerns about cross bores in the utility industry (CTV News, 2013). Evaluating the needs of planners, Köninger & Bartel (1998) proposed that a 3D model would have to accommodate certain characteristics, which can be summarized in the following equation:

3D-urban-GIS = 3D-city model + thematic information + effective data storage and administration + planning analysis functionality (Köninger & Bartel, 1998, p. 81).

The equation promotes 3D GIS as a viable solution if all the components of the equation contribute to a final model. Köninger & Bartel (1998) emphasized that semantics and object oriented modelling can factor into the equation. Semantics give an object meaning, function, and context. With regard to movement, a door is not just a geometric object but has a transitory function, while a window is a barrier (OGC, 2012). A door is a component of a room, which is further defined by walls, ceilings, windows, and entrance ways (Gröger & Plümer, 2012). The concept can be extended out in scope beyond individual structures. A building has a contextual relationship to other buildings and adjacent roads, and so on. Logical relationships in semantic models has some overlap with spatial topologies that are inherent in GIS. The representation of features and their relationships at various levels of detail has the potential to contribute to the effectiveness of 3D modelling for service provision.

An enterprise solution has to look beyond traditional GIS for content, and GIS must contribute to other models and business processes to remain relevant. Building Information Models (BIMs), which are used in architectural and engineering sectors, represent different aspects of building design, construction, and management. BIMs go beyond traditional 2D architectural profiles, floor plans, and elevations and offer a rich source of information in three dimensions (van Nederveen & Tolman, 1992; Autodesk, nd). BIMs can enhance a spatial information system intended for asset management. For example, interior and exterior surfaces can be built from various materials that perform different functions, the modelling of which can improve facilities management and building maintenance processes (Morgan, 2009). Considering the value of semantic models and the potential source of information available from engineering and architectural domains, complementary aspects of BIMs and GIS will be further investigated in Section 4.6.

2.1.1 Public Participation in the Generation of Content

Although sophisticated models such as BIMs can enhance an enterprise GIS, more modest solutions might offer opportunities to improve content. User Generated Content (UGC) is a broad term that encompasses publicly contributed data and information, but more detailed and specific terms in the field of geography do exist. Public Participatory GIS (PPGIS), a term first defined in 1996 (Seiber, 2006), is a process that encourages the public to contribute to and participate in GIS projects and technology (Longley, Goodchild, Maguire, & Rhind, 2011). Volunteered Geographic Information (VGI) categorizes spatial content collected and submitted by the public at large, and as such is more difficult to validate than content prepared through traditional systematic processes (Longley, Goodchild, Maguire, & Rhind, 2011).

2.1.1.1 Casting a Critical Spotlight on PPGIS

If user-generated content is not subjected to the rigour of trained technicians and analysts, then it will be treated with a level of scepticism (Elwood, Goodchild, & Sui, 2012). Heipke (2010) suggests that crowd sourced data sets are developed by people with a common interest that subject themselves to different standards than what might be conventionally used. Similarly, Rama and Joliveau (2009) note that many advocates of neogeography, which is collaborative content facilitated by Web 2.0 technology (Longley, Goodchild, Maguire, & Rhind, 2011), are non-academics that collect data for a specific purpose, and they do not necessarily support the general framework that traditional geographers strive to develop. These points suggest that the definition of a validation process would pose a challenge.

Goodchild (2007) questioned the motivations of contributors of VGI, suggesting that the same human qualities that create viruses, denial of service attacks, and other negative aspects of the broader internet would subvert the efforts of altruistic contributors of user-generated content. Saroiu & Wolman (2010) validate this concern in their discussion on participatory sensors, noting that falsifying data from sensors could negatively impact the trustworthiness of data to the point that it could thwart application development.

In addition to quality concerns, legal liability could undermine the relevance of user-generated content (Elwood, Goodchild, & Sui, 2012). If users donate information and the use of such information results in an injurious mishap or an infringement of intellectual property, then it could be difficult to establish who is accountable. As Goodchild asked "Can a user who contributes geospatial information be eventually held liable for damages that result from its use?" (Goodchild, 2008, p. 8). The examination of conventional mapping processes demonstrates that the breakdown of responsibility between users and providers is not straightforward. The division of responsibility might be difficult to establish when various providers contribute to the data chain, through which errors can be introduced at any point in the data preparation process (Janssen & Crompvoets, 2013). Janssen & Crompvoets (2013) provide a number of cases where the omission of data from navigational and aeronautical charts was deemed to be the responsibility of product developers that absorbed incomplete or inaccurate data from government suppliers into their products. Janssen & Crompvoets (2013) also provided examples where end users had a duty to exercise professional judgement and caution in interpreting spatial data. If there are liability challenges in the development of conventional spatial products, then UGC would likely make things even murkier. Terms of use clauses that remove liability from the producers of content likely will not be acceptable in mission critical systems.

2.1.1.2 Focusing on the Positive

From a geospatial perspective, public participation through user-generated content can contribute to the timeliness of information. User-generated content has been shown to provide valuable information in dynamic environments, such as fire outbreaks (Elwood, Goodchild, & Sui, 2012) or with the identification of the location of landmines (Dunn, 2007), thus underscoring that volunteered information can enhance the quality and quantity of relevant information. At the very least, public input can highlight areas that have undergone change since the most recent authoritative release of information (Elwood, Goodchild, & Sui, 2012).

Elwood et al. (2012) go further to defend UGC. They temper concerns regarding data quality by noting that the internet delivers a platform that can facilitate a peer review, although not necessarily one based on rigorous standards. False, misleading, and inaccurate data are visible to a broader community that over time will ensure that contributors provide value to a repository of user-generated content.

2.1.1.3 A Review of Successful Participatory Projects

Elwood et al. (2012) highlighted the Audubon Society Christmas Bird Count as one of the earlier examples of public participation. This informal survey is restricted to people that have been subjected to an adequate level of training (Goodchild, 2007). In another example, Elwood et al. (2012), citing Irwin (1995), also noted on the use of volunteers to collect meteorological data, with support from expert meteorologists. Both cases demonstrate the legitimacy of user-generated content and both cases highlight that value is improved when guidance from experts is incorporated into the participatory framework.

2.1.1.4 The Application of User Generated Content for Service Location

Some of the benefits and risks of user generated content have been discussed. User Generated Content (UGC) has tremendous value in benign applications, especially when the availability of data is sparse, or when timeliness is a factor and traditional means are not available. The use of UGC in critical applications is more problematic due to issues of liability and reliability. If, however, content is not readily available, or existing data need to be validated, or processes can be established to enhance the collection of data by public participants, then UGC can play a functional role. Although caveat emptor does not strictly apply to UGC, since data is free, the principle has to be respected. The principle should be a driving force behind quality indicators and metadata development.

2.2 The Relationship of Telecommunications Infrastructure to Location

The public switched telephone network (PSTN) provides the framework for public land-line solutions (Goleniewski, 2007). Typically, residential users in established neighbourhoods directly connect to the public network using traditional plain old telephone service (POTS) copper wire lines. Copper wire attenuates and thus needs to be twisted to increase functional distance. Even with twisting, copper wire is only practically usable to 6000 feet (Goleniewski, 2007), and unavailable after 18,000 feet from the CO (Grubesic, Matisziw, & Murray, 2011). The limitation of copper is the source of the commonly known term “the last mile”. Amplifiers or repeaters on the network maintain sufficient quality beyond this distance (Goleniewski, 2007).

The first point of access to the larger telephone network through the local exchange. Long haul telecommunications with distances greater than 100km are handled by inter-city trunk lines that are reliant on fibre optic cabling (Agrawal, 2010). Fibre optic cabling is a high capacity technology that is functional between continents with the assistance of repeaters (Agrawal, 2010). Fibre optic cabling is also more commonly used in new residential installations (Koonen, 2006) since it overcomes many of the limitations of copper wire.

Regardless of which type of cable is used, wireline customers are fixed in location.

2.2.1 Telecommunications Infrastructure for Emergency Services

This paper does not attempt to offer any improvement to core emergency services from a technological systems standpoint, as they are well established; however, the results of this work should provide a process to improve data quality and content in support of emergency services (Peter Glenday, personal communication, January 24, 2013). The challenges for emergency services are also worth considering since there are likely issues that overlap with asset management service location.

Emergency services (911) were first introduced in 1968 in the United States (Schulzrinne & Arabshian, 2002). In North America, emergency service calls are routed through the public service answering point (PSAP). Emergency services through a wireline to the PSAP are capable of determining phone numbers using Automatic Number Identification (ANI) and caller's billing address using Automatic Location Identification (ALI) (Bennett & Regan, 2002). ANI/ALI are used in the event that a caller cannot provide their phone number or address, for example when the caller is distressed or suddenly incapacitated. If a connection were dropped, ANI/ALI enables call centres to re-establish the connection (Gow & Ihnat, 2004).

The expansion of wireless services has resulted in the removal of hard-wired links between phone number and residence and as a consequence has eliminated the certainty of caller location. Location information cannot be reliably retrieved using ANI/ALI with wireless technology (Bennett & Regan, 2002). This limitation is a concern since wireless customers make considerable use of 911. Over half of 911 calls in Canada (CWTA, 2012) and about 70% of all 911 calls in the United States (FCC, n.d. b) were made using mobile devices. In Canada, not all call centres have enhanced 9-11 (E-911) capabilities, which automatically assigns a user to a location. According to the CRTC, subscribers in Newfoundland, Yukon Territory, and the Northwest Territories must rely on Basic 911 over which the caller must relay their location to the operator (CRTC, 2012). Callers in Nunavut have neither Basic nor enhanced 911, and must use regular phone service in an emergency (CRTC, 2012). Some wireless subscribers do not have GPS capabilities, so triangulation must be used to identify emergency service location (CRTC, 2012).

Voice over Internet Protocol (VoIP) is a technology that enables communications over Internet Protocol (IP). There are two categories of VoIP: fixed and nomadic/non-native. Fixed VoIP is similar to wireline communications and as such is supported by E-911 (CRTC, 2011b). Nomadic VoIP, however, does not behave like wireline communications and as a result is not supported by E-911. A CRTC ruling demanded that service providers notify clients of the limitation of nomadic/non-native VoIP (CRTC, 2011a).

In Canada, telecommunications companies are required to prepare service location information to the municipal operations of E-911 services, which rely on geocoding processes (Peter Glenday, personal communication, Jan 24, 2013). Geocoding is a process of determining geographic location using address information (Longley, Goodchild, Maguire, & Rhind, 2011), but is not a complete solution since there are inconsistencies in quality of data in support of the process (Peter Glenday, personal communication, Jan 24, 2013; Murray, Grubesic, Wei, & Mack, 2011). The recognition of a quality gap rationalizes allocating time to develop solutions that address the issue.

2.3 Capacity Issues

The Organisation for Economic Co-operation and Development (OECD), in reference to a study by Shargal and Houseman (2009), evaluated the requirements of tracking household consumption of utility services and determined that the amount of information required to monitor hourly usage of electricity for one million homes "could pose a significant data warehousing and management challenge to utilities" (OECD, 2009, p. 19).This information technology burden would be further increased if clients supplying energy to the smart grid demand individual status reports. The OECD report also noted that regular polling of data could put a strain on internet communications channels that depend on variable load to deliver information. The limitation of spectrum capacity (BBC, 2012), in addition to limits in physical infrastructure, will also temper optimism for unbounded growth.

Implementation of any system is dependent on whether cost recovery is possible. The OECD (2009) noted that telecommunications infrastructure suppliers will invest in locations that are more likely to provide higher returns, thereby leaving locations most needy of investment even further behind the information divide. Furthermore, the paper highlighted that the lack of return on investment in these areas reduced competition, and as a result these areas, if serviced, lost the benefit of competitive markets. The relationship between infrastructure and competition was echoed by Welfens who noted that “insufficient investment in physical digital infrastructure” (Welfens, 2008, p. 87) is an impediment to competition. How then is investment to be encouraged?

One point of view is that government policy related to investment and regulation in telecommunications infrastructure can be used to establish an inventory of under-served areas (OECD, 2009). A contrasting view is that deregulation, such as that experienced in the United States and the UK in 1984 and in the EU in the 1990s (Welfens, 2008, p. 88), is more conducive to the development of new digital service technologies and businesses (Reichl, 2006; Welfens, 2008). Jones & Cowan (2007) suggest that deregulation within the telecommunications industry was in part driven by the rapid growth in the unregulated information technology sector. Jones & Cowan (2007) promote the concept that the convergence of information technology and telecommunications supports the need for a less regulated industry. This paper is of the opinion that technology has to work within both frameworks, and an open solution is more capable of accommodating both points of view.

Another challenge with service location, as proposed in this paper, is the overhead that 3D modelling and network and computer processing demands would bring to an enterprise. A small division dedicated to 3D modelling within an enterprise likely would not be noticed from a computational capacity perspective, but a roll-out of a solution to a larger set of end users could pose a problem. Either a solution would need to be rolled-out in step with upgrades to computer infrastructure, or enterprise stakeholders may be required to use a stripped down version of the model, or repositories would need to be strategically replicated to isolate impacts.

2.4 Into the Future

An examination of business processes must not only look at past requirements, but must consider probable future trends based on current leading edge technologies. A convergence of part or all of the technologies noted so far might contribute to the development of Mobile Spatial Interaction (MSI) technology (Fröhlich, Oulasvirta, Baldauf, & Nurminen, 2011), which on the whole can provide a more robust understanding of location. As an example of MSI, Fröhlich et al. (2011) discuss the use of technology that enables a user to obtain information by simply orienting oneself and pointing a mobile device towards an object of interest. Conceptually, MSI provides contextual information of physical objects in 3D space. The concept is intriguing but limited, since accuracy and performance are issues in a technology that is still in its infancy (Fröhlich, Oulasvirta, Baldauf, & Nurminen, 2011). If performance and accuracy issues can be resolved, the technology could offer opportunities to assist service calls in the field.

Advances in technology do not just improve the experience of users in isolation. Extending the value of smart technology, ubiquitous networks, also referred to as "pervasive computing" or "ambient networks" (ITU, 2005, p. 31), are an evolutionary advancement of the internet, embracing the always on behaviour of the internet to objects. A myriad of devices can be polled to support diverse requirements. Devices could, for example, be used to monitor environmental conditions, such as water usage and pollution levels, or transportation congestion (Desmond, 2010). Of the four criteria outlined by the ITU (2005) that define ubiquitous networks, one has particular geographic significance: devices can form topologies through "the ad hoc detection and linking of mobile devices into a temporary network" (ITU, 2005, p. 31). The report further notes that miniaturization will ensure an expansion of network aware devices. One could also extrapolate that geographic coverage and scope would also expand. Determining "the precise localization of objects" (p. 35) and reducing energy consumption, which can affect the range and life of a device, are identified as primary areas of research (ITU, 2005), suggesting that geography can play a pivotal role in a successful deployment of miniaturized technology by maximizing the spatial network of devices while minimizing energy resource requirements. If "[f]ixed phones define a network of places, mobile phones lead to a network of people" (Reichl, 2006, p. 304), then pervasive computing leads to a "network of things", which has the potential to become embedded into our collective routines (Technology Review, 2013).

2.5 Summary of the Location Problem

A number of methods can be used to detect location, but none offer a complete solution for service location. While various technologies have the potential to complicate the delivery of services and the management of underlying data, demands for more location based services will require that supporting solutions be spatially aware. In their analysis of the merits and challenges of convergence in telecommunications and information technology, Jones and Cowan stated that "[t]he ability to identify the actual user of a device will become increasingly important for service providers to be able to deliver the required services to the end user" (Jones & Cowan, 2007, p. 71). Considering this observation, established processes based on legacy address data and address range geocoding offer an incomplete solution since they do not provide seamless spatial coverage and are dependent on the accuracy and completeness of underlying street network data. These limitations provide an opportunity for alternative approaches to service location that do not replace but enhance geocoding.

There are examples, such as the One Call system in Ontario or the E-911 system in North America, that have demonstrated the relationship between geography and telecommunications, but the solutions have been applied to specific business processes. Even established processes like these need to be regularly evaluated to ensure that a chosen technology will support the intended life cycle (Krutchen, 2000) of the infrastructure the system is supposed to manage. A re-evaluation of data holdings can identify value added opportunities, which can introduce change. Changes in user requirements place new demands on a system (Krutchen, 2000). Ideally, each new innovation should be introduced seamlessly while retaining the capabilities of older technologies. New technologies may trigger the revision or development of standards, which enable future transitions in an infrastructure management system. New innovations must be compatible and current with technologies offered by other vendors (Desmond, 2010), introducing more project details for management.

This research has to demonstrate the capabilities of open standards based 3D geospatial models and processes in support of service location. Service location in telecommunications functions in 3D space, and a model that accurately accommodates overhead and underground assets should provide operational benefits to an enterprise. Key to demonstrating these capabilities is understanding service location, which is one of the identified objectives of this paper. The literature review in this section meets that objective. One of the key concepts coming out of this review is that models must work with a variety of scales and spatial environments and must be flexible enough to accommodate changing user requirements, business processes, and technological advancements. A second concept arising from this review is that data in support of service location can be collected using various technologies and can be submitted by different contributors. Although each source can bring benefits, they also have limitations.

A review of the service location concepts, challenges, and opportunities in this chapter sets the groundwork for a phased development of a service location model validated through collection of field data, geocoding, and detection processes using orthoimagery. The process must show that database entities can be integrated to support multiple business functions. A resulting model must be extensible to incorporate data from different sources, and be capable of supporting legacy data and future requirements.

3 Methodology

This section will describe the processes required to satisfy the remaining objectives of this paper, which are 1) to evaluate, modify, and extend standards-based data models, and 2) to evaluate methods to collect, correct, and assess service location information. Standards and technologies that support the methodology will be discussed. An overview of the methodology will be followed by more detail in dedicated subsections, outlining challenges that will need to be addressed before a successful execution of a step can be completed.

A methodology is a comprehensive approach to problem solving. The methodology for this project includes the architectural process, which involves planning, modelling and building a system (Krutchen, 2000). The assumed benefit to integrating spatial, CAD, and business data in support of service location is that a multi-faceted repository should serve as a stable framework from which disparate users can reap benefits that are not possible in isolation. From a physical and policy implementation, integrated technology can be used to systematically restrict access and limit functionality. These apparent contradictory qualities are necessary to ensure that an extensible solution is also a trustworthy solution.

3.1 Overview of the Project Methodology

Figure 1 outlines the general methodology. The executions paths were aligned with research objectives to 1) evaluate, modify, and extend standards based models, and 2) evaluate data collection, correction, and assessment processes in support of service location. The major processes in the methodology correspond to vertical swim lanes in Figure 1.

An initial state and inventory of data holdings needed to be defined. An interim model had to maintain core data sets, once identified, that provide context for service location. Once an interim model was established, core data from the model were used in processes that simulated the collection of service location data. An evaluation of simulated service location was used as an input to a final revision of a standards-based model.

CityGML was chosen as the template model to support a service location solution, because CityGML is an established open geospatial standard that can accommodate 3D modelling. Details on CityGML are presented in Section 4.2.3.3.

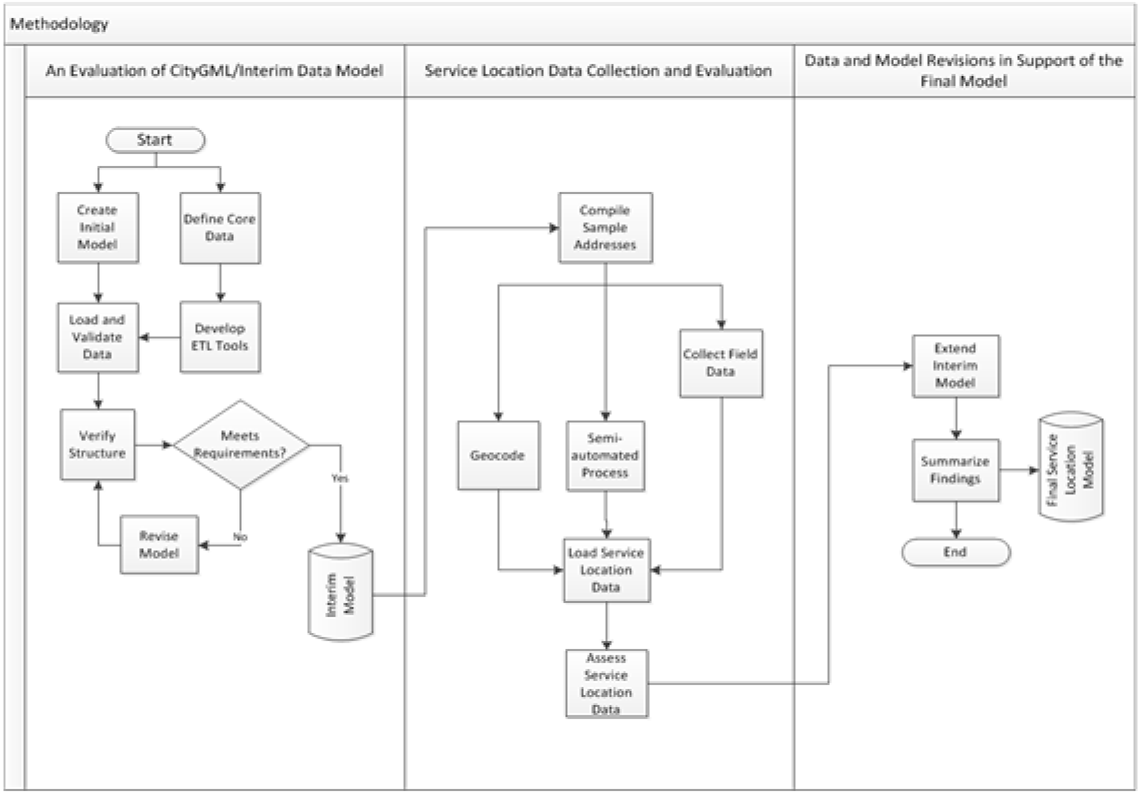

Figure 1 - A High Level Methodology

3.2 An Evaluation of CityGML in Preparation for an Interim Model

An evaluation process covered considerable activities that formed the foundation on which the methods for the collection of service location were dependent. Most sub-processes described in this section are contained in the first swim lane of Figure 1, but a few preliminary processes are not shown for clarity purposes. If there are a number of steps in a sub-process, a tabular summary is presented.

3.2.1 Identify Study Area

A study area was defined as one way to contain the scope of the research. Data availability and study area characteristics influenced the methodology. A prime consideration of this research was that the study area had to be accessible to verify results.

3.2.2 Examine Standards, Data Structures and Technologies

Technologies, standards, and existing frameworks were reviewed before any tangible data preparation was completed. The research for this paper strived to embrace international standards for Geographic Information Systems. Standards established by the Open Geospatial Consortium (OGC), a leading organization for GIS standards, the International Organization for Standardization (ISO), and the European

Petroleum Standards Group (EPSG), which developed standards for coordinate systems, were examined and used to the extent possible. Technologies were chosen primarily on their potential to function within the telecommunications and utilities sectors, as well as to work with CityGML. Open source alternatives and open standards were examined to either extend the scope of the primary technologies or to streamline the integration of the technologies. The steps are outlined in Table 2.

Table 2 - Steps to Evaluate Standards and Technology

PROCESS	DESCRIPTION
Evaluate and analyse standards and data structures	The components of the system were analysed and evaluated to develop an inventory of standards.
Evaluate and analyse technology solutions	Technologies were identified that supported chosen standards and data structures.

3.2.3 Create an Initial Data Model

The standards, data models, and technologies chosen through processes outlined in Section 3.2.1 were used to create an initial model. The initial data model worked in tandem with data requirements gathering and provided a conceptual framework for service location processes. The steps involved are outlined in Table 3.

Table 3 - Summary of Steps for Initial Modelling

PROCESS	DESCRIPTION
Create and modify model	A conceptual model further refined the scope and planned detail for the research. The initial logical model defined how the system met the needs of end users (Krutchen, 2000) and proved that the chosen models could co-exist. The logical model served as the groundwork for the physical model.
Identify and resolve access issues	Client configuration ensured network connectivity. Roles were created in the database. Access issues over network shares and mount points were resolved.
Revise structure in response to core data requirements	Model revisions worked in tandem with the compilation of core data.

3.2.4 Define Core Data

The compilation of core data was influenced by the initial data model, but the compilation process also fed back into the definition of the initial model. Specifications for projection and spatial accuracy were in part established after an examination of core data layers. These specifications were dependent on metadata associated with these data sets. As the requirements for civic addresses became more obvious, the expectations on source data containing address information were better defined, and deviations from the expected quality required additional attention.

The ability to stub data was a factor in data development. In this research, data were isolated to a municipality and therefore the province was created as a non-spatial table; however, a model developed for an enterprise would have to account for regional and possibly provincial and national geographies to at the very least facilitate QA processes (e.g. containment within jurisdictional boundaries).

3.2.5 Develop Extract, Transform, and Load Tools

Data had to be downloaded, uncompressed, and moved into directories structured to differentiate between data sources. Applications that managed the transformation process had to be installed and configured. Once these applications were installed, scripts that carried out Extract, Transform, and Load (ETL) tasks had to be developed. The development of these scripts supported an assessment of the identified ETL technologies.

The hierarchical dependence of assets determined priorities for the transformation of external source data to a target database schema. Using the construction industry as an analogy, a foundation has to exist to support the building frame, and a frame is needed before windows, doors, trusses, and a roof can be added. Similarly, some tables were populated before dependent tables to enable the establishment of relationships. For example, road names were imported into the database model before road segments to ensure that a single road name identifier could be used consistently for one or more similarly named segments.

A metadata process was established to communicate the state of the database, which reduced duplication and data degradation. Duplication was avoided by building processes that confirmed that an external data source had not been ingested as an existing data set in the database. Data degradation was avoided by ensuring that processes did not introduce data of poorer quality or from an earlier vintage by examining the contents of metadata beforehand.

Table 4 - Steps in the Development of Extract, Transform, and Load Components

PROCESS	DESCRIPTION
Prepare data	Data had to be downloaded and prepared for loading.
Configure ETL software	Software had to be installed and configured.
Evaluate ETL tools	An examination of data provided an opportunity to examine capabilities and strategies for extract, transform, and load. This could only be completed by developing scripts that performed ETL tasks.

3.2.6 Load Data

ETLs, which were discussed in Section 3.2.5, were executed in this process. Data transformation occurred in stages, with a QA/QC process at the end of each stage. The representation of data was validated before dependent features were introduced.

3.2.7 Validate Data

Validation included an examination of log files, which would contain data load issues. Systematic errors, such as primary key or foreign key violations, resulted in model revisions. Other errors, for example, data type mismatches that occur when loading a string element into a numeric field, resulted in revisions to ETL components. Validation included verifying data counts on source and target data. Sampling of resulting features identified data domain issues or data mapping issues (i.e. data being entered into the wrong field). Extraneous or missing features were visually identified in an external application, such as ArcGIS. Further analysis using layer and record level validation functions in Oracle Spatial were used.

Table 5 - Steps to Load Data

PROCESS	DESCRIPTION
Load data	Execution of ETL processes.
QA results of load	Examine log files and content in the interim model.

3.2.8 Verify Model Structure

An evaluation of a populated data model indicated whether it was capable of supporting service location processes. The model needed to be structured to support address range geocoding, one of the core capabilities of the interim model, while respecting other elements of the model. Sample addresses were collected and executed against the database to confirm this capability.

CAD elements in datasets provided by the telecom carrier partner were sampled to determine which features were affected by editing processes that moved assets to improve the presentation of information over its accurate spatial location. Of note here is that while sampling of the CAD elements was completed, content was not available for Ottawa. As a result, a template process was developed to improve content, which could be applied to CAD derived data. As such, service location processes were deemed to be transferable to business processes adopted by the telecom carrier partner.

An export process confirmed that business data met the requirements of disparate users, including CAD modellers and designers. Visualization was used to verify export processes. AutoDesk products LandXplorer and Revit were used to validate results.

Table 6 - Steps to Verify Interim Model

PROCESS	DESCRIPTION
Geocoder configuration	This step examined the requirements for geocoder technology with respect to data structures and data completeness and was used to validate the data model and make revisions needed for geocoding.
Prepare file based sample addresses	Some data and data formats were modified to properly support address information. External datasets were modified to encapsulate address information. These datasets were not part of core data.
Establish that service location points can be supported	Define an interim structure to contain points collected in the field and through processing. Compiled addresses were moved to a modified CityGML address database object.
Confirm model data can be used by applications	Export data for use in desktop applications.

3.2.9 Revise Model

Actions taken in this process included the creation or modification of tables or views, the creation of database objects, such as triggers and sequences, and the removal of unnecessary elements. Changes were made only if they respected the structure of the CityGML and Oracle Geocoder models.

The resulting changes, when invoked, required validation similar to that described in Section 3.2.7.

Table 7 - A Summary of Data and Model Evaluation Processes

PROCESS	DESCRIPTION
Modify database structural components	Add database objects and modify data structures as required.
Modify ETL tools	If table structures change, ETL tools might require modifications.
Adjust data to accommodate structural changes	Data could either be purged and reimported or modified in the database using SQL or stored procedures.

3.3 Service Location Data Collection and Evaluation

Point data were collected through field work and were derived from geocoding and semi-automated processes based on orthoimagery. A civic address identifier was a common element that facilitated an assessment of these strategies. Stratified sampling was used to develop discreet sets of address locations in areas representing different land uses. Different land uses affected parcel densities and parcel configuration and therefore were assumed to influence the effectiveness of strategies to identify service location. The study area was divided into downtown commercial, industrial, suburban, and rural sample areas, with the general area of each sample area identified through author familiarity. A minimum of 50 sample addresses were initially targeted for each area. An adequate sample size ensured that addresses were dispersed on an adequate number of road segment sides to reflect the diversity of configurations common in each sample area. The extent of each sample area was dictated by the coverage of multiples of 1 kilometre x 1 kilometre orthoimagery tiles obtained indirectly from the City of Ottawa through Scholars Geoportal (http://geo1.scholarsportal.info).

A post collection assessment of sample data was undertaken. The accuracy of GPS was used to define thresholds of acceptability. The specifications for the GPS unit used in this research indicated that it was accurate to 3 metres. A vector produced by connecting both points collected manually and points developed through geoprocessing to a peer GPS point, that is a GPS point representing the same location, defined an offset distance. A standard deviation of the offset from the GPS point for both methods was calculated to identify outliers. The deviation together with thresholds defined by GPS was used to categorize address points and assign appropriate qualifiers. The accuracy of each method was tabulated for each sample area. The findings identified the most suitable approach for each type of high level land use category sampled for this paper.

Table 8 - A Descriptive Summary of Service Location Identification Processes

PROCESS	DESCRIPTION
Identify sample areas	Select orthoimagery tiles that cover locations that represent the identified land uses.
Compile sample Addresses	Randomly select addresses within the identified sample areas.
Collect GPS locations for sample addresses	Handheld GPS coordinates were collected for the selected addresses.
Develop semi-automated process for service location	Semi-automated address discovery processes determined if address location was improved, or non-addressable entrance ways were identified.
Geocode sample addresses	Using the selected addresses, execute a geocode process to establish a set of associated coordinates.
Develop parcel level geocoder	A parcel level geocoder used spatial buffer and nearest neighbour spatial functions to find alternative address locations as part of the evaluation. This process served as a proxy for a LOT/BLOCK/PLAN geocoder.
Assess results of service location collection processes	Analyse the results of the collection process using the GPS points as the baseline to establish the accuracy of other processes.

3.4 Data and Model Revisions in Support of a Final Service Location Model

The evaluation of the point data collected by processes outlined in Section 3.3 was then used to extend and modify the interim database model to better support service location. The results of these processes defined a final service location model. These processes not only defined changes in the database model, but also identified issues in the collection of data in the field, which could be used to define policies for data collection.

4 Execution of Data Modelling and Analysis Processes

This section will describe the execution of the steps that were outlined in the methodology. The results from which should satisfy objectives 1) to evaluate, modify, and extend standards-based data models, 2) to evaluate methods to collect, correct, and assess service location information, and 3) to incorporate collected field service information into a revised final service location model. The first listed objective will be partly met by physically implementing an interim 3D geospatial model in support of service location through a number of phases. Modelling will include a review of technologies, data sources, and standards. The interim model will support the collection and assessment of service location data. The last objective will be realized by incorporating the assessment of field work to tune the interim model, with the result being a final 3D geospatial model in support of service location. The execution of the methodology will reveal the challenges and benefits of a proposed open standards based data model.

This section will at first describe the study area for this research and related data, and include a discussion on data sources and data issues. Modelling will be subsequently detailed, with a discussion of standards and technologies. The discussion will be followed by an examination of data collection and validation procedures for geocoding, field work, and semi-automated processes. Finally, a section on the challenges of interacting with Building Information Models will be briefly outlined.

4.1 Study Area and Available Data

Personal familiarity was a prime consideration in choosing Ottawa as a study area for this research. Access to the study area supported data validation. Supplementing secondary data with primary data collected in the field was also an important consideration. The City of Ottawa in Ontario, Canada is a geographically large municipality, a result of amalgamation with municipalities in the Regional Municipality of Ottawa Carleton in 2001 (City of Ottawa, 2012a). The City of Ottawa covers an area of 2,796 square kilometres (City of Ottawa, 2012b). The characteristics of Ottawa have various factors that impact service location. The area is 20% urban and suburban (City of Ottawa, 2012b). The remainder is exurban and rural with a variety of land uses, which includes farms, large estates, corridor development, and small towns and hamlets. Different land uses and population densities result in property configurations that need to be accounted for in service location processes.

The City of Ottawa is officially bilingual (City of Ottawa, 2012c), providing a good test case for maintaining multi-lingual services. In addition, municipal amalgamation introduced a new urban corporate hierarchy that had the potential to affect the way locations were referenced. Data models should declare an official name for a street or municipality, but unofficial names may still be used by the public.

Figure 2 - Extent of study area

Since Ottawa is large, the study area was isolated to the west end of the city, as shown in Figure 2.

The research was heavily dependent on secondary data available through the City of Ottawa Open Data website (City of Ottawa, 2013). After an initial examination of available and relevant data sets, additional data were obtained from the University of Waterloo Map Library and directly by request from the City of Ottawa (Table 9). Both sources requested the acceptance of a signed license agreement.

Data from various sources were not consistent in coordinate reference system, accuracy, and vintage. Since the City of Ottawa was the most common source of data, the coordinate reference system used by the City of Ottawa, Modified Transverse Mercator (MTM), was chosen for the initial default spatial reference system for detailed components of the project. Data to support geocoding made use of a geodetic coordinate reference system, which factored later on in the selection of a coordinate system.

Data from commercial data suppliers, namely Navteq (to be described in detail later on) and DMTI, were examined. Navteq data were used to validate the Oracle Geocoder. DMTI enhanced point of interest (EPOI) data served as an evaluation benchmark for data derived from processes developed during this research. DMTI address point data relevant to this research were downloaded under license from Scholars GeoPortal (http://geo1.scholarsportal.info). DMTI EPOI data contained points were rendered on top of the Ottawa orthoimagery to get an idea of coverage. Coverage was good but incomplete in commercial and industrial areas, but sparse in rural and suburban areas. In commercial and commercial areas, points were not always consistently placed. In some cases, points were located at entranceways, while in other cases points were positioned at building locations.

Data were also obtained from crown and public organizations. Teranet, which is a provincial crown corporation entrusted to manage land registration data for the province of Ontario, provides the framework for municipal property assessments in Ontario (Teranet, 2013). Inquiries were also made regarding optional municipal datasets. Driveway data from the City of Ottawa, available for about $1600 (Steve Perkins, City of Ottawa employee, personal communication), was not used but is noted for future research or implementation. Driveways digitized from orthorectified aerial photography should be more accurate than street geocoding or even parcel level geocoding.

Table 9 highlights the main properties that were considered as part of the data preparation stage of the research.

Table 9 - Data Sources

Data	Source	Type	Date	Projection	Datum	Resolution or Accuracy	Count
Roadways	City of Ottawa (Open Data)	Line vector	Last updated April 2012.	MTM Zone 9	NAD83 GRS 1980 spheroid	+/- 1m	24334
Large Buildings	City of Ottawa (Open Data)	Polygon vector	Updated using 2005 orthoimagery.	MTM Zone 9	NAD83 GRS 1980 spheroid	Based on orthoimagery	14508
Buildings with elevation	City of Ottawa (direct request)	Polygon vector	Updated using 2005 orthoimagery.	MTM Zone 9	NAD83 GRS 1980 spheroid	Based on orthoimagery	62271
Orthoimagery	City of Ottawa through Scholars GeoPortal	Raster	2008	UTM Zone 18	NAD83	0.2 m	14
Lot Concession and Township	City of Ottawa (Open Data)	Polygon vector	Information sourced from documents dating from 1860	MTM Zone 9	NAD83 GRS 1980 spheroid	+/- 1m	3628
Parcel	Teranet	Polygon vector	2009	LCC	NAD83 GRS 1980 spheroid	n/a	238359
Address Points	DMTI	Point vector	2012	Geographic	NAD83 GRS 1980 spheroid	n/a	26424 (clipped for Ottawa)
Sample Road Network	Navteq	Line vector	Unknown	Geographic	WGS 1984	n/a	n/a

4.2 Evaluating CityGML

One of the objectives of this paper is to evaluate, modify, and extend standards-based data models to support service location. CityGML, as an Open Geospatial Consortium (OGC) standard that supports 3D models, was chosen as the initial framework model for this paper. An applied evaluation took into account database structure and content.

4.2.1 Standards

Standards from two organizations were reviewed: the International Organization for Standardization (ISO), and the Open Geospatial Consortium (OGC).

4.2.1.1 ISO

There were a number of ISO standards, itemized in Table 10, that were worth reviewing. None of the published standards claim to be exhaustive so users are able to extend the standards as they see fit. End products cannot, however, claim to be compliant unless the standard is adhered to completely. Three ISO standards were examined in detail: ISO 19138, ISO 19113, and ISO 19115.

Table 10 - A Summary of Relevant ISO Standards

Standard	Description
ISO 19105:2000	Geographic information -- Conformance and testing
ISO 19109	Geographic information -- Rules for application schema
ISO 19111:2007	Geographic information -- Spatial referencing by coordinates
ISO 19111-2:2009	Geographic information -- Spatial referencing by coordinates -- Part 2: Extension for parametric values
ISO 19113:2002	Geographic information -- Quality principles
ISO 19114:2003	Geographic information -- Quality evaluation procedures
ISO 19115	Geographic information – Metadata
ISO 19116:2004	Geographic information -- Positioning services
ISO 19138:2006	Geographic information -- Data Quality Measures

ISO 19138 provides a set of standard data quality measures that were adopted as part of a quality assessment. The standard outlines quality measurements for excessive or missing features and duplicate features. Measures for logical consistency, domain compliance, topological relationships, positional accuracy, and geometric anomalies are outlined in ISO 19138. Thematic and attribute accuracy measures outlined in the standard were considered due to their relevance to semantic modelling. Quality measures were applied at the feature level.

ISO 19113 outlines quality principles related to geographic datasets, including measures that define the fitness for use of datasets for specific applications. The principles in this standard are closely aligned with ISO 19138. This standard provides detail of specific test cases for conformance rather than quality measurements.

ISO 19115 has high visibility in the geospatial world. The standard is composed of metadata sections, which contain metadata entities. The entities contain metadata information in elements, which are analogous to UML attributes. At a high level, ISO 19115 provides a metadata framework for structuring data for source, extent, quality, spatial and temporal dependencies, and distribution information. Metadata cover quality aspects defined in ISO 19138. Metadata for spatial reference systems are also defined, being of primary importance in the exchange of geospatial information.

4.2.1.2 OGC

The Open Geospatial Consortium (OGC) is an active organization that has developed standards to make geospatial information more consistent and available. The OGC represents the interests of commercial vendors, government, non-profit organizations, educational institutions, and individuals (OGC, 2013). The OGC is responsible for defining many web service standards including Web Mapping Services (WMS) and Web Feature Services (WFS). Of interest to this paper is the development of GML and CityGML, which will be described in detail later.

Client requests are submitted to web services, which require parameters in the form of key value pairs to tailor a response. Explicit requests may include getMap (OGC, 2006), getFeature, or getCapabilities (OGC, 2010). A getCapabilities response provides the user with metadata on available web map and feature services.

OGC standards are publicly available documents and can be used without fee or royalty as long as proper attribution of the standard is applied. The full text of the legal obligations for using OGC standards can be found at http://www.opengeospatial.org/ogc/document.

4.2.2 Technologies

Commercial and open source technologies are reviewed in the following subsections.

4.2.2.1 Oracle

Oracle Corporation is the second largest commercial software company in the world (Forbes, n.d.), but does release some open source products (Oracle, n.d.). Many of the open source products now promoted by

Oracle Corporation were obtained as part of its acquisition of Sun Microsystems in 2009, including Java, NetBeans, and MySQL. The core commercial product for the company is its relational database management system (RDBMS) product, called Oracle. The last few major releases have introduced object-relational concepts, which are applied substantively to its Oracle Spatial product and the Oracle Geocoder module. Oracle RDBMS is commonly used in conjunction with GE Smallworld, a dominant GIS vendor in the utilities and telecommunications sector.

NetBeans, JDeveloper, and SQL Developer are Integrated Development Environment (IDE) applications. NetBeans and JDeveloper primarily support Java development, but can also be used to create, modify, and validate XML and other common web documents and applications. JDeveloper has a useful data modelling tool that supports UML. SQL Developer is intended for developing database components. Both JDeveloper and SQL Developer are available for free for non-commercial use, while a free stripped down version of NetBeans is available.

For this research, Oracle 11g R2 (version 11.2.0.3.0) on a Windows 2008 64-bit operating system was used.

4.2.2.2 Safe Software

Safe Software is the developer of FME, which is an acronym for Feature Manipulation Engine. FME is the de facto standard for spatial Extract, Transform, and Load (ETL), which is a commonly used term in data conversion projects. FME is not open source, but it does support a number of open standards. The FME workbench is a product that allows users to develop ETL applications using a visual environment. There are three main elements to the Workbench: Readers, which read data, Writers which write data, and Transformers, which manipulate data. Transformers have input and output ports, which are used to pipe data through the transformation process. Transformers can be chained to produce a required geometric feature with attribution. Drag and drop functionality enables users to define readers and writers for many spatial formats. Since the Workbench is visual, processes developed with the tool are inherently self-descriptive. Parameters within the workbench can be assigned externally in a script or batch file, allowing production teams to process large quantities of data for similarly structured data sets.

Custom feature types can be created and imported into a Feature Manipulation Engine (FME) workbench once a writer is defined. This capability is valuable for CityGML files. A feature type once defined can be universally available by placing the template in the standard home directory for FME types (Safe Software, 2011b).

4.2.2.3 Autodesk

Autodesk is a primary developer of CAD software. Autodesk has transcended from CAD and flagship product AutoCAD, and has emerged as a developer and leader in BIM software, the next generation of architecture and engineering software. Relevant to this study, a free version of their LandXplorer CityGML Viewer provided enough functionality to visually validate CityGML files. Revit is a more sophisticated Autodesk offering that supports BIM and Industrial Foundation Class (IFC) files (Bokmiller, Whitbread, & Londenberg, 2011).

AutoDesk supports collaborative work through their Revit Server product (Bokmiller, Whitbread, & Londenberg, 2011). Revit Server, and similar CAD offerings, might challenge the need for GIS integration and the need to adopt open standards. A strong business case must be prepared to demonstrate that openness is a way to extend functionality and not undermine the autonomy of established business units.

4.2.2.4 OpenLayers

OpenLayers is a freely available JavaScript library that can be used to incorporate 2D maps and georeferenced imagery in a web page. Since OpenLayers is JavaScript, it functions strictly within the client web browser. OpenLayers has no web server capabilities, but can consume and request OGC compliant services.

4.2.2.5 OpenStreetMap

OpenStreetMap is a base map service that is delivered as a WMS that can be consumed by web or desktop applications. OpenStreetMap layers can also be downloaded (OpenStreetMap, 2012) and stored locally, which eliminates reliance on OpenStreetMap servers. There are no royalty fees and licensing is liberal (OpenStreetMap, n.d.). The quality of the OpenStreetMap base map varies since volunteered information forms the basis of content. More importantly, there appears to be no systematic process of content compilation, at least from a traditional mapping perspective, resulting in inconsistent coverage. Base maps derived from imagery are limited.

OpenStreetMap layers can be integrated into a Postgres/PostGIS database using utilities such as Osmosis (OpenStreetMap, 2013).

4.2.2.6 MapServer and Apache

MapServer is middleware application, which runs within the Apache web server environment. Middleware is software that operates between server processes and client applications (NATO, 1969), MapServer is an open source solution that delivers spatial content using OGC compliant services, such as WMS, WFS, or WCS. ESRI Shapefiles are a commonly used format for data sources. MapServer supports Oracle Spatial, but some configuration is required.

4.2.2.7 BIMServer

BIMserver is an open source solution that facilitates the delivery and exchange of IFC and CityGML data. BIMserver was not on the radar at the beginning of the research, but was explored once limitations on the exchange of IFC and CityGML formats from other offerings were realized.

4.2.2.8 ESRI

ESRI is the dominant vendor in GIS. Their flagship product, ArcGIS, figures prominently in many sectors. ArcGIS is enhanced through the use of extensions, which provide business specific functionality. ArcGIS is capable of reading Oracle Spatial, GML, and CityGML through the Data Interoperability Extension.

ArcScene is a 3D visualization and analysis tool within ArcGIS Desktop products. ArcScene is capable of rendering CityGML files.

4.2.3 Data Structures

Published and publicly available spatial data models, although extensive, cannot meet the needs of all business requirements for all sectors. As such, spatial models should be considered as a component of an integrated target business model, one that Enterprise Resource Planning (ERP) systems for example can embrace. If treated as a component or module of a system, then the opinion of this author is that the adoption of spatial framework would be less intrusive to established systems.

The CityGML standard was explored with some comparison to IFC. CityGML arose out of the Special Interest Group 3D (SIG 3D) of the Geodata Infrastructure North-Rhine Westphalia (GDI NRW) in Germany (Gröger & Plümer, 2012). IFC came from a similar desire to develop open standards to improve the exchange of data (BuildingSmart, 2013). The Building Interior Space Data Model (BISDM) promoted by ESRI is noted but was not examined.

Some elements in IFC and CityGML overlap, which is desirable for data exchange. Before a discussion can be completed, an overview of frameworks on which these models are built is required.

4.2.3.1 XML

XML is a vendor independent document format that supports distributed computing and information exchange between web services and applications (Salminen & Tompa, 2011). XML is designed to encapsulate “complete information about the content” and supports “re-use of information resources for multiple purposes” (Salminen & Tompa, 2011, p. 8). An XML element contains metadata and content, and elements can be nested within other elements, resulting in a hierarchical document structure (Salminen & Tompa, 2011). XML works well in heterogeneous environments since it is platform independent and supports a variety of encoding standards (W3C, 2006). XML can be transformed through Extensible Stylesheet Language Transformations (XSLT) to create new XML or other formats, such as text and HTML

(W3C, 2010). This transformation process is useful when complex XML needs to be simplified to accommodate specific tasks. The eXtensible Stylesheet Language (XSL) is a complementary document that, with Cascading Style Sheets (CSS), can be used to improve the rendering of XML (Salminen & Tompa, 2011). Finally, specific data can be extracted from XML using the XQuery language (Salminen & Tompa, 2011).

The structure of XML is enforced through the use of one or more referenced document definition languages, such as Document Type Definition (DTD) or XML Schema Definition (XSD). These definition languages constrain data and provide validation rules for the XML processor, which prepares content for a target application, although XSD and DTD focus on different aspects of validation (Salminen & Tompa, 2011). In either case, these languages provide a framework that supports the semantic intentions of data. XSD appears to be similar to XML but is not intended for content storage, as would be expected of a document that defines structure and constraints.

XSD has some improvements over DTD. XSD supports namespaces, which reduce ambiguity when different schemas are referenced within an XML document (Salminen & Tompa, 2011). A namespace is a prefix identifier that is applied within XML elements. Schema documents, in particular, also improve searches against the XML document (Salminen & Tompa, 2011).

XML is based on readable text, for example ASCII or Unicode (W3C, 2006). Considering that document structure and metadata are embedded in XML, file size is generally larger than a similar file in binary format and as such file size could impact performance (Microsoft, 2012). XML documents also require more processing for rendering (Salminen & Tompa, 2011). For this reason, any proposed solution should determine when best to use XML for content storage and delivery.

Why is an understanding of XML important? Since XML is a framework document structure for GML and CityGML, extending these structures requires at least a cursory understanding of XML. XML is also used to configure applications. Occasionally, XML fragments are used within applications.

4.2.3.2 GML

Geography Markup Language (GML) supports the exchange of spatial information. GML is an implementation of ISO 19107, which defines 3D geographic representations (Kolbe, König, Nagel, & Stadler, 2009; OGC, 2012). GML is a structured response to a WFS request. GML can be parsed and formatted using many processes available to XML. Geographic content embedded in GML can be rendered as a map overlay or presented as a textual element in a web page.

4.2.3.3 CityGML

CityGML, an extension of GML, is capable of storing 3D representations of objects in a geospatial context. CityGML supports five different Levels of Detail (LoD), with each level directly correlated to data accuracy. (Kolbe, Gröger, & Plümer, 2005). An extended CityGML model should be able to support surface, subsurface, and overhead telecommunications and utility hardware. Furthermore, telecommunications projects can take advantage of CityGML capabilities that "enable 3D visualizations and facilitate localization in indoor and outdoor navigation" (Kolbe, Gröger, & Plümer, 2005, p. 1), and data management and integration (Gröger & Plümer, 2012). CityGML features are organized into thematic modules (e.g. buildings) which can be combined into profiles (Gröger & Plümer, 2012), and an appreciation of this hierarchy contributes to an understanding of CityGML UML diagrams.

There are a few items one must consider to effectively use CityGML. Although the authors of CityGML recognized the need to support OGC web services (Kolbe, 2007), a model of the built and natural environment must also support requests submitted by applications that can only efficiently consume 2D representations of features, which is typical of many mapping applications. In addition, CityGML is based on XML and thereby inherits the same limitations, as well as the benefits. If one does not respect the overhead inherent in CityGML, a file can become so large that it is no longer a workable format. File size is a limitation (Morgan, 2009), which is further aggravated with domain extensions (El-Mekawy, Östman, & Hijazi, 2011).

4.2.3.4 Opportunities for CityGML in the Database

The authors of CityGML developed a database model called 3DCityDB. Data definition language (ddl) scripts made available by the authors can be used to create a database schema that serves as a repository of CityGML files (Institute for Geodesy and Geoinformation Science, nd). An open source CityGML import/export utility (Institute for Geodesy and Geoinformation Science, nd) supports data exchange. The utility also recognizes Keyhole Markup Language (KML) and COLLADA file formats. COLLADA is an acronym for Collaborative Design Activity (Collada.org, nd).

This research took advantage of an instantiation of the CityGML schema to develop an understanding of CityGML, and to extract and analyse feature level and aggregate data using SQL and Oracle Spatial functions and operations. The top object in the CityGML model is *ObjectClass*, which is a schema level data dictionary table for other classes. The *ObjectClass* is a simple table that has an ID, class name, and an entry for the super class for all other CityGML object classes. The first child of *ObjectClass* is *CityObject*, which is an abstract class. The *CityObject* table contains metadata inherited from GML, including name and description, as well as other details, such as creation, validation, and retire dates, that are related to data management processes (Gröger & Plümer, 2012). If there are modifications to the underlying geometry of an object, the content in the LINEAGE column in the *CityObject* table should reflect these changes. Based on observed behaviour, an object that undergoes changes requires a new entry in the *ObjectClass* table.

All instantiated spatial objects reference the *CityObject* table by way of the ID of the respective object table (e.g. *Building*). The number of records in the *CityObject* table should match the number of records for all the active instantiated spatial objects in the schema. The ENVELOPE field on the *CityObject* table defines the extent of a valid data model (Gröger & Plümer, 2012). Terrain Intersection Curves (TIC) on some objects, such as *Building*, correspond to level of detail and ensure that a continuous surface exists with intersecting objects (Gröger & Plümer, 2012) that might not be at the same level of detail. TICs prevent objects "from floating over or sinking into the terrain" (OGC, 2012, p. 9).

In general, the geometry of an object, e.g. walls, that intersects with a terrain surface is maintained in the *SurfaceGeometry* table and not within the namesake object. The geometry for each object is referenced through an identifier that corresponds to level of detail. Exceptions to the use of *SurfaceGeometry* for geometric storage are LOD0 features, which exist on *Transportation_Complex*. In CityGML 2.0, the *Building* object has an LOD0 representation (OGC, 2012), which should be able to simulate 2D content requirements and avoid the need to create a separate building footprint or building roofline object.

AbstractFeature is a direct child of *ObjectClass*. Child classes of *AbstractFeature* do not contain geometry. Of all the non-geometry classes, the *Address* table is of interest to service location. The independence of *Address* from any class containing geometry presents an opportunity to apply business concepts and add business data before geometry classes are prepared.

Established relationships in CityGML attempt to conceptually model object behaviour; for example, a building cannot be contained in a room, but a room must be contained in a building. The model is flexible enough to handle exceptions, for example, when the "topological correctness of the boundary cannot be guaranteed", an object can be defined as a *MultiSurface* object (OGC, 2012, p. 74). Orientation of an object, such as whether a wall represents an interior or exterior surface, is critical to ensure proper rendering of photo-realistic appearances (OGC, 2012).

There are common attributes across a number of objects. Relevant to this paper, CLASS, FUNCTION, and USAGE need to be managed on *Building, City_Furniture, Building_Installation, Landuse, Room, Transportation_Complex*, and *Traffic_Area*. The values for these fields are entity specific. Appropriate values for these fields are listed in Annex C of the CityGML 2.0 standard (OGC, 2012).

If an external source of data is loaded into a table and if the external fields have no target, then attribute content is loaded into a *CityObject_GenericAttrib* object. Process and data gaps should be eliminated so that content does not end up in this table.

CityGML does not store solid geometry (3D) or volumetric objects explicitly; it breaks down components into separate rows. As a result, if one desires to exploit Oracle Spatial solid geometry functions, a process would have to re-compose the objects.

4.2.3.5 Industry Foundation Classes

Industry Foundation Classes (IFC), which is copyrighted work of BuildSmart International Limited, is an open standard in support of BIMs. IFC covers a number of domains that correspond to building lifecycle management, including, for example, Building Controls, HVAC, Construction Management, and Architecture. The Electrical domain (ifcElectricalDomain) has the potential to model cable carrier fittings, junction boxes, and switching devices with cable routers and network cabling being even more abstractly defined. As of Release 4, IFC does not explicitly model telecommunications equipment (BuildingSmart, 2013).

4.2.3.6 Differences between IFC and CityGML

There are over 900 classes in IFC that define different aspects of engineering and architectural models (El-Mekawy, Östman, & Hijazi, 2011; Van Berlo & de Laat, 2011), but by 2011 only 60 to 70 of these directly relate to GIS and more specifically have a relationship to elements in CityGML version 1.0 (El-Mekawy et al., 2011 referencing van Berlo, 2009). The differences in objects and entities present a challenge in the integration of CAD and GIS processes that use a common model. The loss of information being one concern (El-Mekawy, Östman, & Hijazi, 2011). One notable difference was that IFC explicitly defines a *BuildingFloor* object, while with CityGML model, a floor is an attribute of a building (El-Mekawy, Östman, & Hijazi, 2011). Some objects that are solid objects in IFC are decomposed into sub-components in CityGML. For example, an IFC wall is a solid feature, but CityGML breaks the feature down into interior and exterior walls (El-Mekawy, Östman, & Hijazi, 2011). This can result in a many-to-many relationship in elements that span multiple storeys (El-Mekawy, Östman, & Hijazi, 2011).

A new version of CityGML, version 2.0, was released in March 2012 (Gröger & Plümer, 2012). Data mapping based on the new standard should identify cases where the retention of data has been improved. For example, OUTERCEILINGSURFACE and OUTERFLOORSURFACE features have been added to the version 2.0 of CityGML (Gröger & Plümer, 2012).

Annex F of the CityGML 2.0 specification outlines enhancements and additions to the standard (OGC, 2012).

4.2.3.7 An Overview of Layers in Support of Service Location

The research attempted to map data to a CityGML profile, while accommodating the requirements of Oracle Spatial and other business processes. Some CityGML profiles were ignored. Surface datasets, such as

DTMs, TINs, and spatially referenced images, part of the relief module (Gröger & Plümer, 2012), were not used to create an explicit surface but should be considered in a real world implementation to provide context. DTMs and TINS could also be used as a source of baseline vertical data to enhance 2D datasets.

The initial logical model was developed in tandem with an analysis of data requirements.

4.2.3.7.1 Municipalities and Provinces

Municipalities and other jurisdictions are handled differently in CityGML and by the Oracle Geocoder. Administrative area names are embedded in an *Address* record within the CityGML schema, while in the Oracle Geocoder, areas are stored in a corresponding area table, the structure of which will be described in detail in Section 4.1.6.1.

Amalgamation, annexation, and demalgamation are municipal business processes that need to be considered as part of the support of addresses and service location. Although cadastral data in Ontario reference properties as lots contained in a larger block and plan, property parcels reference a civic address that are associated with an administrative unit. Data from for this research from Teranet contained a number of administrative names that have an historical and local context. A query on the imported data revealed 35 different combinations for the City of Ottawa. 'KANATA', which is a pre-amalgamation municipality, or 'GOULBOURN/STITTSVILLE', which is a combination of an old township name with a locality, were identified sample entries. These names needed to be mapped to the official name for the City of Ottawa. These original place names were entered into the *Municipality* table using an SQL statement and assigned to a current official municipality and assigned a retired date. These entries were needed to support the bulk import of address information associated with a parcel.

Municipalities serve as a reference for street networks, and in Canada municipalities are the authority for street names (Canada Post, 2012). A municipality object in the database is essential for geocoding and location identification. With this in mind, this paper proposes that both the geometry and the attributes of the municipality need to be maintained.

4.2.3.7.2 Streets and Networks

Street network data are built from line segments that typically represent street centrelines. In a number of commonly known datasets, street name, type, and direction are isolated in separate fields. Some data sources do not separate any street name components and as a result require more attention. The nature of road centrelines, as a boundary for blocks of residential units or as boundaries between municipalities, require separate fields to accommodate variations in postal codes, municipality names, or even street names as managed by different municipalities. TIGER files demonstrate this approach by storing a right and left

ZIP code on each segment. Sophisticated data sets provide options to store types and directions before the street name.

Street network data obtained from the City of Ottawa followed the name/type/direction nomenclature. As part of data preparation, a distinct set of types and directions were placed in look up tables. Typically, in the French language a street type is a prefix whereas in English the type would qualify as a suffix. Canada Post (2005b) represents the prefixes and suffixes as codes that point to a reference table containing variations of a type or direction. An official street name can be assembled on-the-fly depending on the language of choice. For example, a person with an English mother tongue would prefer 'Richmond Road' whereas a person with a French mother tongue would expect 'Chemin Richmond'. Canada Post (2005b) provides alternative names in cases where the street name has another locally recognized name.

The tabular representation of street data was prepared using the `POPULATECITY` stored procedure, which is in Appendix A. Some post processing was still required after executing the procedure, for example when a standard street direction code was missing from the direction cross reference table. This gap in the original processing of street data was easily corrected with an update statement.

Roads and streets are almost identical in their respective meanings and the terms have been used interchangeably in this document.

4.2.3.7.3 Land, Parcels, and Buildings

Cadastral data, or land parcels, and dedicated right-of-ways are valuable layers in support of service location. They can be used to establish a relationship between road network and buildings and provide context for other telecommunications and utility assets. Transportation and utility networks and some elements of telecommunications infrastructure function within right-of-ways, while buildings do (should) not. Therefore, the separation of these entities is necessary from a conceptual and functional level.

In Ontario, data on over 4 million parcels are managed by Teranet (Ontario Parcel, n.d. c). Land ownership, property assessments, and crown land are managed as separate data sets (Ontario Parcel, n.d. b) as they support different business functions. British Columbia is in the process of assembling a seamless cadastral data set, which currently includes approximately 1.7 million parcels of private and crown land (GeoBC, 2013). Sample data for Ontario obtained for this paper do not spatially differentiate between assessments and ownership, but there are differences in attribution. The main concept to consider is that ownership might apply to a parcel that has different functions or uses, and therefore different assessments. The Assessment Role Number (ARN) is the link to these different assessments. The other factor to consider is that condominiums have many individual ownership parcels, basically the inner living space, within a parcel owned by a condominium board, but these relationships are also not represented in the spatial data. As the data are 2D, it would be difficult to properly convey multi-storey ownership parcels.

Since parcel data is managed by Teranet at the Ontario provincial level, data should be easily available for other municipalities within the province. Data are available at no cost to utilities if the project is "not retained by the utility company, and is for the benefit of the municipality" (Ontario Parcel, n.d. a). Until such time that data are universally available, there will be a need to extract information from various sources that are not as extensive and have varying degrees of quality and accuracy.

The *Landuse* object in the CityGML schema can be used to model parcels (Kolbe, König, Nagel, & Stadler, 2009), but the ownership of condominium properties would eliminate *Landuse* as a candidate placeholder for parcel information. There are two reason for taking that position. First, high density residential landuse classifications are useful in modelling aggregate residential units as classes, but an explicit residential unit inside a building is less conventional; mixing cadastral information with other forms of land use classes when the focus should be on service location was considered unnecessary overhead. Second, the *Landuse* object in CityGML does not directly relate to a building, but to a *Surface_Geometry* and a direct explicit relationship is preferred.

The core CityGML schema can be augmented to support cadastral information through Application Development Extensions (ADEs), such as the one developed for Immovable Property Taxation (CityGMLWiki, 2012). ADEs are XML schemas that can be directly referenced by a CityGML document. Modification of the database schema would need to consistently reflect ADE requirements if adopted.

Buildings are self-contained structures that belong to a parcel. The approach for the proposed model is to assume that not all buildings are associated with a road. For example, in Canada, many cottages are on islands that are only accessible by boat. All buildings, however, are in the cadastre. It is through the cadastral framework that buildings are able to establish adjacency, rather than just proximity, to right-of-ways, waterbodies, and to other parcels. Service provision on large tracks of land, within transmission or transportation corridors, or on islands would benefit from the parcel fabric. Mobile users can then be associated with a geographic location regardless of their movements.

The benefit of modelling buildings in 3D, as opposed to just parcels or 2D buildings, is that interior service locations and associated paths can be identified. Buildings intersect a surface at the footprint and vertically extend from the footprint to occupy space above and/or below ground level. Building protrusions and insets alter the vertical profile. The Ontario College of Art and Design in Toronto is a prime example of how a vertical profile deviates from its footprint, and in CityGML would likely be represented as a *Building* with *BuildingPart* features (OGC, 2012). A *BuildingPart* may also be used to model condominiums.

Underground cities, as they exist in Montreal and Toronto, can be assigned properties that define measured height or stories above and below ground level. These cases demonstrate how information systems based on conventional large scale 2D mapping are functionally constrained.

In architectural and CAD drawings, structures are more detailed but have arbitrary coordinate origins, and even floors within buildings might have different origins (Morgan, 2009). In CityGML and Building Information Models, buildings are composite geometries that are spatially referenced and can optionally be composed of multiple building parts. In CityGML, a *CITYOBJECTGROUP* object assembles objects into a single group. Assigned roles to components in a cluster can support specific business functionality, such as a "sequence number … in an escape route" (OGC, 2012, p. 15).

Three dimensional objects can be created using extrusion methods in Oracle Spatial, ArcGIS and FME based on a height above ground level if available. Alternatively, one can use architectural software, such as Autodesk Revit, to manually add 3D content. If data were in CityGML format, one could go through the torturous task of manually adding a third dimension to XML content, or use XSLT as a more simplified option.

Oracle Spatial requires that 3D data be assigned a three-dimensional coordinate system. If a 3D Spatial Reference Identifier (SRID), which represents a coordinate system, and metadata are not defined, validation processes on data will generate errors. Data with validation errors cannot be reliably used within some Oracle Spatial functions. To determine the limitations of projected data in Oracle Spatial, data in a Modified Transverse Mercator (MTM) projection were extruded, or assigned a vertical dimension, then loaded into Oracle Spatial. When exported from Oracle Spatial, the data were not apparently affected, so Oracle Spatial can be used as a repository, even if the data are invalid in the database.

4.2.3.7.4 Parcel Data Analysis and Preparation

A basic FME workbench with a ShapeFile reader, a REPROJECTOR transformer and an Oracle Spatial writer loaded Teranet parcel data into a staging table called *Ottawa_Parcels*. The parcel data contained lot description and street address information, in addition to geometry. Street name components (name/type/direction) were contained in a single field. Variations in the application of street types and directions made it difficult to match against the street names on the City of Ottawa road network. This was further complicated by the inconsistent application of types and directions on streets. The following sample query extracted street name elements from the parcel data for Melbourne Avenue, with results of the query are shown in Table 11:

```
SELECT DISTINCT SUBSTR(street, 0, INSTR(street, ' ', -1, 1)),
       TRIM (SUBSTR(street, INSTR(street, ' ', -1, 1))) street
  FROM ottawa_parcels
 WHERE street LIKE '%MELBOURNE%';
```

Table 11 - Name and Type Variations

Street Name	Street Type
	MELBOURNE
MELBOURNE	AVE
MELBOURNE	STREET
MELBOURNE	AVENUE

To complicate matters, further sampling of the dataset revealed that types can appear to be part of the street name (e.g. 'CEDAR LANE TERRACE'). Variations in street type illustrate the risk in accepting data from various sources. The approach taken for this paper was to develop a filtering process followed by visual confirmation.

A topological adjacency relationship could be used to assign road information from the right-of-way to parcel data (Doug Dudycha, personal communication, May 1, 2013). This step though would not easily account for corner lots, lots that have different roads on the front and back lot line, and even some cases where a property is bound by three or more roads, which does occur more often in rural areas. In these minority cases, the correct road cannot be assigned without relying on additional information, such as the street name on the parcel. Considering these cases, a logical relationship between parcel addresses and city streets was adopted.

Data tables for known street types and directions were prepared to standardize street name elements. Street direction, typically at the end of the full street name, was extracted and removed from a copy of the street name and placed into a new street direction column on the *Ottawa_Parcels* table. The altered street name was also placed in a new column, allowing the original name to be used for comparison and quality assurance purposes. Street type, if present, was likewise processed but it used and modified the content of the new street name field. The stored procedure that accomplished this task is highlighted in Appendix B.

The results were iteratively reviewed to identify abnormalities in the way types and directions were extracted. The process had to be careful to not remove legitimate parts of a name. For example, when replacing the street direction from 'BELL ST S', only the last occurrence of 'S' needed to be removed, otherwise the result would be 'BELLT'. Properly defined regular expressions in SQL can be used to avoid this problem. A regular expression can be constrained so that it only replaces the *nth* occurrence of a pattern, which can be identified using the `REGEXP_COUNT` function (Oracle, 2010b):

```
SELECT TRIM(NVL((SELECT REGEXP_REPLACE(name, rd.dir_en, '', 1,
      REGEXP_COUNT(name, rd.dir_en))
  FROM street_directions rd
```

```
  WHERE TRIM(SUBSTR(r1.name, INSTR(r1.name, ' ', -1, 1))) =
        rd.dir_en), name)) x,
       (SELECT REGEXP_REPLACE(name, SUBSTR(r1.name, 0, INSTR(r1.name, '
', -1, 1)),'')
          FROM street_directions rd
         WHERE TRIM(SUBSTR(r1.name, INSTR(r1.name, ' ', -1, 1))) =
rd.dir_en) y,
       name
  FROM street r1
 WHERE name like 'BELL%'
```

Once street names were prepared, the street name, type, and direction could be used to match against the street network. Type and direction IDs were assigned to parcels using the same concept as applied earlier to roads and an explicit update was executed:

```
UPDATE ottawa_parcels p
   SET type_id = NVL((SELECT formal_id
                        FROM street_types
                       WHERE type_en = p.street_type), 1)
 WHERE type_id is null;

 UPDATE ottawa_parcels p
   SET dir_id = NVL((SELECT formal_id
                       FROM street_directions
                      WHERE dir_en = p.street_dir), 1)
 WHERE dir_id is null;

UPDATE ottawa_parcels p
   SET street_id = (SELECT street_id
                     FROM road r
                    WHERE r.base_name = p.new_streetname
                      AND type_suffix_id = p.type_id
                      AND dir_suffix_id = dir_id)
 WHERE street IS NOT NULL
   AND street_id IS NULL;
```

The assignment of a street ID on the parcel confirmed that parcels and roads were referencing the same street details. The number '1' in the code represented a null value for both type and direction.

Of the 184799 parcels that had a street name, 168732 received a street ID that was originally assigned to the road network, resulting in a linkage between parcel and road. With limited preparation, over 91% of the parcel addresses matched the road network. Although a 100% match rate was preferable, this study was only concerned with parcels that fell within the previously defined sample areas. A manual process resolved the remaining road name issues within the sample areas. A process to identify potential unlinked roads made use of the following two queries for the industrial sample area:

```
SELECT distinct r.name
  FROM road r, sample_areas ta, address_range ar
 WHERE SDO_ANYINTERACT(ar.geometry, ta.geometry) = 'TRUE'
   AND ar.street_id = r.street_id
   AND ta.name = 'Industrial';
```

```
SELECT distinct r.name
  FROM ottawa_parcels r, sample_areas ta
 WHERE SDO_ANYINTERACT(r.geometry, ta.geometry) = 'TRUE'
   AND r.street_id is null
   AND ta.name = 'Industrial';
```

The first query generated a list of 32 names. The second query identified nine parcel street names within the same sample area that did not have a matching ID. Some of the names were obvious typographic mistakes and needed manual effort to establish a link. Some highlighted the need for name aliases, and that an additional means of linking street names was necessary. Table 12 provides a sample cross reference of the difference in the two sources, even after the initial preparation.

Table 12 - Variations in Street Names between Parcel and Road Networks

Parcel Street Name	Road Network Street Name
MOIRA COURT	MORIA CRT
CITIPLAVE DRIVE	CITIPLACE DR
CONCOURSE GATE	CONCOURSE GT
VIEWMOUNT ROAD	VIEWMOUNT DR
MOIRA CT	MORIA CRT
COACH HOUSE GATE	COACH HOUSE GT
ASSINIBONE DRIVE	ASSINIBOINE DR
HIGHWAY 16	PRINCE OF WALES DR

Some manual correction was needed for all sample areas. The differences between representations of street name highlighted that a data dictionary for standard items, such as street types, and directions, is necessary when data are first entered at call centers, in the field, or in engineering services. A corresponding cross reference table to reduce variations of these items is a likely requirement.

Once the records for the suburban sample area were cleaned, an initial geocoding process was invoked.

4.2.3.8 Address Considerations

The representation of address information in a database can be moderately complicated. In Canada, mail delivery is facilitated through the use of postal codes, which are used by electronic equipment to automatically bulk sort mail (Canada Post, 2013). The postal code is composed of two parts: Forward Sortation Area (FSA) and Local Delivery Unit (LDU). FSAs are containers for LDUs.

Addresses can be civic, which is identifiable from a blockface, or mailing, such as a post office box. In the latter, the mailing address, where mail is picked up, is not even physically in the same location as the dwelling or business. The same applies to General Delivery (GD) Boxes and Community Mail Boxes (CMBs), which are more commonly known as super mailboxes (Canada Post, 2005a). Mail can also be addressed to Large Volume Receivers (LVRs), which are postal clients that receive substantial mail and therefore are assigned their own postal code (Canada Post, 2005a). The characteristics of mailing address in isolation are not essential for service location and are not implemented explicitly for this paper, but an enterprise system will need to establish links between mailing and civic address.

With regard to service location and the challenge faced by geography, existing solutions have functional limitations and as such need evaluation. Address range geocoding, while used extensively, does occasionally reveal spatial data quality and completeness issues (Murray, Grubesic, Wei, & Mack, 2011). Street network dataset addresses are heavily dependent on offset distances that are calculated from an interpolation of address ranges. Interpolation assumes that addresses are evenly spaced and assigned in a logical order with consistent parity. Consistency, however, cannot be guaranteed. Gaps in street numbers or the irregular application of a sequence of civic numbers is not uncommon (Doug Dudycha, personal communication).

Murray et al. (2011) outline the issues with probabilistic and deterministic geocoding as part of their advocacy for hybrid geocoding. As part of their study, successful geocoding did not necessarily translate to spatial accuracy, although it is implied. Murray et al. (2011) correctly note that geocoding needs to be confirmed through supplemental information. Spatial accuracy can be improved through parcel level geocoding, but parcel geocoding could have inconsistencies in the way addresses are textually structured with parcel data (Murray, Grubesic, Wei, & Mack, 2011) as verified earlier in the assessment of Ontario parcel data. In addition, the availability and consistency of cadastral coverage is a concern (Murray, Grubesic, Wei, & Mack, 2011).

4.2.3.8.1 CityGML Addresses

Addresses in the CityGML schema are stored in a single record. Buildings "may be assigned zero or more addresses" with exact positions assigned through an associated address record containing a multipoint feature (Gröger, Kolbe, & Czerwinski, 2007, p. 46; OGC, 2012). Considering the previous discussion on street name components, a partially implemented model should remove the addressing components into a separate table. Multiple addresses can be represented in a CityGML file structure through the use of links to external files or through a WFS (OGC, 2012), so removing street name content from the *Address* table is not without precedent.

The CityGML schema supports the relationship between *Address* and *Building* through the *Address_To_Building* table. Many-to-many relationships represent configurations that might occur in campus-like settings for example.

4.2.3.8.2 OASIS xAL

Addresses in CityGML support the Organization for the Advancement of Structured Information Standards (OASIS) extensible Address Language (xAL) standard (Kolbe, König, Nagel, & Stadler, 2009) developed through ISO TC211 (OASIS, 2012b). OASIS xAL (OASIS, 2012a) is a standard for consolidating various address components from over 200 countries. OASIS has made available a schema document to help achieve conformance (OASIS, 2012c). OASIS xAL is also used by Google Maps as an AddressDetail, but not as an address (Google, 2012). The specification is not used by Google Earth (Google, 2013).

In FME, xAL address elements that are contained in the Writer Type XML document are not visually presented in the transformer. The xAL data used by the transformer is an XML fragment (ESRI, nd). This means that there is no direct explicit mapping between input street address components, such as street name, street type, street direction, municipality, province, country, and postal code in FME. A user must correctly define an XML fragment and trust that FME will understand it.

The process to build a XML fragment (Safe Software, 2011a) requires FME specific functions that accept address inputs into a transformer. The user must define a way to retain the namespace references in a fragment that would not violate the XSD. Using a prepared CityGML document as a template, FME will only represent the address generically as a single element in a transformer. The user must define the XML template code that will pull the address components from the source data to populate a tabular address that conforms to the xAL structure.

4.2.3.8.3 ESRI Addresses

A telecommunications data model produced by ESRI (2012) includes a location entity that contains address information. The telecommunications model contains all the address information in one tuple of the *Location* table. This might not be sufficient from a data structure standpoint, but is not dissimilar from the treatment of addresses in CityGML.

4.2.3.8.4 Summary of Address Models

Table 13 below provides a comparison of various address standards that were discussed earlier. The main consideration with the address component of the model was that it must support geocoding and that it could act as a bridge between parcels and road data. Mailing address components, which are important for billing information and other business functions, were not critical for the support of service location. Based on a service location requirement, the first ten rows of Table 13 were included into the model. The best candidate standard was determined to be OASIS aXL.

Table 13 - Address Elements supported by Different Standards

Element	CityGML	OASIS xAL	ESRI	Oracle
Civic Number	✓	✓	✓	✓
Apartment#	✓	✓	✓	
Floor	indirectly		✓	
PO Box	✓	✓		
Prefix	possible	✓	✓	✓
Name	✓	✓	✓	✓
Suffix	possible	✓	✓	✓
Municipality	✓	✓	✓	✓
State/Province	✓	✓	✓	✓
Country	✓	✓		✓
Postal Code	✓	✓	✓	✓
Delivery Routes		✓		
Status	possible	✓	possible	possible
Valid Timeframe		✓		
LVR		✓		

Source: partly sourced from OASIS (2012c).

4.2.4 Initial Data Model and Core Data Objects

A local implementation of the CityGML schema was generated within Oracle. An initial logical model is presented in Figure 3. The CityGML entities that were the focus of this research are shown in dark green on the left side. Preliminary tables added to the model are shown in the center of the figure. Oracle tables that support geocoding are in orange on the right. The initial model for all intents and purposes is a hybrid model, one that establishes a framework for 3D geocoding.

Preliminary tables were added to support incoming data as well as to act as a bridge between the Oracle Geocoder data tables and CityGML tables that are directly relevant to service location. The metadata tables used by the Oracle Geocoder database application direct geocoding queries to the appropriate set of tables.

Geocoding in Oracle relies on the existence of seven data tables and three metadata tables (Oracle, 2011). Data used to support geocoding can be purchased or prepared as part of an internal process. Oracle strongly advocates that users obtain data from a known data distributor for use with their product (Oracle, 2011). From an operational perspective, obtaining vendor data is good advice. The data and relationships between the tables that represent different aspects of geocoding are time consuming and difficult to develop and validate. Regardless of source, the geocoding data structures need to be created and content needs to be imported before the Oracle Geocoder can be functional.

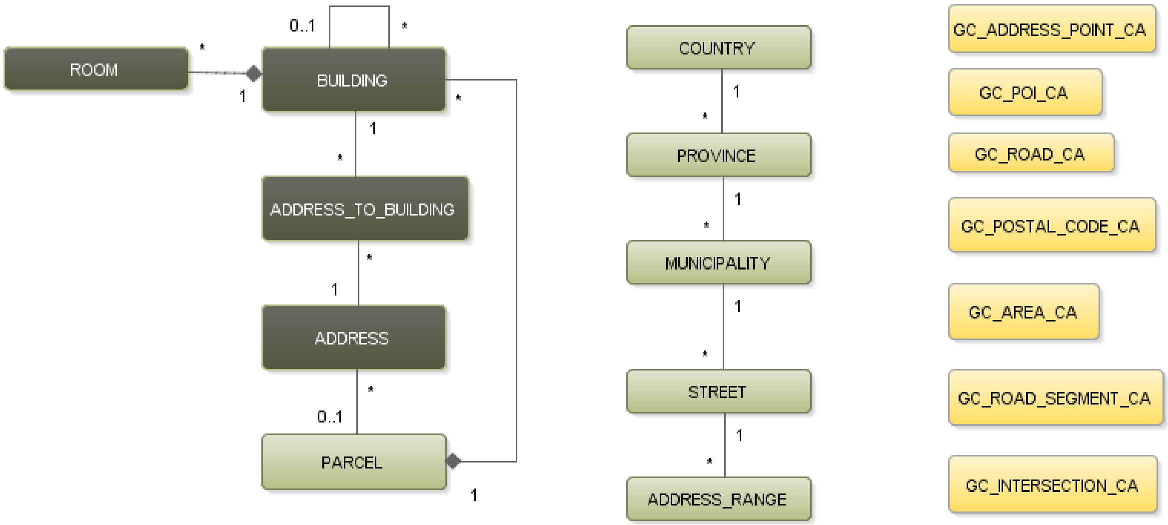

Figure 3 - Initial Model

Oracle Spatial Geocoder metadata tables contain profiles and parsing rules for address matching, including standards adopted by various countries (Oracle, 2011). These metadata tables are shown to the right in brown in Figure 4. The Oracle Geocoder metadata tables should not be confused with Oracle Spatial metadata tables. These metadata tables can be access using system-defined views, the most notable being *USER_SDO_GEOM_METADATA*. Oracle Spatial views and tables are owned by the MD_SYS schema (a schema in Oracle is analogous to an account, in addition to a series of related database objects). Layer specific metadata tables are used to maintain resolution, indexing, and coordinate reference system information for all Oracle Spatial tables.

The two profile metadata tables, *GC_PARSER_PROFILEAFS* and *GC_PARSER_PROFILES*, can be created by invoking a stored procedure called `SDO_GCDR.CREATE_PROFILE_TABLES` and the contents of these metadata tables can be populated using an external script, sdogcprs.sql, which is supplied by Oracle (Oracle, 2009b). The *GC_PARSER_PROFILEAFS* table contains records that define the expected structure of an address for an associated country. Records in the *GC_PARSER_PROFILES* table, which could number in the hundreds, contain alternatives for place names, street types, street directions, civic address numbering (e.g. 1st Avenue instead of First Avenue) and other similar aliases for address components for each specific country. The default contents generated by the script represent profiles for a small set of countries. For a country not represented in the default profile, new entries can be created by modifying copied content from existing records.

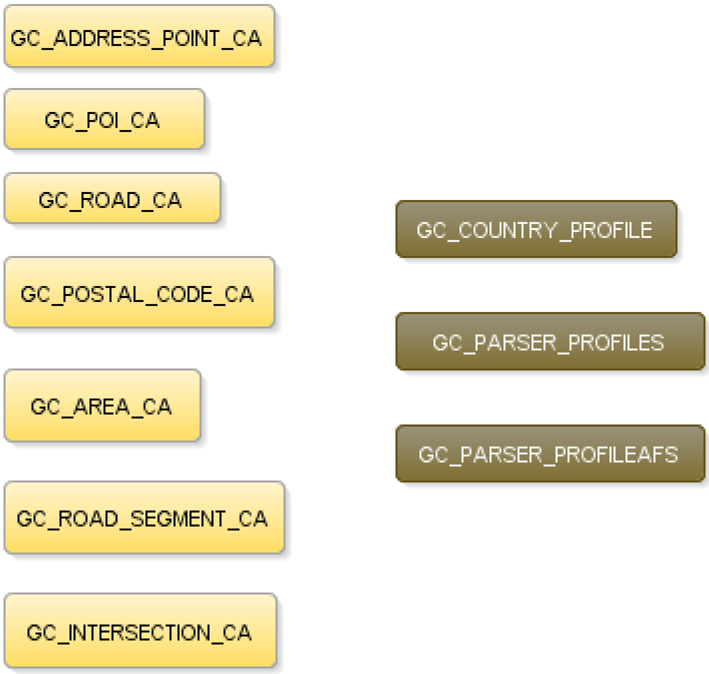

Figure 4 - Oracle Geocoder Data and Metadata Tables

For the Oracle Geocoder, a country code needs to be consistently applied at the row level in all data tables. Maintaining a country code at the record level allows one to aggregate countries into one set of data tables if desired. For example, data both Canada and the United States can be contained in one table. One mistaken or missed country code entry, however, can impair results, and if a metadata entry for a country exists and the corresponding data tables do not exist, the Geocoder will fail to initialize.

Data tables support different geocoding methods used at various spatial resolutions. Data tables contain roads, road segments, administrative areas such as municipalities and provinces, points of interest, road intersections, postal codes, and address points. Users can define versions of these tables by conforming to a naming convention and standard structure. The naming convention is *GC_[table]_[country code]*, with the *[country code]* component often referenced as *XX* in Oracle documentation. Administrative areas in Canada, for example, would be stored in the *GC_AREA_CA* table. The *CA* extension was used was this research.

The *GC_ROAD_CA, GC_ROAD_SEGMENT_CA*, and *GC_AREA_CA* tables are needed for address range geocoding of standard addresses. Street names, types, and directions, as well as identifiers for three hierarchical administrative levels in which a street is contained, namely the settlement, municipality and parent area of the municipality, are stored in the *GC_ROAD_CA* table. Address range geocoding relies on start and end civic address numbers for both sides of a road corresponding to the digitized direction of the representative line segment. Line segments represent sections of road between intersections, or in the case of a cul-de-sac a single line segment joined at one end to an intersection. A geocoding process interpolates a geographic location along the length of a segment. Interpolation is based on the offset of a civic address relative to the values of the address ranges on the line segment. The accuracy of interpolation is reduced when address ranges are irregular or inconsistent, as might be the case for a cul-de-sac or a curvilinear street (Doug Dudycha, personal communication). Road segments are stored in the *GC_ROAD_SEGMENT_CA* table, which is the only geocoding data table that contains an SDO_GEOMETRY column and therefore the only Oracle Spatial table. Road segments in *GC_ROAD_SEGMENT_CA* are contained within administrative areas, which are represented as records in the *GC_AREA_CA* table, with higher level areas containing lower level areas (Oracle, 2011). If the country profile defines a settlement as a level 3 administrative area, Ottawa would be a level 3 area contained in a level 2 Ontario, which would be contained in a level 1 administrative area representing Canada. If counties were to be supported, a municipality would be then be classified as a level 4 administrative area and counties would be assigned to level 3. Some administrative levels might not exist in some countries; however, the *GC_COUNTRY_PROFILE* table can be customized to reflect the administrative hierarchy for a country.

Three of the remaining four tables do not need to be populated for address range geocoding or explicit address representation. *GC_POSTAL_CODE_CA* is used for geocoding that relies on a postal code and works independently from address range geocoding (Oracle, 2011). *GC_INTERSECTION_CA* contains relations of connecting road segments (Oracle, 2011). The *GC_POI_CA* table contains information on point (e.g. monuments) and areal (e.g. parks) landmarks, that can be associated to a civic number, road, administrative area and postal code (Oracle, 2011).

The *GC_ADDRESS_POINT_CA* table stores explicit locations if exact location is preferred and available. The *GC_ADDRESS_POINT_CA* table requires a priori knowledge of a location, so it is in essence a geocoding results table. The content of the *GC_ADDRESS_POINT_CA* table represents in many ways the concept of service location.

Now that the requirements for the Oracle Geocoder have been identified, the foundation for integration with CityGML can be established. A *COUNTRY* table and other administrative level tables were added as stubs in the initial model to facilitate the construction of relationships between CityGML addresses and Oracle Geocoder table content.

4.2.5 The Data Transformation Process

The activities in this section correspond to processes described in sections 3.2.5, 3.2.6, 3.2.7 and 3.2.8.

There are different approaches to structure data for a target database model that integrates the CityGML schema and takes advantage of the Oracle Geocoder. A data developer can use:

- database objects to manipulate existing database content;
- a COTS (Commercial Off The Shelf) ETL application;
- a modified version of the 3DCityDB Importer/Exporter; or
- a custom application.

The first option is the only real candidate for manipulating third party data specifically prepared for geocoding in Oracle, such as the Navteq data described earlier. For Oracle databases, these files are available as dump (.dmp) file format, which is used for importing, exporting, and backing up Oracle databases. The other options in the list are useful for data that require massaging and/or are in various formats. CAD data and non-Oracle legacy data would fall into this category. A COTS ETL tool, such as FME, is the strongest candidate of the latter three approaches because the product can support various formats. If CityGML XML were to be strictly used as an exchange format, then the 3DCityGML importer/exporter would be the preferred candidate. In general terms, custom applications are time consuming and potentially difficult to maintain compared to the other solutions, but do provide exact required functionality.

4.2.5.1 A Review of Database Components and Processing Requirements

Municipal administrative areas, road network, and buildings were the first set of layers to receive spatial data. Although buildings conceptually were to be contained in a parcel, the initial load relaxed this constraint to get a better understanding of how to best integrate road network, addresses and buildings. These were the entities that would bind the CityGML model to the Oracle Geocoder tables. The building data import process was straightforward but the municipal road centerline data from the City of Ottawa did not have a target CityGML object that could retain all the associated information. By default, municipal road centreline geometry that were loaded using the CityGML importer/exporter found a home in the LOD0_NETWORK field in the *Transportation_Complex* object. Road names and address ranges did not have a dedicated target field in the table. Later findings revealed that data with no target ended up in the *genericAttribute* class. The *Transportation_Complex* was not extended to include address ranges because the object represents more than road network. *Transportation_Complex* is intended to contain detailed objects, such as auxiliary traffic areas, which have no association with address ranges. As a result, an *Address_Range* object, as identified in the logical model, was created to accommodate this requirement.

To support the geocoding requirement of the research, City of Ottawa street network data were parsed so that the *GC_ROAD_CA, GC_ROAD_SEGMENT_CA*, and *GC_AREA_CA* tables could be populated. The *GC_AREA_CA* table, used to represent the coarsest level of geocoding, is a parent to the *GC_ROAD_CA* table. The *GC_ROAD_CA* is similarly a parent to the *GC_ROAD_SEGMENT_CA* table. As mentioned earlier, parent-child relationships were implied, but not explicitly defined, in Oracle Geocoder table structures. Considerable cross referencing between fields complicated the process. For example, both the *GC_AREA_CA* and the *GC_ROAD_CA* reference a road segment, a child to both, thus demonstrating that a linear data development process was not possible.

With the relationship established between road segments and roads, the *Address_Range* table was now positioned to carry the TRANSPORTATION_COMPLEX_ ID. The identifier established a connection between CityGML features and business data. Similarly, a road identifier on the *Address_Range* table established a connection to *GC_ROAD_CA* and therefore to the Oracle Geocoder structures.

The Oracle Geocoder required information related to address range, including the highest and lowest civic address per segment. This information was obtained from the FROM_LEFT, TO_LEFT, FROM_RIGHT, and TO_RIGHT fields on each municipal road segment. An addressing scheme or parity for each side, which indicated if the address range was even ('E'), odd ('O'), or mixed ('M') (Oracle, 2011), was also required.

The Oracle Geocoder expected a start, centre, and end coordinate pair for a road, that is, a continuous pavement with a consistent name. The incoming data did not have coordinate information as attributes, so a process had to generate values from the geometry. Regardless of import method, the sequence of road segments that define a road from start to finish had to be established. The natural geometry of a sequence of segments was exploited. An assumption was that any series of segments, regardless of record number, that define a road begin with a vertex that is not shared with any other segment on that road. Likewise the last vertex is not be shared with any other segment on that road. Using SQL, the initial vertex was found using a query based on samples on the Oracle Forum (Oracle, 2010a):

```
SELECT t.y, t.x
  FROM street r, address_range ar,
       TABLE(SDO_UTIL.GETVERTICES(ar.geometry)) t
 WHERE r.street_id = ar.street_id
   AND name = 'BRIDLE PARK DR'
```

```
    AND (t.x, t.y) IN (SELECT t2.x, t2.y
                         FROM address_range ar2,
TABLE(SDO_UTIL.GETVERTICES(ar2.geometry)) t2
                        WHERE ar2.street_id = ar.street_id
                          AND t2.id = 1)
  GROUP BY t.x, t.y
 HAVING COUNT(*) = 1;
```

The last vertex was found using the following query:

```
SELECT t.y, t.x
  FROM street r, address_range ar,
       TABLE(SDO_UTIL.GETVERTICES(ar.geometry)) t
 WHERE r.street_id = ar.street_id
   AND name = 'BRIDLE PARK DR'
   AND (t.x, t.y) IN (SELECT t2.x, t2.y
                        FROM address_range ar2,
TABLE( SDO_UTIL.GETVERTICES(ar2.geometry)) t2
                       WHERE ar2.street_id = ar.street_id
                         AND t2.id = (
                              SELECT MAX(t3.id)
                                FROM address_range ar3,
TABLE(SDO_UTIL.GETVERTICES(ar2.geometry)) t3
                               WHERE ar3.street_id = ar2.street_id)
                        )
 GROUP BY t.x, t.y
HAVING COUNT(*) = 1;
```

With this approach, the nested queries on the latter query were necessary to 1) find the identifier of the last vertex for a segment, and 2) get the coordinate of the coordinates of the last vertex. If the identified vertex were shared or joined to another segment on the street, then the count for that coordinate pair would be greater than 1, and such an outcome highlighted the need for a HAVING clause as a condition on the grouping.

There were a few issues with the previous two queries. This approach did not work with “P” configurations, i.e. streets that close on themselves. Additionally, if the digitized direction of the first and last segment were not the same, then one query returned two points, while the other query returned nothing. The calculation of a center latitude and longitude for a series of road segments was a bit more complex. The average of start and end points was sufficient to calculate mid-point coordinates when collective line segments were straight and not very long. If a road were not straight, the average could not be used. More importantly, however, was that despite the technically intriguing approach, these queries killed performance.

As a result of these concerns, a different approach was needed. A process joined like-named segments to produce a continuous road. In Oracle Spatial, lines were concatenated using the following query:

```
SELECT SDO_AGGR_CONCAT_LINES(SDO_CS.MAKE_2D(geometry, null))
  FROM address_range ar, road r
 WHERE name = 'BRIDLE PARK DR'
```

```
        AND ar.street_id = r.street_id;
```

Line concatenation in Oracle Spatial appeared to only work with 2D data, so a nested `MAKE_2D` function was applied. Once lines were concatenated, the first, last, and mid-point were identified using the `SDO_UTIL.GETNUMVERTICES` function, but the middle coordinate was not guaranteed to be halfway along the road. The definition of a mid-point was simplified by taking the centroid of a polygon created from a buffer of joined lines (Oracle, 2009c), although the centroid was not guaranteed to be coincident to the resulting line. Since the intention of geocoding was to match a civic address, an accurate street mid-point was not a critical requirement and a buffer centroid was considered to be adequate.

The adopted solutions for preparing the street data are similar in function to what can be accomplished using LINEJOINER and BUFFER transformers in FME. The advantage of a using SQL over FME is that it can be used for non-ETL transaction processes. The advantage of FME is that transformers encapsulate many manual and semi-automated processes, thus simplifying tasks.

4.2.5.2 Other Database Considerations

A number of database objects can be used in conjunction with external applications. Database functions can be used to completely replace some functionality offered by external applications. A database trigger is one such object. A trigger is independent of external software upgrades and to some extent model modifications. Triggers are instantly available and are not impacted by external application compilation.

Since CityGML is still a relatively new standard with potential for expanded scope, structural variations are inevitable on future releases, and external applications will likely need to change to accommodate revisions to the standard. The process of synchronizing custom code and interfaces to new releases of open source software would be more cumbersome than implementing custom code in the database as long as the database interface remains constant.

4.2.5.3 3DCityDB Importer/Exporter

For the initial transformation process, the 3DCityDB importer/exporter was explored to determine CityGML functionality and limitations. The application was used to trace record level object creation in the database. Although UML diagrams available from the authors of CityGML were used to develop an awareness of the relationships between objects, an actual physical implementation of classes in the database using the importer/exporter solidified an understanding of object behaviour. The CityGML importer/exported had considerable value as a validation tool or for importing pure CityGML. The export functionality of the 3DCityDB importer was used to confirm that FME ETLs and custom SQL code produced expected results.

4.2.5.4 Interacting with BIM Projects

Two workflows were considered with BIM. The telecommunications partner on the project invested considerably in CAD and as a result there was a need to accommodate transformations from CAD to GIS, while the opposite workflow had to be anticipated. As mentioned earlier, CAD projects generally rely on a local coordinate reference system. Software such as AutoDesk Revit are increasing their support for geospatial data in conjunction with a local coordinate system. Using Revit as an example, projects can be oriented based on an externally linked file, such as a DWG file. Figure 5 shows the results, albeit crude, of a house that was based on a building footprint digitized in ArcGIS and exported as a DWG CAD file. After the building was composed in Revit, the project was exported as an ifcXML file, which was then imported into BIMServer. BIMServer was capable of translating the data to CityGML version 1.0.

Figure 5 - Sample house rendered in AutoDesk Revit

There were two critical criteria that determined if the process worked. The process had to retain semantic information and the spatial reference system. The process was partially successful. BIMServer was able to translate the file to a JSON file that was readable by BIMSurfer as shown in Figure 6. Doors, windows, and the roof were represented in the resulting JSON file, showing promise.

The translation to KMZ and Collada did not work so well. The translation of the KMZ into ArcGIS through the ArcToolBox failed, while Google Earth did not render any of the KMZ content. In fact, Google Earth rotated the globe to prime meridian and the Equator, thus confirming concerns regarding the retention of the coordinate reference system.

QGIS rendered the CityGML output, although again using a local coordinate system. The coordinate settings for Revit need to be used to define a project location. An interoperability test though was not conducted to confirm if a re-applied coordinate location resolved the issue. Future releases of FME should improve data exchange for Revit files in the future.

Another tests had to confirm that data could be loaded and exported as functional GML and CityGML. A building with address information was exported from the CityGML schema in Oracle and rendered in ArcScene. A successful test case established a GML structure, shown in Figure 7, for use as a template in ETL construction. The nodes in Figure 7 are collapsed to illustrate how building and address elements are nested.

Figure 6 - Sample house rendered in BIMSurfer

```
▼<cityObjectMember>
  ▼<bldg:Building gml:id="UUID_0dba1fe8-af43-4e38-b74a-badf7849cd44">
     <gml:description>house</gml:description>
     <gml:name>3 Bridle Park</gml:name>
   ▶<gml:boundedBy>...</gml:boundedBy>
     <creationDate>2012-09-09</creationDate>
   ▶<externalReference>...</externalReference>
     <bldg:usage>primary</bldg:usage>
   ▶<bldg:lod4Solid>...</bldg:lod4Solid>
   ▶<bldg:interiorRoom>...</bldg:interiorRoom>
   ▶<bldg:interiorRoom>...</bldg:interiorRoom>
   ▶<bldg:address>...</bldg:address>
    </bldg:Building>
  </cityObjectMember>
</CityModel>
```

Figure 7 - Sample CityGML Element containing Building with Address Information

4.2.6 Model Revisions

The initial logical model, as represented through UML diagrams, contained three main conceptual components: buildings, roads, and road segments. The elimination of duplicate data was a priority. A revised model had to accommodate Oracle geocoding structures without affecting the CityGML model. Views based on content in the CityGML tables simulated geocoding tables. Metadata for Oracle Spatial objects were created for the road segment view. A geocoding query validated that a view-based solution that inherited the index of the source spatial table worked. The *Road* table was adopted for the street names to conform to Oracle terminology.

Although CAD and BIM typically use a local reference system, an integrated solution required a reference system functional at the continental level. Two reference systems were adopted. The LL-CSRS geographic coordinate system (EPSG 4955 for 3D) replaced the previously assigned EPSG 4326 by offering improved precision for Canadian content. A regional coordinate system, MTM Zone 9, was initially adopted as a regional coordinate system, but the limitations of a 2D projection in Oracle Spatial for BIM and IFC source data dictated that a 3D coordinate system be adopted. A drawback to mixing local and geodetic coordinate systems was identified. Transformations to a geospatial world from a local coordinate system is not natively supported in Oracle Spatial, and logical linkages, such as common field identifiers, are required to bridge the two worlds (Oracle, 2012).

The UML for the interim model is presented in Appendix F.

4.3 Obtaining Service Location Data

Three strategies were taken to compile data in support of service location: geocoding, field work using GPS, and a semi-automated process to derive points from orthoimagery. These strategies do not represent an exhaustive list, as many other technologies described in Section 2.1 could be used, but the chosen set was feasible with the available technology.

A set of sample addresses were compiled to facilitate a comparison between strategies. Of the four sample areas, a subset of parcels was used for the suburban and commercial sample areas. For the each of these sample areas, fifteen percent of parcels with addresses were randomly selected using the Oracle sample function on a temporary parcel table. Because the commercial downtown sample area included residential areas, the sample size in this area was reduced again by only taking a subset of points that better represented the central business district. The industrial and rural samples areas were treated a bit differently. Since the rural area was sparsely populated, positional readings were collected for all identified property frontages for parcels contained within the bounds of the image mosaic. In the industrial area, all commercial properties were observed.

The boundaries of each sample area used in the detection process were prepared externally (ArcGIS) and imported into Oracle Spatial using an FME workbench. An Oracle Spatial query was executed to confirm that the sample areas correctly overlapped previously loaded addressable segments. The number of intersecting addressable segments in Oracle were identified using the following query:

```
SELECT ta.name, count(*)
  FROM address_range ar, sample_areas ta
 WHERE SDO_OVERLAPS(ar.geometry, ta.geometry) = 'TRUE'
 GROUP BY ta.name;
```

The query produced the numbers outlined in Table 14. Results were roughly in-line with expectations as the street names associated with the selected segments appeared to be correct. Segments for each sample area were saved as a new layer using the SQL below, and metadata and a spatial index were added to the table.

Table 14 - Addressable Segments per Sample Area

Sample Area	Addressable Segments
Rural	8
Suburban	28
Industrial	17
Commercial	44

The contents of the layer were read by FME and viewed using a Data Inspector to confirm the expected road segment features were selected.

```
CREATE TABLE [sample area]Roads as
   SELECT ta.name, r.name "ROADNAME", ar.geometry
     FROM address_range ar, sample_areas ta, road r
    WHERE SDO_ANYINTERACT(ar.geometry, ta.geometry) = 'TRUE'
      AND ta.name = '[sample area]'
      AND ar.street_id = r.street_id;
```

4.3.1 Field Collection

GPS was used to collect positions on the ground for rural, suburban, industrial, and downtown sample areas. A Trimble Juno 3D receiver with a 3 metre accuracy was used. Accuracy was improved with post processing that incorporated concurrent readings from a GPS base station. At least 15 one second samples were collected for each location, although longer readings were made for stubborn positions. A minimum of 7 satellites were used, and most observations were based on 9 to 11 satellites. Data were collected over five days under clear conditions. The GPS unit was also configured to optionally collect civic number, street name, and street type, all of which needed to be manually input through a form for each recorded position.

Collecting positions for suburban properties was a mundane and straightforward process. The only concern in the suburban sample area was that vector data and imagery were slightly out of date. Some currently completed construction projects were shown as in progress or not even started in the data.

Rural properties had slightly more to consider. A few properties had more than one point of entry. Some points of entry were fenced and gated, some were not functional in winter, while others that were present in the aerial photography were no longer evident.

The complex nature of service location became more apparent in the industrial sample area. Often the parcels had more than one point of entry. Some points served a single parking lot, while in other cases service bays were separated from employee and customer parking. Sometimes different units within a complex were serviced by different access points. One case had a time restriction on entry, while a bus servicing depot had one way designations.

Downtown commercial areas highlighted additional issues with the identification of service location. Underground or surface parking lots, parking garages, or street level parking that was 10s of metres from a pedestrian access point were commonly observed. One building, typical of many downtown buildings, was examined in detail. The building had emergency exits, street level exterior access points for different leaseholds, and points of access with limitations, such as stairs. The building complex occupied a full city block. As a result, there were multiple addresses for a single building complex.

These observations are not startling, but the physical characteristics need to be identified and modelled. Various building and parcel configurations made the evaluation of access points obtained through other means, such as through the previously described detection process, more complex. Parcel number or ARN were not enough to establish a link to a single GPS location. A single geocoded point needed to be associated with one or more field collected points. Variations in buildings, parcels, and access points required that ancillary data be collected in the field.

4.3.2 Geocoding

During this research, address point data through the City of Ottawa Open Data initiative became available. There were two sets of data: primary addresses, and sub addresses. There were 246841 main addresses and 52585 sub addresses. These address points were derived from parcel centroids. With posted monthly updates by the City of Ottawa, the incentive to develop a separate address maintenance process was reduced, but not eliminated. The value of these data sets needed to be confirmed through a comparison using other means. In addition, not all municipalities participate in an Open Data initiative. A good assumption is that more will come online, considering the growth at which larger municipalities are exposing their data.

4.3.2.1 Initial Geocoding Troubleshooting, Modelling, and Data Preparation

Unless data do not exist for some component of a civic address, every address should geocode, just not necessarily to a desired granularity. A critical analysis of Oracle geocoding requires that the match code, which is embedded in an SDO_GEO_ADDR object that is returned by the Oracle Geocoder, be examined. A match code of 1 is the most desirable. Lower values proportionally identify degraded matching. A match code of 4, for example, indicates that only the postal code and municipality match (Oracle, 2009e).

To confirm that the Oracle Geocoder was functional and that there were no schema specific barriers, a sample dataset covering San Francisco from Navteq was downloaded and examined (Navteq, 2013). The Oracle dump file was loaded using the *imp* command into a Navteq schema. A manual that accompanied the data outlined the steps required for functional geocoding. The following query was executed to confirm that the Oracle Spatial Geocoder responded to the Navteq data. In this query, the inner query returned a virtual table with a single SDO_GEO_ADDR object. Latitude and longitude values were extracted from the object using dot notation on the object through a table alias. The resulting coordinate pair was verified using a Google search.

```
SELECT x.z.latitude, x.z.longitude, x.z, x.z
  FROM (SELECT SDO_GCDR.GEOCODE('NAVTEQ',
               SDO_KEYWORDARRAY('1350 Clay ST','SAN FRANCISO' ),
               'US','DEFAULT') as z FROM dual) x;
```

Although Navteq data worked with the Oracle Geocoder, Ottawa data sourced through a CityGML document did not produce the expected results. The difference in behaviour was traced to the dimensionality of the geometry. Navteq data were stored with 2D coordinates while the extruded Ottawa data, imported into CityGML, contained 3D coordinates. To confirm that the dimensions made a difference, the data were converted to 2D. Since the removal of a vertical dimension would negate the benefit of using a 3D model, a new 2D geometry field of type SDO_GEOMETRY was added to the *Address_Range* table, which was populated using the `SDO_CS.MAKE_2D` function (Oracle, 2009a) on data sourced from *Transportation_Complex*. The conversion process was based on a geographic coordinate reference system, initially WGS 1984 (EPSG 4326). The target column was added manually, then updated using the following statement:

```
UPDATE address_range ar
   SET geometry = (SELECT SDO_CS.MAKE_2D(lod0_network, 4326)
                    FROM transportation_complex tc
                   WHERE tc.id = ar.transportation_complex_id);
```

Since *Address_Range* was one of the source tables for the *GC_ROAD_SEGMENT_CA* view, the data were available to the Geocoder. A query was executed to verify the contents.

```
SELECT x.z.latitude, x.z.longitude, x.z, x.z
  FROM (SELECT SDO_GCDR.GEOCODE('MACHERES',
               SDO_KEYWORDARRAY('3 BRIDLE PARK DRIVE','OTTAWA' ),
               'CA','DEFAULT') as z FROM dual) x;
```

The Oracle Geocoder produced coordinates based on the MTM projection used by the City of Ottawa, and not WGS 1984. Executing the inner query in isolation confirmed that `SDO_CS.MAKE_2D` ignored the EPSG parameter. The update statement was modified as follows:

```
UPDATE address_range ar
   SET geometry = SDO_CS.MAKE_2D((
                SELECT(SDO_CS.TRANSFORM(lod0_network, 4326))
                  FROM transportation_complex tc
                 WHERE tc.id = ar.transportation_complex_id), null);
```

With this change, the previous select statement returned a value coordinate pair of 45.28468, 75.85614, which was validated using a Google search.

With the Geocoder now properly configured, bulk geocoding was completing by executing a stored procedure. The source code for the procedure is presented in Appendix E.

4.3.3 Semi-Automated Process using Image Classification

A semi-automated process to identify service location points from imagery, from now on referenced as the detection process, and the role of a parcel level geocoder were explored to determine if geocoding provides the best spatial representation for service location. Although driveways were the primary objects to be identified in the detection process, paths and walkways were also needed, and as a result the term access point was adopted. The evaluation of all processes was based on sample areas representing commercial, suburban, industrial, and rural locations. ArcGIS with the Spatial Analyst extension was used to develop most processes except where explicitly noted.

The automated detection of roads using imagery has been researched before (Mena, 2003; Unsalan & Sirmacek, 2012; Wang, Hu, & Zhang, 2005) with a focus on defining linear paths. Driveway and access way detection was identified as a popular area of research, but research was focused on navigation (Huang, Chen, Hsiao, & Fu, 2004). The process to be outlined for this paper focused on defining point locations using basic orthoimagery and vector datasets. The objective was not to build a new mathematical model, but a process model. A supplemental source, and not a definitive source, of point locations was the desired outcome.

A few initial assumptions were made from the outset for the detection process. First, that classification would not likely improve access point identification in dense urban areas due to the obstructive nature of height displacement. Second, the proximity of adjacent parcels in urban areas would favour access points derived from parcel level geocoding over image processing.

Image mosaics for each sample area were used to define consistent datasets from which statistics for classification were gathered. A few attempts were made to define a reasonable classification scheme. After a simple default equal interval classification was applied in a rural area with modest results, the process was applied in a suburban area. The results were disappointing, with less than 10% of the total number of driveways identified. As a result, a supervised classification in the suburban sample area with two basic training areas, one for trenches/ditches and one for driveways, was applied. A visual inspection showed the process did not provide meaningful results. The composition and reflectance of driveway material made it difficult to establish training areas. Edge detection was also attempted. Edge detection was able to detect boundaries of classes, but as with image classification, no consistent process could be easily established.

Although an unsupervised classification using 10 classes did not generate much value, 20 classes generated enough granularity to distinguish between roads, driveways, and roadway embankments. Applying similar symbology to groups of classes provided a visual clue for reclassification. Using the visual guides, a reclassified image, called *[sample area]Reclassified*, was created.

It was apparent that classification alone was not sufficient to derive access points. Another criterion needed to be applied. If driveways and other access points were to be perfectly classified, the expected area would fall within a certain interval. Based on an observation of a single driveway and visually comparing other driveways in an orthoimage, a single car driveway was estimated to be 3 metres wide, which was slightly above the minimum driveway width of 2.4 metres specified in a City of Ottawa by-law (City of Ottawa, 2012d). A buffer created from parcel outlines which overlapped the classified image was used to further restrict access point candidate selection. Classified polygons that had a similar perimeter and area to that of buffered centroids were considered to be candidates for access point detection. The idea was that irregularly shaped polygons would be dismissed from a list of candidate polygons.

Figure 8 shows a candidate access point polygon (bottom circle) on a field boundary that was obviously not an access point. Although a different limit on the perimeter could have been applied to weed out this case, clearly only parcel frontages were needed to identify access points, while still placing some limit on the area and perimeter of classified polygons. The establishment of a process to obtain property frontages to improve extraction of access points from a classified image also produced the groundwork for a parcel level geocoder.

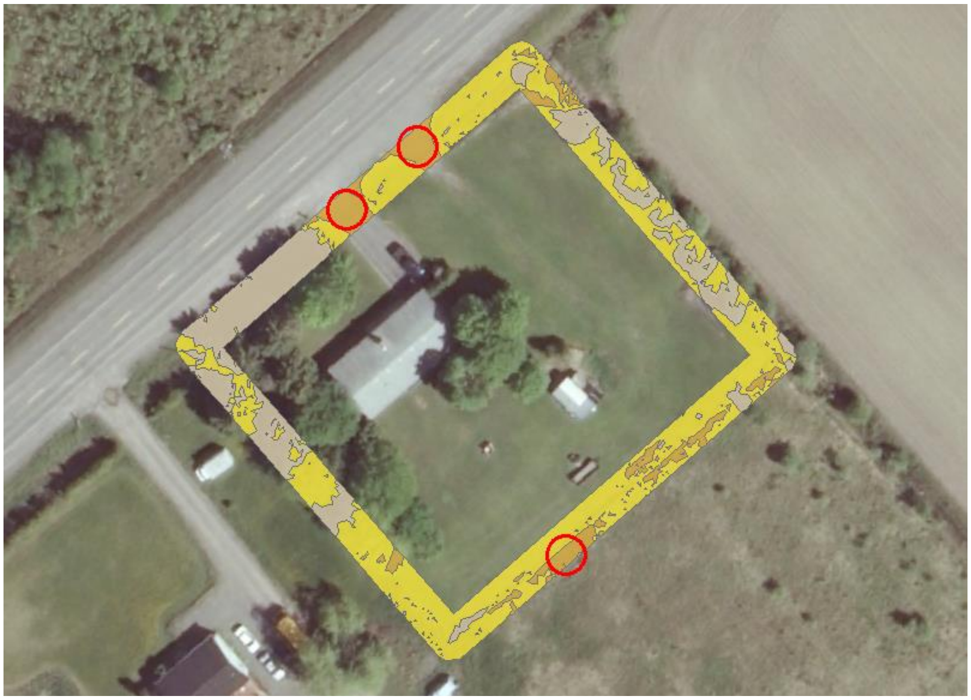

Figure 8 - Sample of a classified parcel buffer

4.3.3.1 A Process to Identify Property Frontages

Parcels were clipped from the city wide layer using the image mosaic boundary for each sample area. The clipped polygons were retained as a reference to validate the creation of complementary layers.

A one-to-many spatial join of parcels to roadways identified parcels that contained one or more road segments. Any parcel with a non-zero join count was saved in a right-of-way feature class. On the odd occasion, a road network segment crossed into a property parcel, which resulted in a misidentified right-of-way. An editing process heavily reliant on visual inspection was required to clean up these cases. After these corrections were applied, a dissolve process was executed on the right-of-way parcels to create a single polygon.

Parcels with a zero join count, the complementary set of the right-of-ways processed earlier, and the corrected cases were saved as a property layer. Ownership parcels contained address information, but not a consistent unique identifier. The Assessment Role Number (ARN) was transferred to the property polygons from the ownership parcels using a one-to-one spatial join on identical features to create *[sample area]AssessmentProperties*.

The *[sample area]AssessmentProperties* feature class was converted to an outline feature class, called *[sample area]outlines* using a polygon to line tool. Pseudo nodes on the outlines were removed (unsplit) using the dissolve function to create a cleaner parcel outline called *[sample area]DissolvedOutlines*.

Figure 9 shows the need for this step. The road right-of-way was buffered to a distance that overlapped property frontages, which were not always a fixed distance perpendicular to road centreline (driveway A in Figure 10). An original 3 metre wide buffer using the right-of-way did not identify all parcel frontages. In some cases, property frontages were located behind a narrow linear parcel that fronted the road across a number of parcels, as shown in Figure 10 at driveway B. As a result, a 6 metre buffer distance was applied to the right-of-way.

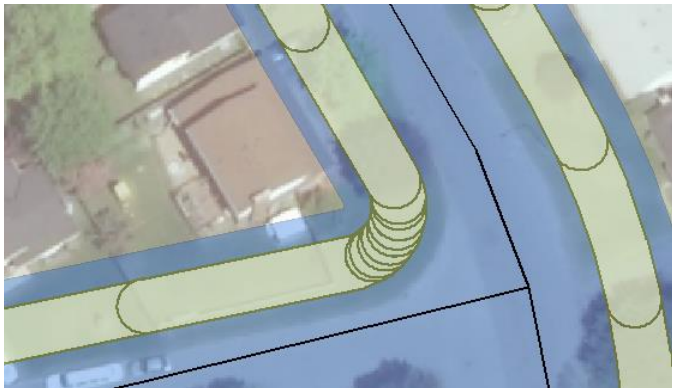

Figure 9 - Case for flat edge buffering and pseudo node removal

Parcel frontages in *[sample area]DissolvedOutlines* were identified using a Select by Location operation with the condition that a line feature had to be completely within the right-of-way buffer, with the selected set being saved as a *[sample area]ParcelFrontages* feature class. The process selected almost all the property frontages for parcels within the orthoimage extent except for the small percentage of property lines that extended beyond the right-of-way buffer, as shown in Figure 11. The situation is an anomaly related to selecting a sample set of parcels and unsplitting lines. The attributes of the parcel polygons were transferred to the identified frontages using a one-to-many spatial join and the SHARE_A_LINE_SEGMENT_WITH method, producing a feature class called *[sample area]ParcelsJoined*, The resulting feature class was then buffered by 2.5 metres with flat edges to create *[sample area]BufferedFrontages.*

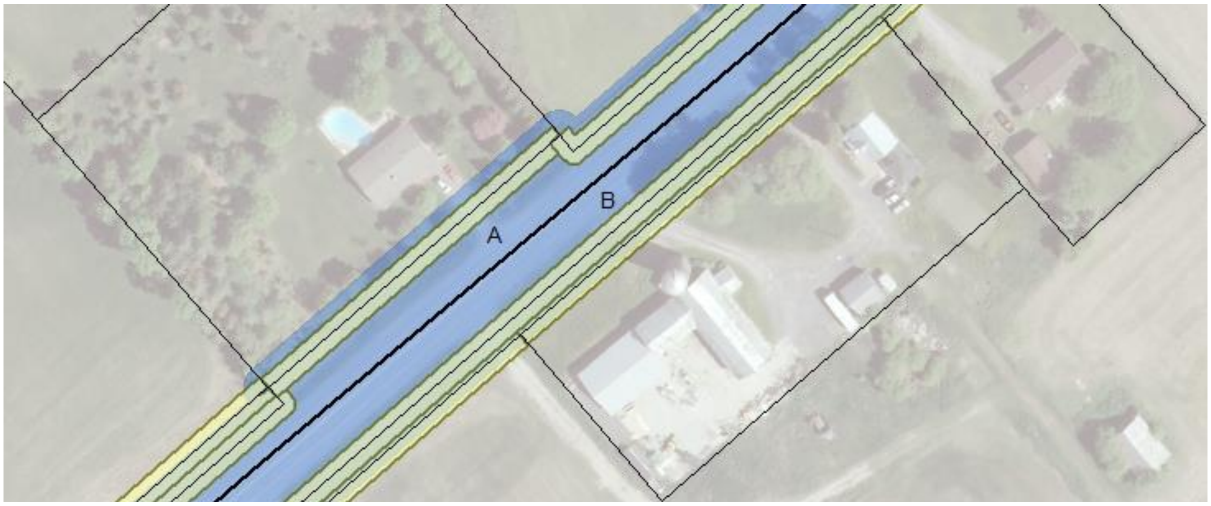

Figure 10 - Right-of-way buffering considerations

At this point the buffered property frontages and the classified image had to be in the same projection. The reprojected feature class, called *[sample area]PropertyFrontsReprojected,* was used in the detection process. The property frontages and classified polygons could have been merged, but that was an oversight identified in the post-process.

Figure 11 - Identifying missing frontages

4.3.3.2 A Process to Identify Access Points

The reclassified raster mentioned earlier was converted to a polygon feature class, called *[sample area]AccesswayPolygons,* which was subsequently clipped using the reprojected property frontages. This process resulted in a feature class called *[sample area]ClassifiedFrontagePolygons*.

Classified polygons for each parcel frontage were then subjected to a filtering process. Driveways were identified by selecting correctly classified polygons of a preferred area as a filter, then filtering smaller and smaller polygons up to a minimum threshold. If no driveways were identified by that point, then a larger sized area threshold above the preferred size was selected. If a filter identified a candidate driveway, a temporary point feature using the centroid of the identified polygon was created. Fields were added to the point and populated with attributes indicating the parcel ID, ARN, run number (an ID to keep track of various test runs), and the filter through which a point was identified. Identified points were then appended to a master results layer. Capturing the parcel ID and the ARN established a logical relationship between access point and parcel and removed any dependence on a spatial relationship. Although some spatial relationships (e.g. nearest) could have worked with some certainty, there was no real processing cost in establishing a linkage using attributes.

There were no absolute values for parameters for the various filters. The distances chosen were based on trial and error, with an attempt to balance the effect of special cases that might have been captured or missed with different tolerances. The property frontage buffer distance was also subjective, with the chosen offset an attempt to reduce cases where one driveway would “bleed” into another. An initial 3 metre buffer, similar to the default driveway width, was reduced to 2.5 metres to remove such cases.

Most of the processing was manual - a reflection of the experimentation that occurred during research. A Python script (see Appendix C) for use in ArcGIS was developed to automate some aspects of the detection process. The visualization capabilities within ArcGIS made it easy to validate results. Figure 12 shows the property buffers (yellow) and access points (red) superimposed over the right-of-way buffer.

There were cases under which the process did not work well. Short driveways in the suburban area resulted in cars in the buffer zone, which impacted classification. A greater percentage of driveway surfaces in the suburbs were obscured by tree canopies compared to other sample areas.

In the commercial sample area, buildings obscured ground level features and long shadows affected classification. Wider paved surfaces adjacent to streets in the commercial area impacted the ability to constrain selection using any standard entrance width. Poor results in the commercial sample area discouraged further attempts to adjust the process, and as a result the detection process was abandoned for the commercial area. If imagery were captured on slightly overcast skies or with an increased flying altitude to reduce height displacement, then results might have been better in the commercial sample area. Aerial photography captured with a light snow cover would also likely have improved classification. These options were not available for this research.

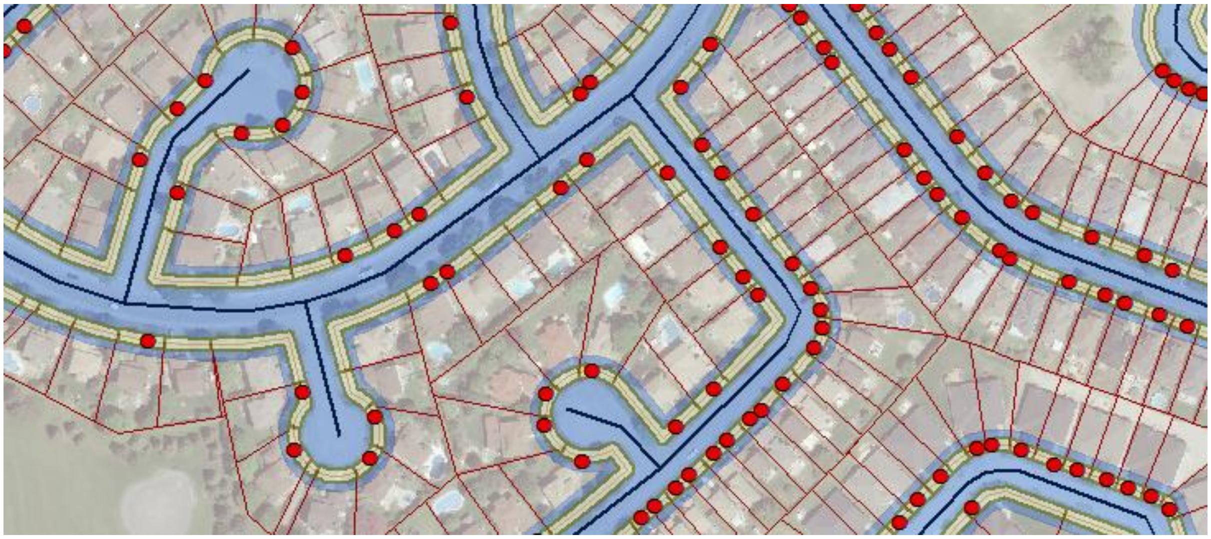

Figure 12 - Identification of access points in the suburban sample area

The intention of the detection process was not to create a final dataset, but demonstrate that conventional location information can be supplemented by data derived by unconventional processes. Even if not all access points were identified, enough were to support QA processes. The detection process was bound to include errors or commission and omission. A process to clean up these errors was unnecessary, since the results were used to identify outliers from various sources; likewise, various sources were used to validate data derived from the detection process.

The parcel frontage polygons were deemed to be sufficient to eliminate development of a similar Oracle Spatial process, which would use the `SDO_BUFFER` operation on parcels and right-of-ways. An Oracle nearest neighbour search method using the parcel centroid and road centreline data was informally explored, but the process only identified the nearest road segment, which was more granular than a centroid from a buffered frontage.

4.3.3.3 Implementation Considerations

The detection process was completed using a personal version of ArcGIS 10 on a Windows 7 laptop running a Pentium dual core 2.3 GHz processor with 4GB of RAM. During development, the processing time for the suburban sample area varied from between 3 hour 21 minutes to 16 hours, 5 minutes, but these were not rigid observations. The deletion of temporary datasets, which was inconsistently applied, affected processing time. Other processes, such as overnight anti-virus software scans, likely impacted performance. Modifications during development also improved processing time.

Implementing a similar strategy for national coverage would be dependent on the availability of comparable or better data. For City of Ottawa imagery, one 1 kilometre x 1 kilometre orthoimage costs $100 (Stephen Perkins, personal communication, January 24, 2013), which would be considerable and likely prohibitive for an area of any reasonable size. In-kind contributions by utilities in exchange for improved E-911 data could reduce or eliminate acquisition costs. Alternative image formats and suppliers should be worthy of consideration. Panchromatic imagery as coarse as 0.5 metre pixel size may be suitable for driveway detection (Jensen & Cowen, 1999). GeoEye and QuickBird imagery should be options based on these guidelines.

4.3.4 A Comparative Evaluation of Service Location Detection

The collection of points using various strategies highlighted a need for a temporary dedicated point table to support data analysis. Once the data were loaded into this point table, a method was required to establish a connection between all relevant points for a given address. Detected points associated with a parcel and related GPS points needed to be associated with a common address identifier. With a common address identifier between points collected using different strategies, analysis could be initiated.

An address identifier was not enough to establish a comparative evaluation. The displacement between detected points and geocoded points to a GPS point needed to be determined. A process to create an Oracle Spatial table containing displacement vectors is included in Appendix D.

Since more than one GPS point was collected on certain parcels, especially in the commercial and industrial locations, the shortest link between a GPS collected point and its peer point for each method was kept. For example, although 100 GPS points were collected in the industrial area, only 60 were needed to evaluate an addressable parcel. All GPS points had a peer geocoded point, while the detected points found a peer GPS point only 67% of the time in the best scenario (Table 15). A cursory examination can be deceiving. Less than 30% of the geocoded points matched all criterion (e.g. matching street name, street type, municipality, etc.), while over 30% had a match code of only 4. Meanwhile, the match gap with the detected points was related to the quality of the address information in the parcel data.

Table 15 - Points per Sample Area

Sample Area	Collected GPS Points	Parcels	Detected Points with GPS	Geocoded with GPS
Rural	34	29	8	29
Suburban	104	104	64	104
Industrial	100	60	40	60
Commercial	76	45	n/a	29

An expectation of accuracy needed to be defined. Oracle geocoded points were coincident to a road centreline, while detected points were contained within a frontage buffer, which was generally offset from the road edge by a few metres. GPS points, to remain safe while being unobtrusive, were collected at the road/driveway intersection. In the best possible situation, the GPS collected point would be somewhere between a geocoded point and a point in the parcel buffer. Equivalency was therefore an area of tolerance around the GPS point that would include a best fit geocoded point and a detected point. In the rural sample area, the average perpendicular offset of a GPS point to the road centreline based on 10 samples as measured from orthoimagery was 5.1 metres. In the suburban sample area, 16 samples were used to calculate a perpendicular offset of 3.4 metres, and 16 samples in the industrial sample area yielded a distance of 7.0 metres. As a result, an area of tolerance of 7 metres was seen to be sufficient.

In an initial comparison of displacement distances between detected points and geocoded points, the standard deviation was over 8 kilometres in the suburban sample area, with a maximum displacement distance of 74 kilometres. These extreme cases were a result of points being geocoded to the municipality centroid. Geocoder sensitivity to variations in street type were a factor in these results. Adjusting the street type domains so that they matched OUTPUT_KEYWORD values in the *GC_PARSER_PROFILE_TABLE* associated with 'STREET_TYPE_KEYWORD' value contained in the SECTION_LABEL field improved results. The standard deviation was reduced to 228 metres and the maximum distance was reduced to 844 metres.

The coarsest match codes were used to identify focus areas in the data. Figure 13 shows that some displacement overlap existed in match codes, suggesting that a focus on standard deviation alone would not resolve issues. There were cases where a geocoded point with a more granular match code of 4 was actually closer to a GPS point than cases with a more precise match code. These cases appeared when a parcel was located near the midpoint of a road, resulting in a coincidence rather than accurate geocoding. Since some location errors were related to the same problem (e.g. street type issues), many poor results were improved with a single modification.

Once an initial data repair was completed, the displacements between GPS collected points and geocoded and detected points were used to identify where further data tuning could be undertaken. Visual analysis

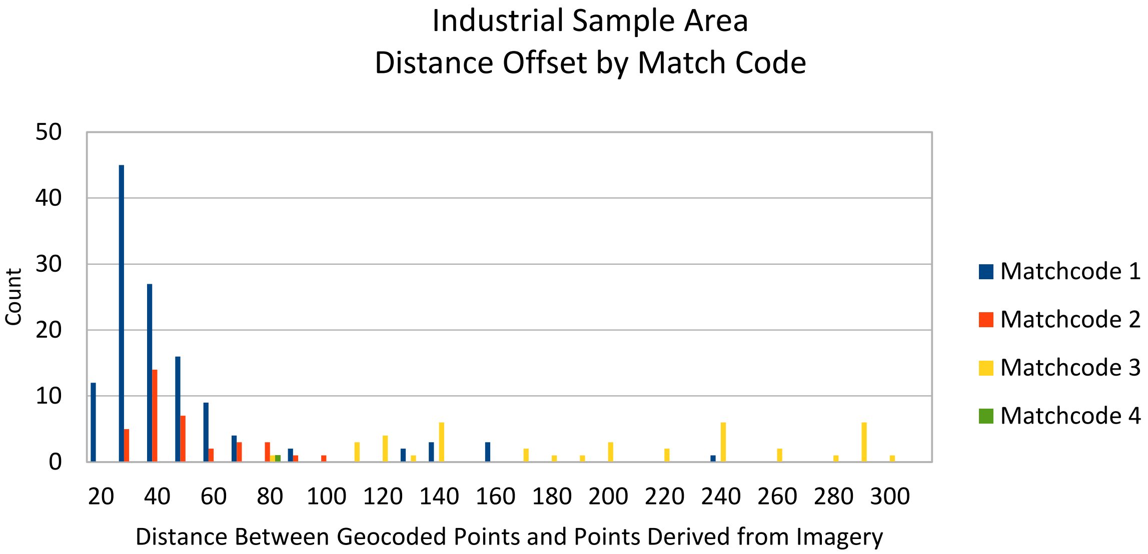

Figure 13 - Sample Geocoding Match Code Distribution

was completed using ArcMap after displacement data were exported to a ShapeFile using FME. Examples of displacement vectors are shown as long red line in Figure 14 to 17.

There were other factors that affected results. The first attempt to match point data in the industrial sample area was a disappointment, since many parcel address entries lacked a street direction component. A manually effort to clean street name was necessary to derive at least partial results. Data could not be cleaned completely since some parcels did not have any address information. As a result, the initial sample size was greatly reduced. In such cases, a topological relationship could have been established, which would have retrieved the identifier of the adjacent road.

In the industrial sample area, geocoded points (red lines) in Figure 15 are considerably further than the detected points, the displacement of each is shown as a green line that is barely visible, and in this case only the black arrowheads are visible.

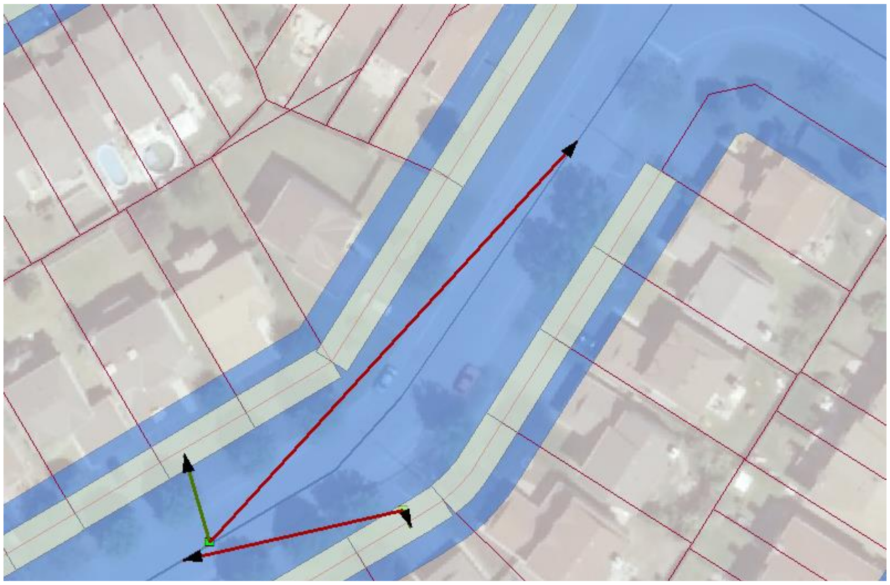

Figure 14 - Displacement of Geocoded Point

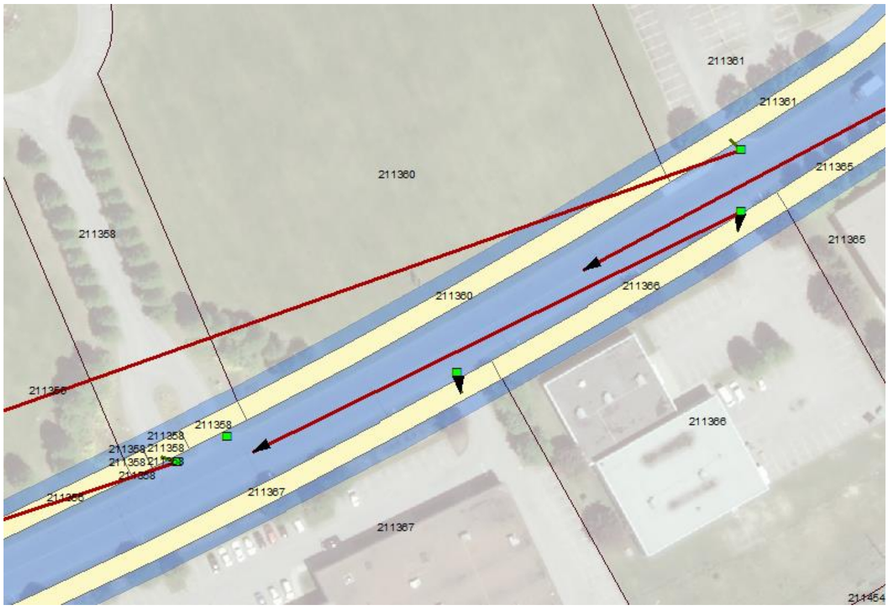

Figure 15 - Example of Point Displacements in Industrial Areas

In the suburban sample area, the detected points appeared to match quite well with respect to the GPS points. One anomaly in this sample in the bottom right quadrant of Figure 16 is obvious. Both the detected points and the geocoded points were considerably distant from a peer GPS collected point. One would conclude that the address for a GPS point (small green squares) embarrassingly was not entered correctly. This case demonstrates that UGC, no matter the intention, can have errors and that different sources can be used to evaluate data quality.

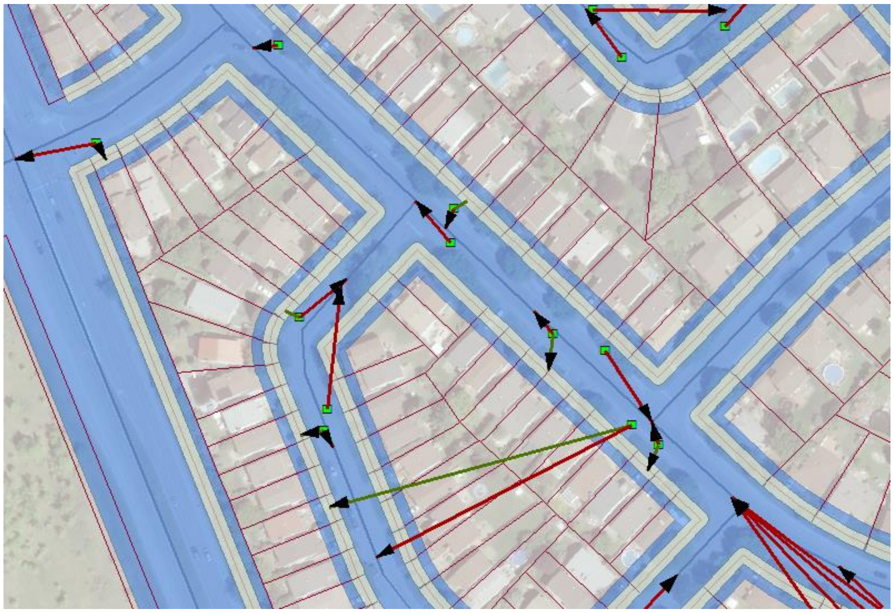

Figure 16 - Example of Point Displacements in Suburban Areas

Finally, results from the rural area (Figure 17) demonstrate that the detection process can work well as illustrated by comparing displacement lines representing detected points (green lines/arrowheads) with displacement lines for geocoded points (red lines with arrowheads).

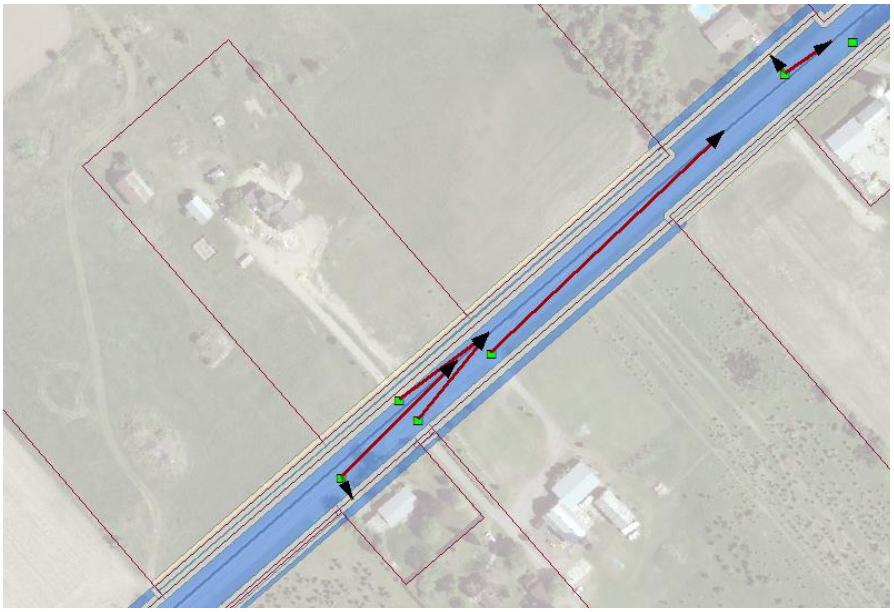

Figure 17 - Example of Point Displacements in Rural Areas

4.3.4.1 Statistically Comparing Strategies for Determining Service Location

The benefit to preparing data in Oracle is that the displacement vector layer could also take advantage of statistical functions in Oracle. Displacement distance were calculated using the following SQL statement:

```
SELECT SDO_GEOM.SDO_LENGTH([geometry column], [tolerance])
  FROM [displacement vector layer];
```

The Oracle `AVG` function was applied to the results of the above statement to calculate the average displacement, while the `STDDEV` function, which takes the displacement length as an argument, was used to calculate standard deviation. Table 16 outlines the average displacement distance and standard deviation of the displacement distances for each data collection method. At first glance, the detected points appeared to be an improvement over geocoded points in the suburban and industrial sample areas.

The results in the rural sample area posed a challenge. In almost all cases, the standard deviation was too high, suggesting that there were still outliers in the data. Upon further examination, only eight detected points matched with an address, which could be linked to a GPS point for comparison. An examination of the parcel data showed that many of the rural areas did not have a civic number but did had a lot and concession number. A nearest neighbour match or topological adjacency test could be used to transfer the road name in such cases, but without a civic number, geocoding matching was limited. Further observations of outliers based on standard deviation showed that two detected points that matched an address through parcel information had a high displacement (132.8 metres and 62.0 metres). This was a result of properties with multiple entry points with long frontages. Although a detected point would be certain to fall within the property frontage, the real value in any proposed process is the ability to link a civic address to a point location.

The detection process in the suburban and industrial areas fared much better, but with still relatively high standard deviations suggesting that considerable outliers were present. The high average distance with the geocoded data indicated that a few street records required further improvements.

Table 16 - Descriptive Statistics for Comparing Collection Methods

Sample Area	Geocoded Points	Geocoding Average Distance	Geocoding Standard Deviation	Detected Points	Detected Average Distance	Detected Standard Deviation
Rural	29	50.599	38.870	8	31.281	46.188
Suburban	104	46.828	95.230	64	8.961	13.120
Industrial	60	358.319	253.654	40	5.949	8.566

A comparison favoured detected points over unfiltered GPS points. By applying a match code filter on the geocoded points, i.e. civic number matching, geocoding statistics improved (Table 17). The removal of the GPS entry error identified earlier in Figure 16 and minor content corrections also improved geocoding displacements. Missing entries for 'GT' or 'Gate' in the GC_PARSER_PROFILES table improved the match code for affected streets. With minor improvements, 37 geocoded points in the industrial sample area had a high match code. All rural geocoded points matched on the civic address. Over 82% of suburban geocoding matched the civic address with about half with a high match code. With these numbers, a fair comparison between geocoded points and detected points could be carried out. The second column of Table 17 shows the number of points with a high match code after minor corrections. The third column of Table 17 shows the number of GPS and detected points that matched based on address id.

Table 17 - Statistics using Match Code 1

Sample Area	Geocoded Points (match code 1)	Common points	Geocoding Average Distance	Geocoding Standard Deviation	Detected Average Distance	Detected Standard Deviation
Rural	29	7	50.599	38.870	31.281	46.188
Suburban	86	48	21.962	13.718	8.961	13.120
Industrial	56	37	68.860	47.382	6.087	8.769

A one sided t test was used to determine if the means were significantly different. The t-test results were obtained using the `STATS_T_TEST_ONE` function in Oracle. The following SQL is an example of a test for the suburban sample area, which uses the average from the detected points in Table 16:

```
SELECT stats_t_test_one(
            SDO_GEOM.SDO_LENGTH(offsetline, 0.05), 8.961,
            'STATISTIC') t_obs,
       stats_t_test_one(SDO_GEOM.SDO_LENGTH(offsetline, 0.05), 8.961,
            'ONE_SIDED_SIG') t_sig,
       stats_t_test_one(
            SDO_GEOM.SDO_LENGTH(offsetline, 0.05), 8.961,
            'DF') dof
  FROM suburbangeocoding g, geocode_details d, access_point a
 WHERE g.geocoding_id = d.access_point_id
   AND g.geocoding_id = a.access_point_id
   AND not a.address_id = 1291
   AND matchcode = 1;
```

The mean for the detected points was used to represent the population mean because the averages and standard deviations of the displacement values were lower than geocoding displacement values. The hypothesis was that the displacement of geocoded points was not significantly different at the 95% confidence interval from that of the detected points.

Table 18 - Values for a Student T-test to Determine Significance of Mean Displacement

Sample Area	Observed t-value	Approximate t-value threshold	Significance of t	Degrees of Freedom
Rural	1.521	1.943	0.08948	6
Suburban	6.112	1.677	0.00000	47
Industrial	7.283	1.688	0.00000	36

The observed t values and significance values indicated that the means were significantly different in the industrial and suburban sample areas, but not so in the rural sample area. The rural area, however, did not have many records with a common address id. Based on the results in Table 18, the alternative hypothesis was generally true, suggesting that the detected points can be used to validate and improve geocoding data, when available.

4.4 An Implementation of Service Location in the Extended CityGML Model

This section will provide a summary of the research as well as some concluding comments, mostly in response to information collected in the field.

The process of collecting GPS points highlighted some aspects of service location that had to be modelled. Service location should not just describe access to infrastructure, but must also deliver information related to location. Barriers, time based restrictions, and similar attributes associated with access need to be considered. Crews must know hardware infrastructure and environmental conditions before arriving at a location to determine resources for deployment (Mladineo, Knezic, & Jajac, 2011) Crews must also understand the relationship of hardware components to each other, which in part dictates site requirements (Peter Glenday, personal communication).

4.4.1 Data Model Modifications

Data collection, the usage of various software, even at a cursory level, and the exchange of data between applications highlighted a fundamental requirement for the data model. A separation of data preparation and enterprise structures was necessary. Separating CityGML, CAD and core Oracle Spatial functionality into separate schemas was seen as a viable solution. Furthermore, different CAD schemas were proposed to maintain separate projects based on CO work areas.

These suggestions might seem contradictory to the goal of this paper, but the view of the author is that integration does not require absorbing all parts of a system, but should enable information exchange. The CityGML encoding standard acknowledges that “logical subsets of CityGML” could be used for “valid partial implementations” but a model is not conformant unless the core module is fully implemented (OGC, 2012, p. 17). The requirements of the national carrier would benefit from this flexibility.

The exercise of pulling data to and from different systems highlighted the need to support more than one coordinate system. An implementation in 3D dictates that local, geodetic, or compound coordinate systems be used. The CityGML encoding standard states that "each geometry can have its own coordinate reference system" (OGC, 2012, p. 25). Considering the flexibility in the standard, at least two functional coordinate systems are proposed: one based on LL-CSRS geographic coordinates (EPSG 4955 for 3D) and a local coordinate system. Oracle suggests that for geodetic coordinates that a tolerance no finer than 0.05 metres be adopted (Oracle, 2009d). Engineering documents would require finer tolerance. Tolerance requirements alone provide enough justification for at least two coordinate system. Transformations to a geospatial world from a local coordinate system is not natively supported in Oracle Spatial, and logical linkages would be needed to bridge the two worlds. Another option would make use of an "anchor point" (OGC, 2012, p. 28) to spatially relate a local coordinate system to a geodetic coordinate system. Local coordinate systems itemized in the *SDO_SRS_ENGINEERING* view in Oracle Spatial should be used to represent CAD data.

4.4.1.1 Model Modifications to Support Field Operations

The model was augmented to support a number of access points for the same parcel or building that were collected using various devices. In addition, a revised model preserved information on function and usage associated with these points. The function and usage domains contained in the CityGML version 2.0 encoding standard (OGC, 2012) previously reviewed in Section 5.2.3.3.1 was incorporated in the model. Annex C of the encoding standard does mention that codes associated with functions and usage, a concept inherited from the GML 3.1.1 CodeType mechanism (OGC, 2012), are not mandatory and are incomplete. If the codes were to be adopted, attributes would need to be captured in any future workflow. For example, GPS forms should be customized to collect function and usage information in a standard way, and not as free form notes.

Some points that were collected in the field did not fall neatly under one CityGML class. Multi-storey parking (code 1610) and level parking (1620) is an *AbstractBuilding* function and usage, while a parking bay (code 1100) is an *AuxillaryTrafficArea* function. The definition of a car_park (code 7) function for *TrafficArea* and a parking_area (code 1610) or driveway (code 1210) function and usage for *Transporatation_Complex*, illustrated that additional field information will need to be captured.

The introduction of a *Parcel* layer ensures that contained objects, whether that be a *Building* or a *BuildingComplex* can be logically and spatially referenced. In addition, a parcel could indirectly reference an *Opening,* such as a door, which if spatially referenced could define an access point.

4.4.2 Data Management

Different LoD views support a number of geometries for the same object. Processes have to be established to manage the synchronization of different representations of the same physical object, which are aggregated at lower LoDs (OGC, 2012). The CityGML encoding standard provides the framework for data storage, but explicitly states that the user is responsible for synchronizing references to the same real world object across LoDs (OGC, 2012). Any process would need to ensure that objects at different LoDs are assigned a consistent identifier. A GML ID based on a Globally Unique ID (GUID) is proposed.

The use of 3D data framework would also support quality control processes in the development of access point features. Data collected within buildings could be verified using 3D models. Likewise, collected access points could also be used to identify quality issues with 3D CAD and BIM models. The advantage of GPS and other collection methods in GIS is that traditionally they offer a level of independence to a normal architectural workflow.

4.4.3 Data Schemas

An *Access_Point* table, in which data would be cleaned and validated, would store collected data. The content of the *Access_Point* table, as a geometry table, could be validated against other layers of information, such as a *Parcel* or *Building* layer using built-in spatial operators, relational queries, or visual overlay. Data derived from 2D sources would require that the vertical component be transferred from other sources. Data from a 3D *Parcel* or *Building* object could be transferred. If buildings and parcels are 2D, there are opportunities to transfer a vertical component from an underlying digital terrain model if available.

Attribute values could be managed using domains stored in lookup tables, but would not be constrained until workflows ensure that data are in conformance with data model expectations. Access point addresses collected by field personnel would benefit from devices that are capable of synchronizing official name, type and direction with the *Physical_Address* table, which was introduced to distinguish the content from a mailing address and allowed for a more extensive representation of address than that contained in the CityGML *Address* entity. The CityGML *Address* object could then be then implemented as a view based on the *Physical_Address* table.

As a point feature type, the *Access_Point* table, shown in Figure 18, would behave like a LOD0 feature. A relationship to the *Physical_Address* object would allow access points to be logically connected to property parcels even if any are spatially in a right-of-way, as was the case for geocoded points. An address ID would also reduce the number of required joins that are evident in SQL code in Appendix D. Related information on barriers, such as time restrictions, height restrictions, gated access, etc. would be contained in a separate *Access_Barriers* table. The *Access_Type* table would contain content to describe whether a point represents a building, parking lot, service bay, and things of a similar nature. *Access_Priority* would identify whether access is primary (main doors), secondary, or an emergency exit.

A dependency between the *Access_Point* entity and the *GC_ADDRESS_POINT_CA* table could be established. That dependency would ensure that the most accurate point, regardless of collection method, be used as in the *GC_ADDRESS_POINT_CA* table, and as a result enhance the geocoding process.

The concept of the access points could be applied to improve point and vertex positions for physical inventory. Multiple points representing a single utility pole for example could be collected, with an assessment process put in place to identify the most likely position of the pole. A common identifier on all records representing the same physical object, similar to the address identifier used in this paper, would be a requirement, the absence of which would likely be the largest barrier to implementation.

The remaining elements of the proposed model are carried over from the interim model in Appendix F.

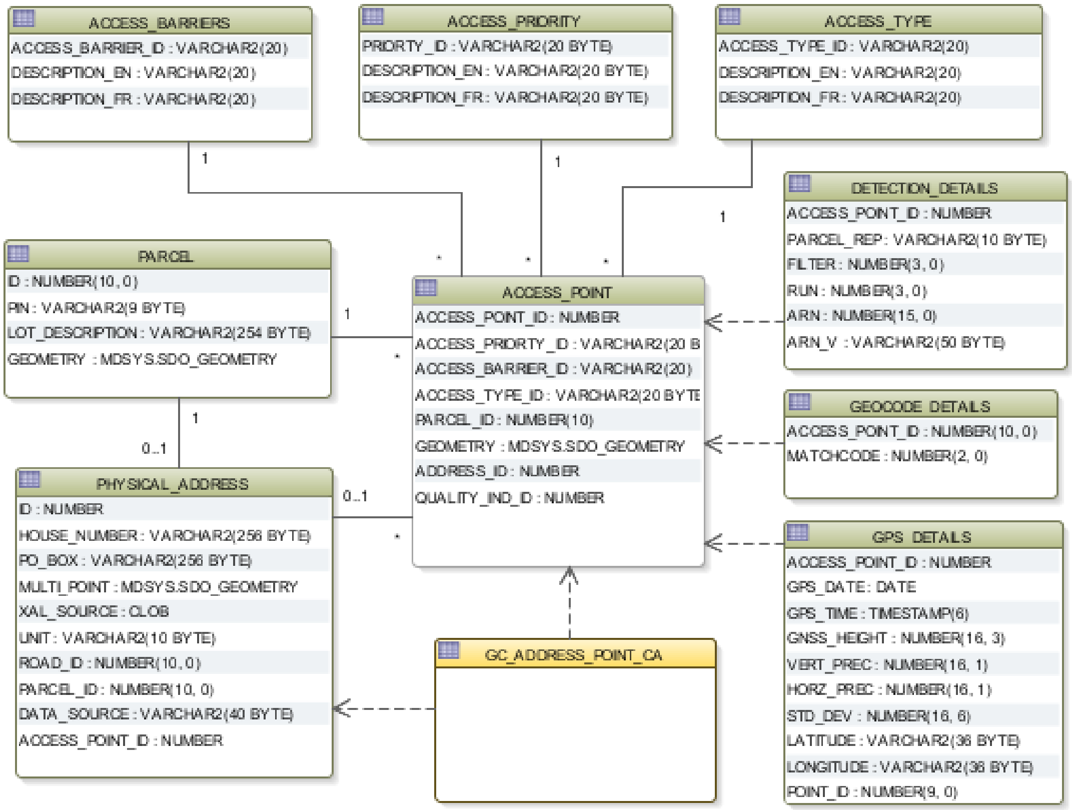

Figure 18 - Access Point

The right-of-way layer prepared earlier in this paper should also be incorporated into the model. The layer would be added as a LOD1 *Transportation_Complex* element and would serve as a framework to build *AuxillaryTrafficArea* and *AuxillaryTrafficArea* elements if needed in the future. The right-of-way could support processes like the automated detection of access points, in the database with supplemental information. More importantly, a right-of-way would be useful to detect candidate topological errors with features representing telecommunications hardware that are currently represented as schematic elements, such overhead wiring and underground cables and ducts, that are expected to be contained in right-of-ways. Further construction of street intersections as a transportation square subclass (OGC, 2012) could be used to ensure that connecting utility infrastructure are contained within intersecting right-of-ways when appropriate.

Finally, the results of the detection process in rural areas highlighted the need to capture access point locations, but also the need to develop a process to assign address to these points when only lot and plan area available. By having the access points in a database, they can act as a cue for data improvements.

5 Conclusions and Discussion

The goal of the research for this paper had to demonstrate that 3D geospatial models based on open standards can simplify and enhance the management of spatial data, specifically with regards to service location and related asset management. This goal was achieved by executing activities that met the stated objectives to:

- develop an understanding of service location and identify data requirements and candidate data sources that would enhance current spatial data holdings in support of service location;
- evaluate and to the extent necessary modify and extend standards-based data models to create an interim 3D service location database;
- evaluate two different methods to collect service location information, one based on geocoding with another based on a detection process using orthoimagery, and assess the results using a statistical comparison to baseline GPS data collected in the field; and
- incorporate collected field service information into a revised final service location model.

The tasks to meet the first objective were executed in Section 2, which defined the scope of service location. Service is the operative word since it implies that a task at a location can only be completed if the related activities can be supported through the deployment of suitable resources. The literature review carried out for this paper demonstrated that the context provided by 3D geospatial information not only enhances service functions, by allowing overhead and subsurface assets to be represented, but in some cases is critical to specific businesses, such as emergency services or utility asset management.

An examination of recent developments in standards, and in particular CityGML, an OGC standard, and IFC, suggests that the convergence of data models can contribute to integrated and interoperable business process. Established and emerging data standards were shown to accommodate service location requirements, and other business functions could no doubt exploit an extended model with technologies such as MSI.

As part of the execution of the methodology in Section 4, a number of open source and COTS applications were shown to support information exchange and consume 3D data products, although the current state of mapping between some models suggests that for the time being only partial exchanges are supported. Continued development of a common frame of reference for IFC and CityGML, as demonstrated by the release of new versions within the time frame of this paper, should promote the active exchange of data in the future through improved element mapping between both standards. The diversity of software options will compel vendors to support the benefits of these open models in order to justify investments in their products by clients. In essence, 3D geospatial data convergence will bring benefits by allowing business units with limited or no previous GIS expertise to apply their core competency using familiar tools, and in the process extend their communication capabilities with other businesses units, business partners, newly acquired companies, and clients, all of which can also contribute to value added activities using the same common framework.

As discussed in the introduction, cultural and technological differences are barriers to convergence. A strategy of openness could expose conflicts. Infrastructure engineering and GIS have common ground but spatial areas of interest and data resolution requirements potentially can create barriers between business units. For example, the maintenance of CAD documents in a local coordinate system requires linkages to geospatial data. The effort to establish linkages is a business process issue because although a technical solution is possible, other issues such as data ownership and project funding might dissuade managers from implementing a solution. This is of course conjecture based on a prudent anticipation of things that can go wrong. To address cultural and technological differences, an enterprise must support the value and culture of 3D geospatial contexts at the project level to achieve success.

The case for data and process integration can be improved by looking at established costs in various sections. A US National Institute of Standards and Technology (NIST) study identified a "loss of $15.8 billion in 2002 in the US capital facilities industry resulting from inadequate interoperability" (Shen, et al., 2010, p. 197). The NIST also identified that the automotive industry wastes over a billion dollars, and about half of that amount results from data exchange issues (Gridwise.org, 2009). Costs associated with the lack of integration and interoperability in the telecommunications industry is difficult to isolate, but one can extrapolate for these examples that they are significant. Considering the potential value of improved integration for these sectors, one can safely assume that there are financial incentives in the telecommunications industry to evaluate processes to improve data integration.

5.1 Reflections on Data Preparation

Certainly, the execution of tasks to meet the second objective, i.e. the creation of an interim service location model, demonstrated that a single target database has the potential to support CAD and GIS, which should in itself bring benefits similar to convergence. This paper also showed that it is possible to build a comprehensive solution, by demonstrating that geocoding applications can occupy the same repository as CAD and 3D GIS data.

The preparation of data in some cases was not a simple process, but the net benefit of integrating parcel, building, and road network data in support of service location processes justified the effort. The bulk of data preparation as demonstrated should be seen as a rare process since data, once loaded, would not need to be similarly processed and if necessary, the utilities developed for this paper or other similar utilities can be re-used. Database models with properly defined relationships and constraints should reduce the risk of data deviating from an established set of values or domains. Although some processing was complex, the complexity associated with the status quo, i.e. islands of data, is not acceptable if a company is to remain competitive and innovative. Taking the steps to improve the data within a single repository will in all likelihood reduce the risk of data degradation and assist in data collection and dissemination activities.

5.2 Reflections on the Collection and Analysis of Field Data

The collection and analysis of service location data, as set out in the third objective, demonstrated that a better product can be developed by incorporating data from various solutions to develop a best-of-breed data solution. A statistical analysis as applied in this research was able to identify preferred sources of service location data for a particular land use. An analytical process could be used to apply a ranking to the results. Although a service location application was demonstrated, a similar approach could be made to incrementally adjust the position of assets that are offset from a true position. A common identifier would be needed for points representing a similar feature, comparable to the use of address identifiers used in this paper.

The process of collecting information in the field revealed that there is value in getting to the core of the problem in a physical sense. Although some of the observations from the field work conducted in this paper could be anticipated, an inventory of issues collected in the field highlighted that requirements gathering has to consider the experience of on-the-ground resources.

5.3 An Assessment of a Final Service Model

Meeting the requirements of the final objective, that is the development of a final service location model, is the one that should provide the most value. An extended model allowed for a separation of activities through Oracle schemas when required but within the framework of a single repository. A separation of activities by schema also responds to one of the benefits of an integrated solution mentioned at the beginning of Section 3, which was that consolidation can also systematically restrict access and limit functionality.

The development of an *Access_Point* object is just one approach to model service location. A single point table, with ranking allows permitted end users to easily see all collected data, while still allowing preferred points to be isolated by rank. Separate point tables for GPS, geocoding, and semi-automated processing could have been made, but if one considers the number of devices and methods for collecting point data outlined in Section 2, then the number of tables could result in a model that is more difficult to manage. The separation of details specific to a methodology allows for metadata to be isolated, reducing the number of unpopulated fields in the core asset point table.

The use of views was adopted in the interim model to bridge between Oracle Geocoder and CityGML models. The use of views created an opportunity to incorporate data required by Oracle Spatial into an enterprise and possibly using existing content. The use of views has been promoted in this paper as a way to build geospatial functionality unobtrusively.

5.4 Data Opportunities

Data to support service location appears to be constantly improving. During this research, additional data sets, such as address points from the City of Ottawa, became available. Parcel data integration is occurring at various government levels, which should improve availability, although differences in parcel definition were apparent with the comparison of BC and Ontario parcel datasets. Robust data sets also provide a mechanism to derive data, such as was demonstrated with the compilation of access points from orthoimagery.

VGI could supplement official content to produce comprehensive data coverage. An OpenStreetMap utility called Osmosis (OpenStreetMap, 2013) can be used to create a POSTGIS repository based on OpenStreetMap, and layers from which can then be selectively channelled to other enterprise solutions through ETL processes. VGI could also be compiled by service workers through handheld devices as part of a service request call. These examples, show the potential use of these data sources, but as discussed in Section 2.1.1.4, these sources of information should not be treated as final but rather as supplemental data. As demonstrated in this paper, these sources of information need a home in the service location model, one that allows for an appropriate assessment with other data that represents similar assets.

Most data for this project was available through a license agreement or as a free download. One temptation might be to ignore data for purchase, but an examination of available data from the City of Ottawa suggests that business processes might be streamlined if such available data sets are incorporated into the business process. Considering the effort in collecting GPS points for the number of points collected, an enterprise should monitor the development of data by private sector data suppliers.

With data representing different spatial entities, this paper highlighted the benefit in using similar data collected by different means to develop candidate features. The same process showed that a reasonable comparison can validate and quantify data. The evaluation also highlighted that such processes could identify gaps in other datasets, as was demonstrated with non-conforming address information associated with parcels and illustrated in Figure 16.

5.5 A Final Thought

This paper hopefully demonstrated that spatial applications should be part of any future data integration process. The cross-domain nature of geography is capable of opening doors to new opportunities. The advantage of the benefits of open standards and data models can only facilitate those opportunities.

5.6 Future Work

In Section 2, a number of devices were itemized that potentially can be used to determine location and contribute to core spatial infrastructure. This paper concentrated on the use of traditional datasets combined with GNSS (GPS) and geocoding. The use of indoor sensors was regrettably excluded. Future work should extend the applications of wireless devices for service location within enclosed spaces, similar to the case study on robotic navigation contained in Annex I of the CityGML encoding specification (OGC, 2012). Outdoor sensors, which cover a much larger spatial extent, should also be researched. LiDAR has become an effective tool in the construction of 3D datasets, and its potential role as a validation tool is one area of future research.

The convergence of standards will offer opportunities to improve on the work outlined in this paper. CityGML and IFC currently allow for a limited set of information exchange. A more intelligent, comprehensive information system is the likely outcome of evolving standards in the near future, and the reliability of such systems will require validation processes.

Earlier in this paper, Oracle Spatial was identified as a core technology due to tight integration with GE Smallworld. This paper did not examine the issues and benefits resulting from a tighter integration between the two technologies with respect to service location. Research using these technologies should be explored.

The potential of linear referencing was not explored as an option in the preparation of datasets. The capabilities of the `SDO_LRS` package in Oracle is substantial and should offer more streamlined methods for data preparation. The incorporation of linear referencing into the data preparation process should generate new opportunities for data and application integration.

Bibliography

Agrawal, G. P. (2010). *Fiber-optic Communication Systems* (4th ed.). New York: Wiley.

Autodesk. (nd). Retrieved November 6, 2012, from http://bimcurriculum.autodesk.com/unit/unit-1-%E2%80%93-bim-modeling-basics

BBC. (2012, November 7). *Twisted light' data-boosting idea sparks heated debate*. Retrieved November 7, 2012, from http://www.bbc.co.uk/news/science-environment-20217938

Bell. (2011). *What is enhanced 9-1-1 (e9-1-1) service*. Retrieved December 20, 2011, from http://support.bell.ca/en-on/Mobility/Rate_plans_features/What_is_Enhanced_911_e_911_service_for_my_mobile_phone?step=6

Bennett, C., & Regan, P. M. (2002, May 30). *What Happens When You Make A 911 Call? Privacy And The Regulation Of Cellular Technology In The United States And Canada*. Retrieved December 17, 2011, from Canadian Political Science Convention, University of Toronto: http://web.uvic.ca/polisci/people/faculty/Bennett/pdf/911Call.pdf

Bokmiller, D., Whitbread, S., & Londenberg, J. (2011). Mastering Autodesk Revit MEP 2012. Hoboken, NJ: John Wiley and Sons.

BuildingSmart. (2013). *Industry Foundation Classes Release 4 (IFC4)*. Retrieved March 13, 2013, from http://www.buildingsmart-tech.org/ifc/IFC4/final/html/

Canada Post. (2005a). *Postal Code Address Data Technical Specifications*. Retrieved December 11, 2012, from http://www.canadapost.ca/cpo/mr/assets/pdf/business/postalcodetechspecs_en.pdf

Canada Post. (2005b). Retrieved September 5, 2012, from http://www.canadapost.ca/cpo/mc/assets/pdf/business/aa_qa_en.pdf

Canada Post. (2012, March 8). *Addressing Guidelines*. Retrieved April 7, 2013, from http://www.canadapost.ca/tools/pg/manual/PGaddress-e.asp

Canada Post. (2013). *Urban and Rural Delivery Area Counts andMaps*. Retrieved April 7, 2013, from http://www.canadapost.ca/cpc2/addrm/hh/doc/about-e.asp

Chan-Olmsted, S., & Jamison, M. (2001). Rivalry through alliances: Competitive strategy in the global telecommunications market. *European Management Journal, 19*(3), 317-331.

City of Ottawa. (2012a). *Profiling Progress - An Introduction to Ottawa*. Retrieved December 4, 2012, from http://ottawa.ca/en/city-hall/accountability-and-transparency/annual-report/2004-city-ottawa-annual-report-0/profiling

City of Ottawa. (2012b). *Land and Climate Statistics, City of Ottawa and Area*. Retrieved December 4, 2012, from http://ottawa.ca/en/city-hall/get-know-your-city/statistics/55-land-and-climate-statistics-city-ottawa-and-area

City of Ottawa. (2012c). *Bilingualism Policy*. Retrieved December 4, 2012, from http://ottawa.ca/en/city-hall/your-city-government/policies-and-administrative-structure/bilingualism-policy

City of Ottawa. (2012d). *Driveways*. Retrieved February 23, 2013, from http://ottawa.ca/en/residents/laws-licenses-and-permits/laws/driveways

City of Ottawa. (2013). *Data Catalogue*. Retrieved April 1, 2013, from http://app06.ottawa.ca/en/city_hall/statisticsdata/opendata/info/index.htm

CityGMLWiki. (2012). *CityGML Immovable Property Taxation ADE*. Retrieved November 24, 2012, from http://www.citygmlwiki.org/index.php/CityGML_Immovable_Property_Taxation_ADE.

Collada.org. (nd). Retrieved from http://collada.org/public_forum/files/COLLADA_Introduction_IGDA.pdf

Colman, G. W. (2001, October 1). *A comparison of the accuracy of TDOA and TOA location algorithms with equivalent receiver geometry*. Retrieved March 29, 2013, from http://cradpdf.drdcrddc.gc.ca/PDFS/unc75/p517521.pdf

CRTC. (2011a). *Telecom Decision CRTC 2011-619*. Retrieved December 16, 2011, from http://www.crtc.gc.ca/eng/archive/2011/2011-619.htm

CRTC. (2011b). *Voice over Internet Protocol (VoIP) 9-1-1 Service*. Retrieved December 18, 2011, from http://www.crtc.gc.ca/eng/info_sht/t1029.htm

CRTC. (2012). *9-1-1 services for traditional wireline, VoIP and wireless phone services*. Retrieved December 5, 2012, from http://www.crtc.gc.ca/eng/info_sht/t1035.htm

CTV News. (2013, February 23). *Beware the maze of underground utilities.* Retrieved from http://www.ctvnews.ca/w5/beware-the-maze-of-underground-utilities-1.1168165

CWTA. (2012). Retrieved from http://cwta.ca/facts-figures/

De Sousa, P. T., & Stuckmann, P. (nd). *Telecommunication Network Interoperability Telecommunication Systems and Technologies, II*. Retrieved from http://www.eolss.net/Sample-Chapters/C05/E6-108-22.pdf

Desmond, C. (2010). *The ComSoc guide to managing telecommunications projects.* Hoboken, N.J.: Wiley through IEEE Xplore (Online service).

Dunn, C. E. (2007). Participatory GIS - A people's GIS? *Progress in Human Geography, 31*(5), 616-637.

El-Mekawy, M. A., Östman, A. B., & Hijazi, I. C. (2011). *An Evaluation of IFC-CityGML Unidirectional Conversion.* Retrieved February 23, 2013, from International Journal of Advanced Computer Science and Applications: http://thesai.org/Downloads/Volume3No5/Paper_25-An_Evaluation_of_IFC_CityGML_Unidirectional_Conversion.pdf

Elwood, S., Goodchild, M. F., & Sui, D. Z. (2012). Researching Volunteered Geographic Information: Spatial Data, Geographic Research, and New Social Practice. *Annals of the Association of American Geographers, 102*(3), 571-590.

ESRI. (2012). *Esri Telecommunications Data Model for ArcGIS 9.3.1.* Retrieved October 6, 2012, from http://www.arcgis.com/home/item.html?id=0ab9a0e3261045e5ac2df30ec3cd7b1b

ESRI. (nd). Retrieved from http://resources.arcgis.com/en/help/main/10.1/004m/pdf/FMEReadersWriters.pdf

Fang, S.-H., & Lin, T. (2012). Principal Component Localization in Indoor WLAN Environments. *Mobile Computing. IEEE Transactions on Mobile Computing, 11*, 1.

FCC. (n.d. a). *Enhanced 9-1-1 - Wireless Services*. Retrieved December 14, 2011, from http://www2.fcc.gov/pshs/services/911-services/enhanced911

FCC. (n.d. b). *Guide - Wireless 911 Services*. Retrieved December 16, 2011, from http://www.fcc.gov/guides/wireless-911-services

Forbes. (2012, September 10). *Google Maps War With Apple Is Over - Google Won*. Retrieved 6 10, 2013, from http://www.forbes.com/sites/tjmccue/2012/09/10/google-maps-war-with-apple-is-over-google-won/

Forbes. (n.d.). *The World's Biggest Public Companies*. Retrieved December 4, 2012, from http://www.forbes.com/global2000/#p_1_s_a0_Software%20&%20Programming_All%20countries_All%20states_

Fröhlich, P., Oulasvirta, A., Baldauf, M., & Nurminen, A. (2011). On the Move, Wirelessly Connected to the World. *Communications of the ACM*, *54.*

GeoBC. (2013). *GeoBC: Integrated Registry Branch*. Retrieved March 30, 2013, from archive.ilmb.gov.bc.ca/irb/Level1/Spatial/ICF/ICF.html

Goleniewski, L. (2007). *Telecommunications Essentials* (2nd ed.). Boston, MA: Addison-Wesley.

Gong, H., Chen, C., Bialostozky, E., & Lawson, C. T. (2011). A GPS/GIS method for travel mode detection in New York City. *Computers, Environment and Urban Systems*. doi:doi:10.1016/j.compenvurbsys.2011.05.003

Goodchild, M. F. (2007). Citizens as Sensors: The world of volunteered geography. *GeoJournal, 69*(4), 211-221.

Goodchild, M. F. (2008). Retrieved May 2012, from http://www.geog.ucsb.edu/~good/papers/435.pdf

Google. (2012). Retrieved October 6, 2012, from https://developers.google.com/kml/documentation/kmlreference

Google. (2013). Retrieved February 4, 2013, from https://developers.google.com/kml/documentation/kmlreference

Gow, G. A., & Ihnat, M. (2004). Prepaid Mobile Phone Service and the Anonymous Caller: Considering Wireless E9-1-1 in Canada. *Surveillance & Society, 1*(4), 555-572.

Gridwise.org. (2009). *Reliability Benefits of Interoperability is a work of the GridWise Architecture Council*. Retrieved October 4, 2012, from http://www.gridwiseac.org/pdfs/reliability_interoperability.pdf

Gröger, G., & Plümer, L. (2012). CityGML – Interoperable semantic 3D city models. *ISPRS Journal of Photogrammetry and Remote Sensing, 71*(Complete), 12-33.

Gröger, G., Kolbe, T. H., & Czerwinski, A. (2007). *Candidate OpenGIS® CityGML Implementation Specification (City Geography Markup Language).* Retrieved from Open Geospatial Consortium Inc.

Grove, N., & Baumann, O. (2012). Complexity in the telecommunications industry: When integrating infrastructure and services backfires. *Telecommunications Policy, 36*(1), 40-50. Retrieved from www.scopus.com

Grubesic, T. H., Matisziw, T. C., & Murray, A. T. (2011). Market Coverage and Service Quality in Digital Subscriber Lines Infrastructure. *International Region Science Review, 34*(3), 368-390. doi:10.1177/060017610386479

Hall, G. B., & Leahy, M. G. (n.d.). Open Source Approaches in Spatial Data Handling Advances. *Geographic Information Science, 2*. doi:10.1007/978-3-540-74831-1

Heipke, C. (2010). Crowdsourcing geospatial data. *ISPRS Journal of Photogrammetry and Remote Sensing, 65*(6), 550-557.

Houseman, J. (1991). Joint Ventures,Strategic Alliances, and Collaboration in Telecommunications. *CATO Review of Business & Government*. Retrieved from http://www.cato.org/sites/cato.org/files/serials/files/regulation/1991/1/v14n1-8.pdf

Huang, S., Chen, C., Hsiao, P., & Fu, L. (2004). On-board vision system for lane recognition and front-vehicle detection to enhance driver's awareness. *IEEE International Conference on Robotics and Automation*, *3*, pp. 2456-2461.

Iansiti, M. (1995). Technology integration: Managing technological evolution in a complex environment. *Research Policy*, 521-542.

Institute for Geodesy and Geoinformation Science. (nd). Retrieved from http://www.3dcitydb.net/index.php?id=1939

ITU. (2005). *The ITU New Initiatives Programme: UbiquitousNetwork Societies*. Retrieved October 5, 2012, from http://www.itu.int/wsis/tunis/newsroom/stats/Ubiquitous_Network_Societies_2004.pdf

Janssen, K., & Crompvoets, J. (2013). Geographic Data and the Law: Defining New Challenges. Leuven University Press. Retrieved from http://books.google.ca/books?id=Jij5xdhLqY4C&pg=PA116&lpg=PA116&dq=LAwsuits+geographic+data&source=bl&ots=rafa6-WkR5&sig=icgVisBQWnGFmFMc37l8bflaf2s&hl=en&sa=X&ei=0asXUc6aFKnvygHd6YGYBQ&ved=0CEoQ6AEwBA#v=onepage&q=LAwsuits%20geographic%20data&f=false

Jensen, J. R., & Cowen, D. C. (1999). Remote sensing of urban/suburb an infrastructure and socio-economic attributes. *Photogrammetric Engineering and Remote Sensing, 65*(5), 611-622.

Jones, A., & Cowan, E. (2007). Convergence. *Information Security Technical Report* , 69-73.

Kolbe, T. H. (2007). Retrieved September 28, 2012, from http://www.citygml.org/fileadmin/citygml/docs/CityGML_Tutorial_Kolbe_Internet.pdf

Kolbe, T. H., Gröger, G., & Plümer, L. (2005). *CityGML – Interoperable Access to 3D City Models*.

Kolbe, T. H., König, G., Nagel, C., & Stadler, A. (2009). *3D-Geo-Database for CityGML*. Retrieved from http://opportunity.bv.tu-berlin.de/software/documents/show/32

Köninger, A., & Bartel, S. (1998). 3D-GIS for Urban Purposes. *Geoinformatica, 2*(2), 79–103.

Koonen, T. (2006). Fiber to the Home/Fiber to the Premises: What, Where, and When? *Proceedings of the IEEE*, (pp. 911-934). doi:10.1109/JPROC.2006.873435

Krutchen, P. (2000). *Thr Rational Unified Process An Introduction* (Vol. 2). (G. Booch, I. Jacobson, & J. Rumbaugh, Eds.) Reading, Massachusetts: Addison-Wesley.

Kwok, Y.-K. R., Lau, & Lau, V. K. (2007). *Wireless Internet and mobile computing: interoperability and performance.* (Wiley-Interscience, Ed.) Hoboken, N.J. doi:ISBN 9780470167960 , 0470167963

Lee, K., & Lee, Y. (2007). A LBS system for cellular phones. *ICCIT 2007*, (pp. 222-228).

Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2011). *Geographic Information Systems and Science* (3 ed.). Hoboken, NJ, USA: John Wiley & Sons, Inc.

Mena, J. B. (2003). State of the art on automatic road extraction for GIS update: A novel classification. *Pattern Recognition Letters, 24*(6), 3037-3058. Retrieved from www.scopus.com

Microsoft. (2012, July 9). Retrieved from http://msdn.microsoft.com/en-us/library/ff647804.aspx

Mladineo, N., Knezic, S., & Jajac, N. (2011). Decision Support System for emergency management on motorway networks. *Transportmetrica, 7*(1), 45-62. doi:10.1080/18128600903244669

Morgan, M. F. (2009). CAD-GIS interoperability issues for facilities management: Enabling inter-disciplinary workflows. (Order No. 1473064, University of South Carolina). ProQuest Dissertations and Theses, 125. Retrieved from http://search.proquest.com.proxy.lib.uwaterloo.ca/docview/304995579.

Murray, A. T., Grubesic, T. H., Wei, R., & Mack, E. A. (2011). A Hybrid Geocoding Methodology for Spatio-Temporal Data. *Transactions in GIS, 15*, 795–809. doi:10.1111/j.1467-9671.2011.01289.x

NATO. (1969). *Software Engineering.* NATO. Brussels: Scientific Affairs Division. Retrieved April 7, 2013, from http://homepages.cs.ncl.ac.uk/brian.randell/NATO/nato1968.PDF

Navteq. (2013). *Develop with Sample NAVTEQ® Map Data and Location Content*. Retrieved from http://sampledata.navteq.com/site/global/developer_resources/sample_data/p_sample_data.jsp#odf

OASIS. (2012a). Retrieved October 6, 2012, from https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=ciq

OASIS. (2012b). *OASIS CIQ V2.0 Committee Specifications / Standards Page*. Retrieved October 6, 2012, from https://www.oasis-open.org/committees/ciq/ciq.html

OASIS. (2012c). Retrieved December 8, 2012, from http://docs.oasis-open.org/election/external/xAL.xsd

OECD. (2009). *Network Developments in support of Innovation and User Needs*. Retrieved October 6, 2012, from http://search.oecd.org/officialdocuments/displaydocumentpdf/?doclanguage=en&cote=dsti/iccp/cisp%282009%292/final

OGC. (2006, June 15). *OpenGIS® Web Map Server Implementation Specification.* Retrieved from http://portal.opengeospatial.org/files/?artifact_id=14416

OGC. (2010, November 02). *OpenGIS Web Feature Service 2.0 Interface Standard.* Retrieved from http://portal.opengeospatial.org/files/?artifact_id=39967

OGC. (2012, April 4). *OGC City Geography Markup Language (CityGML) En-coding Standard.* (T. H. Kolbe, G. Gröger, C. Nagel, & K.-H. Häfele, Editors) Retrieved November 4, 2012, from https://portal.opengeospatial.org/files/?artifact_id=47842

OGC. (2013). *About OGC*. Retrieved March 4, 2013, from http://www.opengeospatial.org/ogc

Ontario Parcel. (n.d. a). *Frequently Asked Questions*. Retrieved March 30, 2013, from http://ontarioparcel.ca/english/index.htm

Ontario Parcel. (n.d. b). *Data Sets*. Retrieved March 30, 2013, from http://ontarioparcel.ca/english/index.html

Ontario Parcel. (n.d. c). *Overview*. Retrieved March 30, 2013, from http://ontarioparcel.ca/english/index.html

OpenStreetMap. (2012, December 29). *Downloading data*. Retrieved 04 18, 2013, from http://wiki.openstreetmap.org/wiki/Downloading_data

OpenStreetMap. (2013, February 16). *Osmosis*. Retrieved March 18, 2013, from http://wiki.openstreetmap.org/wiki/Osmosis

OpenStreetMap. (n.d.). *Copyright and License*. Retrieved March 18, 2013, from http://www.openstreetmap.org/copyright

Oracle. (2009a). *Oracle® Spatial Developer's Guide*. Retrieved December 11, 2012, from SDO_CS Package (Coordinate System Transformation): http://docs.oracle.com/cd/B28359_01/appdev.111/b28400/sdo_cs_ref.htm#CHDBCAEE

Oracle. (2009b). *Oracle® Spatial Developer's Guide 11g Release 1 (11.1)*. Retrieved from Geocoding Address Data: http://docs.oracle.com/cd/B28359_01/appdev.111/b28400/sdo_geocode_concepts.htm

Oracle. (2009c, August 24). *Oracle Forums*. Retrieved from https://forums.oracle.com/forums/thread.jspa?threadID=215037

Oracle. (2009d). *Spatial Concepts.* Retrieved from http://docs.oracle.com/cd/B28359_01/appdev.111/b28400/sdo_intro.htm

Oracle. (2009e). *Geocoding Address Data*. Retrieved from http://docs.oracle.com/cd/B28359_01/appdev.111/b28400/sdo_geocode_concepts.htm#autoId3

Oracle. (2010a, August 12). *Oracle Forums*. Retrieved from https://forums.oracle.com/forums/thread.jspa?messageID=4499110

Oracle. (2010b, August 12). *Oracle Forums*. Retrieved February 12, 2013, from https://forums.oracle.com/forums/thread.jspa?threadID=1113588

Oracle. (2011). *Oracle Spatial 11g Geocoder An Oracle Technical White Paper*. Retrieved October 24 , 2012, from http://download.oracle.com/otndocs/products/spatial/pdf/spatial11gr2_geocoder_twp.pdf

Oracle. (2012). *Oracle® Spatial Developer's Guide 11g Release 2 (11.2)*. Retrieved from Coordinate Systems (Spatial Reference Systems): http://docs.oracle.com/cd/E11882_01/appdev.112/e11830/sdo_cs_concepts.htm

Oracle. (n.d.). *Free and Open Source Software*. Retrieved April 7, 2013, from https://oss.oracle.com/

Plevyak, T., & Sahin, V. (2010). *Next Generation Telecommunications Networks, Services, and Management.* Hoboken, New Jersey: John Wiley and Sons.

PWC. (2006). Technology executive connections - Shaping digital convergence through mergers & acquisitions. *2*. Retrieved from http://www.pwc.com/en_TH/th/publications/assets/techexecconn.pdf

Rana, S., & Joliveau, T. (2009). NeoGeography: An extension of mainstream geography for everyone made by everyone? *Journal of Location Based Services, 3*(2), 75-81.

Region of Waterloo. (2010). *Region of Waterloo Open Data*. Retrieved from http://www.regionofwaterloo.ca/en/regionalgovernment/OpenDataHome.asp

Reichl, W. (2006). Balancing innovation and preservation in telephony. *Elektrotechnik Und Informationstechnik, 123*(7-8), 302-309. Retrieved from www.scopus.com

Rogers. (2011). *911 Information*. Retrieved December 14, 2011, from http://www.rogers.com/web/content/wireless-911?content10=wireless-911

Sadoun, B., & Al-Bayari, O. (2007). LBS and GIS Technology Combination and Applications. *International Conference on Computer Systems and Applications, 2007. AICCSA '07*, *13-16*, pp. 578-583. doi:doi:10.1109/AICCSA.2007.370940

Safe Software. (2011a). Retrieved October 17, 2012, from http://fmepedia.safe.com/articles/Samples_and_Demos/Creating-and-Validating-XML-Documents-with-the-XMLTemplater-and-XMLValidator-Transformers

Safe Software. (2011b). Retrieved from http://docs.safe.com/fme/2011/html/FME_ReadersWriters/citygml/tutorial/writing_citygml_from_fme.htm

Salminen, A., & Tompa, F. (2011). *Communicating with XML.* New York: Springer.

Saroiu, S., & Wolman, A. (2010). *I am a sensor, and I approve this message*. Retrieved from http://research.microsoft.com/pubs/136533/tenor.pdf

Schulzrinne, H., & Arabshian, K. (2002). Providing emergency services in internet telephony. *IEEE Internet Computing, 6*(3), 39-47.

Seiber, R. (2006). Public participation geographic information systems: A literature review and framework. *Annals of the Association of American Geographers, 96*(3), 491-507.

Shaik, M., Das, O., Zhao, L., & Liao, Z. (2009). Inter-vehicle range smoothing for NLOS condition in the persistence of GPS outages. *4th IEEE Conference on Industrial Electronics and Applications*, 3904-3909, 25-27. doi:10.1109/ICIEA.2009.5138939

Shargal, M., & Houseman, D. (2009). *Smart Grid News, "The Big Picture of Your Coming Smart Grid".* Retrieved from www.smartgridnews.com/artman/publish/commentary/The_Big_Picture_of_Your_Coming_Smart_Grid-529.html

Shen, W., Hao, Q., Helium, M., Neelamkavil, J., Dickenson, J., Thomas, R., . . . Xue, H. (2010). Systems integration and collaboration in architecture, engineering, construction, and facilities management: A review. *Advanced Engineering Informatics, 24*, 196-207.

Space Daily. (2009, January 19). *Alternative Positioning Technologies Will Provide 25% Of All Positioning Solutions By 2014*.

State of New York. (2011). *Statewide Communications Interoperability Plan*. Retrieved October 5, 2012, from http://www.dhses.ny.gov/oiec/documents/2011-NYS-SCIP.pdf

Technology Review. (2013, January 4). *2013: The year of the Internet of Things*. Retrieved February 5, 2013, from http://www.technologyreview.com/view/509546/2013-the-year-of-the-internet-of-things/

Telus. (2011). Retrieved December 13, 2011, from http://www.telusmobility.com/en/ON/Being_mobile/calling911.shtml

Teranet. (2013). *About Teranet*. Retrieved June 09, 2013, from http://www.teranet.ca/our-company/about-teranet

Thiagarajan, A., Biagioni, J., Gerlich, T., & Eriksson, J. (2010). Cooperative transit tracking using smart-phones. *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems*, (pp. 85-98). Zurich, Switzerland.

Thill, J.-C., Dao, T. H., & Zhou, Y. (2011). Traveling in the three-dimensional city: applications in route planning, accessibility assessment, location analysis and beyond. *Journal of Transport Geography, 19*(3), 405-421.

Trevisani, E., & Vitaletti, A. (2004). Cell-ID location technique, limits and benefits: an experimental study. *Mobile Computing Systems and Applications*, 51-60. doi:10.1109/MCSA.2004.9

Unsalan, C., & Sirmacek, B. (2012). Road network detection using probabilistic and graph theoretical methods. *IEEE Transactions on Geoscience and Remote Sensing, 50*(11 PART1), 4441-4453. Retrieved from www.scopus.com

Van Berlo, L., & de Laat, R. (2011). Integration of BIM and GIS: The development of the CityGML GeoBIM extension Advances. *3D Geo-Information Sciences*. doi:ISBN 978-3-642-12669-7

van Nederveen, G. A., & Tolman, F. P. (1992). Modelling multiple views on buildings. *Automation in Construction, 1*(3), 215-224. Retrieved from www.scopus.com

W3C. (2006). *Character encodings*. Retrieved November 4, 2012, from http://www.w3.org/International/O-charset.en.php

W3C. (2010). *Transformation*. Retrieved September 25, 2012, from http://www.w3.org/standards/xml/transformation

Wang, R., Hu, Y., & Zhang, X. (2005). Extraction of road networks using pan-sharpened multispectral and panchromatic quickbird images. *Geomatica, 59*(3), 263-273.

Welfens, P. J. (2008). *Digital integration, growth and rational regulation.* Berlin: Springer.

Wireless News. (2009, January 20). *ABI Research Examines Alternative Positioning Technologies*.

APPENDIX A – Code to Populate Road Segments

This stored procedure takes data from a temporary road segments layer to produce a record in *Road*, *Address_Range* and *Transportation_Complex* layers in the interim model.

```
CREATE PROCEDURE POPULATECITY
AUTHID CURRENT_USER AS
  CURSOR c1 IS SELECT street_name, street_type, street_dir,
                      left_from, left_to,
                      right_from, right_to, geom
                 FROM road_segments;

  v_streetname       VARCHAR2(50);
  v_street_type      VARCHAR2(50);
  v_street_dir       VARCHAR2(50);
  v_left_from        NUMBER;
  v_left_to          NUMBER;
  v_right_from       NUMBER;
  v_right_to         NUMBER;
  v_geom             road_segments.geom%TYPE;
  v_envelope         SDO_GEOMETRY;
  v_cityobjectID     cityobject.id%TYPE;
  v_transcomplexID   transportation_complex.id%TYPE;
  v_roadID           street.street_id%TYPE;
  v_munID            municipality.id%TYPE;
  v_provID           province.id%TYPE;
  v_count            NUMBER := 0;
  v_count2           NUMBER := 0;
  v_x                NUMBER(10,5);
  v_y                NUMBER(10,5);
  err_msg            VARCHAR2(1000);
BEGIN
DBMS_OUTPUT.ENABLE(1000000);
OPEN C1;
    LOOP
       FETCH c1 INTO v_streetname, v_street_type, v_street_dir,
            v_left_from, v_left_to, v_right_from, v_right_to, v_geom;
       EXIT WHEN c1%NOTFOUND;
       BEGIN
          SELECT cityobject_seq.nextval INTO v_cityObjectID FROM dual;
          v_roadID := null;
          SELECT p.id into v_provID
            FROM province p WHERE p.name_en = 'ONTARIO';
          SELECT m.id into v_munID
            FROM municipality m
           WHERE m.name_en = 'OTTAWA'
             AND m.province_id = v_provID AND retired IS NULL;
          BEGIN
             SELECT street_id INTO v_roadID
               FROM street r, municipality m
              WHERE NAME = TRIM(TRIM(v_streetname) || ' ' ||
                    TRIM(v_street_type) || ' ' || TRIM(v_street_dir))
                AND r.municipality_ID = m.id
                AND m.province_id = v_provID
                AND m.id = v_munID;
```

```
          EXCEPTION WHEN OTHERS THEN
             raise_application_error(-20101, 'WARNING: Could not find street
in address range loader');
          END;
          IF v_roadID IS NULL THEN
             SELECT street_seq.NEXTVAL INTO v_roadID FROM dual;
             INSERT INTO road(road_id, name, base_name, municipality_id,
                  parent_area_id, lang_code, type_suffix_id, dir_suffix_id)
             VALUES( v_roadID, TRIM(TRIM(v_streetname) || ' ' ||
                  TRIM(v_street_type) || ' ' || TRIM(v_street_dir)),
                  v_streetname,
                  v_munid, v_provid, 'ENG',
                  (SELECT NVL(formal_id, 5)
                     FROM street_types WHERE type_en = v_street_type),
                  (SELECT NVL(formal_id, 2)
                     FROM street_directions WHERE dir_en = v_street_dir));
             v_count := v_count + 1;
          END IF;
          INSERT INTO cityobject (id, class_id, gmlid, gmlid_codespace,
                 envelope, creation_date, termination_date,
                 last_modification_date, updating_person,
                 reason_for_update, lineage, xml_source)
            VALUES(v_cityobjectID, 45, 'UUID_' || sys_guid(), 'UUID', null,
                 SYSDATE, null, SYSDATE, user, 'Initial insert', '', null);
          SELECT transportation_complex_SEQ.NEXTVAL
            INTO v_transcomplexID FROM dual;
          INSERT INTO transportation_complex (id, name,
                  name_codespace, description, function,
                  usage, type, lod0_network)
            VALUES (v_cityobjectID, ',','','','','',
               SDO_CS.MAKE_3D(v_geom, null));
          INSERT INTO address_range(transportation_complex_id,
               from_left, to_left, from_right, to_right, street_id, geometry)
            VALUES(v_cityobjectid, v_left_from, v_left_to,
               v_right_from, v_right_to, v_roadid,
               SDO_CS.TRANSFORM(v_geom, 4326));
       EXCEPTION WHEN OTHERS THEN
           err_msg := sqlerrm;
           raise_application_error(-20100, 'General error in populating
address ranges');
       END;
    END LOOP;
CLOSE c1;
DBMS_OUTPUT.PUT_LINE('Inserted ' || v_count );
COMMIT;
END populatecity;
```

APPENDIX B – Stored Procedure to Prepare Road Names

This stored procedure was used to parse parcel address information, so that parcel records could be matched to sample addresses. This process was executed against a staging table in the interim model.

```
CREATE PROCEDURE PREPARESTREETNAME
AUTHID CURRENT_USER AS
  CURSOR c1 IS SELECT parcel_rep
                 FROM ottawa_parcels
                WHERE NOT street IS NULL
                  AND new_streetname IS NULL;
  v_parcel_rep     ottawa_parcels.parcel_rep%TYPE;
  v_count          NUMBER := 0;
BEGIN

OPEN c1;
LOOP
 FETCH c1 INTO v_parcel_rep;
 EXIT WHEN c1%NOTFOUND;
 v_count := v_count + 1;
 UPDATE ottawa_parcels r1
    SET new_streetname = TRIM(NVL((SELECT REGEXP_REPLACE(street,
                         rd.dir_en, '', 1,
                         REGEXP_COUNT(street, rd.dir_en))
                         FROM street_directions rd
                         WHERE TRIM(SUBSTR(r1.street,
                         INSTR(r1.street,' ',-1,1)))= rd.dir_en),street)),
            street_dir = (SELECT REGEXP_REPLACE(street,
                         SUBSTR(r1.street,0,INSTR(r1.street,'',-1,1)),'')
                         FROM street_directions rd
                         WHERE TRIM(SUBSTR(r1.street,
                         INSTR(r1.street,'',-1,1))) = rd.dir_en)
  WHERE parcel_rep = v_parcel_rep;

 UPDATE ottawa_parcels r1
    SET new_streetname = NVL((SELECT REGEXP_REPLACE(new_streetname,
                         SUBSTR(r1.new_streetname,
                         INSTR(r1.new_streetname,'', -1,1)),'')
                         FROM street_types rt
                          WHERE TRIM(
                         SUBSTR(r1.new_streetname,
                         INSTR(r1.new_streetname,'',-1,1))) =
                         rt.type_en), new_streetname),
            street_type = (SELECT REGEXP_REPLACE(new_streetname,
                         SUBSTR(r1.new_streetname,
                         0,INSTR(r1.new_streetname,'',-1, 1)),'')
                         FROM street_types rt
                         WHERE TRIM(SUBSTR(r1.new_streetname,
                         INSTR(r1.new_streetname,'',-1,1))) = rt.type_en)
  WHERE parcel_rep = v_parcel_rep;
END LOOP;
COMMIT;
END preparestreetname;
```

APPENDIX C – Python Script to Identify Parcel Access Points

This Python script carries out the identification of parcel based service locations for the sample areas outlined in this paper. Each sample area is defined by an image mosaic, and thus the mosaicName argument is used to identify which sample property frontages are to be processed. The script takes advantage of data preparation activities outlined in Sections 4.2.3.1 and 4.2.3.2. Lines are broken for document formatting purposes only.

```
import arcpy
from arcpy import env
import sys
import os
import arcgisscripting
import time

gp = arcgisscripting.create()
arcpy.env.overwriteOutput = True

gp.CheckInExtension("spatial")
if gp.CheckExtension("spatial") == "Available":
   gp.CheckOutExtension("spatial")

mosaicName = gp.GetParameterAsText(0)
runID = gp.GetParameterAsText(1)
parcelLayer = mosaicName + "PropertyFrontsReprojected"

GRIDCODE = "1"
mxd = arcpy.mapping.MapDocument("CURRENT")
arcpy.env.workspace = "mosaicName + ".gdb"

accessways = "accessways"

counter = 0
for row in gp.SearchCursor(parcelLayer):
   counter = counter + 1
   objIntValue = row.getValue("OBJECTID")
   fcIntValue = row.getValue("Parcel_Rep")
   fcArnIntValue = row.getValue("ARN")
   fcLength = row.getValue("Shape_Length")
   objValue =  str(objIntValue)
   fcValue =  str(fcIntValue)
   fcArnValue = str(fcArnIntValue)
   gp.addMessage(str(counter) + ") Checking " + fcValue)
   whereClause = "OBJECTID = " + objValue
   boundaryFC =  "Boundary_" + objValue
   boundaryImage = "BoundaryMask_" + objValue
   boundaryPoints = "BoundaryPts_" + objValue
   boundaryPolygons = "BoundaryPolys_" + objValue
   # Extract from image using buffer
   gp.addMessage("Making layer " + mosaicName + ", " + boundaryImage)
   if gp.Exists("test_" + objValue):
      arcpy.Delete_management("test_" + objValue)
   gp.MakeFeatureLayer_management (parcelLayer, "test_" + objValue,
```

```
whereClause)
   try:
      if gp.Exists(boundaryPolygons):
         gp.addMessage("Deleting boundary polygons " + boundaryPolygons )
         gp.Delete_Management(boundaryPolygons)
      if gp.Exists(boundaryPoints):
         gp.addMessage("Deleting boundary points " + boundaryPoints )
         gp.Delete_Management(boundaryPoints)
      gp.Clip_analysis(mosaicName + "ClassifiedFrontagePolygons", "test_" +
objValue, boundaryPolygons, "")
      addLayer = "ParcelPolygon_" + objValue
      gp.addMessage("Creating layer...")
      filter = 20
      arcpy.MakeFeatureLayer_management (boundaryPolygons, addLayer)
      gp.addMessage("Selecting areas over 20 square metres...")
      condition = "Shape_Area > 20 AND grid_code = " + GRIDCODE + " AND
                   Shape_Length < 50"
      arcpy.SelectLayerByAttribute_management(addLayer,
                   "NEW_SELECTION", condition)
      result = int(arcpy.GetCount_management(addLayer).getOutput(0))
      if result == 0:
         filter = 16
         gp.addMessage("Selecting areas over 16 square metres...")
         condition = "Shape_Area > 16 AND grid_code = " + GRIDCODE + " AND
                      Shape_Length < 45"
         arcpy.SelectLayerByAttribute_management(addLayer,
                      "NEW_SELECTION", condition)
         result = int(arcpy.GetCount_management(addLayer).getOutput(0))
         if result == 0:
            filter = 12
            gp.addMessage("Selecting areas over 12 square metres...")
            condition = "Shape_Area > 12 AND Shape_Area < 45 AND
                grid_code = " + GRIDCODE + " AND Shape_Length < 40"
            arcpy.SelectLayerByAttribute_management(addLayer,
                "NEW_SELECTION", condition)
            result = int(arcpy.GetCount_management(addLayer).getOutput(0))
            if result == 0:
               filter = 10
               gp.addMessage("Selecting areas over 10 square metres...")
               condition = "Shape_Area > 10 AND Shape_Area < 40
                  AND grid_code = " + GRIDCODE + " AND Shape_Length < 30"
               arcpy.SelectLayerByAttribute_management(
                  addLayer, "NEW_SELECTION", condition)
               result = int(arcpy.GetCount_management(
                  addLayer).getOutput(0))
               if result == 0:
                  filter = 24
                  gp.addMessage("Selecting areas over 24 square metres...")
                  condition = "Shape_Area > 20 AND Shape_Area < 24
                    AND grid_code = " + GRIDCODE + " AND Shape_Length < 65"
                  arcpy.SelectLayerByAttribute_management(
                    addLayer, "NEW_SELECTION", condition)
                  result = int(arcpy.GetCount_management(
                    addLayer).getOutput(0))
      gp.addMessage("The process found " + str(result) + "
          candidate polygon(s).")
      if result > 0:
```

```
         gp.addMessage("Preparing " + str(result) + "
            candidate polygon(s) for " + fcArnValue)
         arcpy.FeatureToPoint_management(addLayer,
            boundaryPoints,"INSIDE")
         arcpy.AddField_management(boundaryPoints, "Parcel_Rep",
             "LONG", "", "", "")
         arcpy.AddField_management(boundaryPoints, "ARN",
             "TEXT", "", "", 50)
         arcpy.AddField_management(boundaryPoints, "COUNT",
             "SHORT", "", "", "")
         arcpy.AddField_management(boundaryPoints, "Filter",
             "LONG", "", "", "")
         arcpy.AddField_management(boundaryPoints, "RUN",
             "SHORT", "", "", "")
         gp.addMessage("Calculating values...")
         arcpy.CalculateField_management(boundaryPoints,
             "Parcel_Rep", fcValue, "PYTHON")
         arcpy.CalculateField_management(boundaryPoints,
             "COUNT", result, "PYTHON")
         arcpy.CalculateField_management(boundaryPoints,
             "Filter", filter, "PYTHON")
         arcpy.CalculateField_management(boundaryPoints,
             "RUN", runID, "PYTHON")
         arcpy.CalculateField_management(boundaryPoints,
             "ARN", "\"" + fcArnValue + "\"", "PYTHON")
         gp.addMessage("Appending " + str(result) + " point(s)")
         arcpy.Append_management(boundaryPoints,
             accessways,"NO_TEST","","")
   except:
      e = sys.exc_info()[0]
      gp.addMessage( "Error: %s" % e )
gp.CheckInExtension("spatial")
del mxd
```

APPENDIX D – SQL to Create Offset Lines for Quality Assessment

The following code created a temporary spatial object in Oracle that was used to render differences in the rural sample area between a GPS point and a point derived from image processing as a line segment. Similar code was created for geocoded points. The complex code accommodates 3D and 2D coordinate conversion, the basics of which are outlined under Example 6-4 at:

http://docs.oracle.com/cd/E16338_01/appdev.112/e11830/toc.htm

Since there were potentially more than one point per parcel, the shortest distance between GPS point and detected point was selected for comparison. SDO functions were not incorporated because they required some processing that would have been complicated in a single create table statement. In addition, some SDO functions require a spatial index, which could only be applied after layer creation.

The vectors created by this routine are 2D, but the SDO_ORDINATE_ARRAY constructor can be modified to produce a 3D vector by extracting the Z value from the respective 3D SDO_POINT objects.

This version reflects the creation of a table using the final model.

```
EXEC geocode('Suburban');
DROP TABLE suburbangeocoding;
CREATE TABLE suburbangeocoding AS
  SELECT access_point_id,
         SDO_CS.TRANSFORM(
           SDO_GEOMETRY(2002, 4617, NULL, SDO_ELEM_INFO_ARRAY(1,2,1),
                SDO_ORDINATE_ARRAY(pt.gps_geom.sdo_point.x ,
                                   pt.gps_geom.sdo_point.y,
                                   pt.geocoding_geom.sdo_point.x,
                                   pt.geocoding_geom.sdo_point.y)),
                32189) offsetLine,
         geocoded_address_id,
         gps_address_id
    FROM (SELECT geocoded_point.access_point_id access_point_id,
                sdo_cs.make_2d(geocoded_point.geometry) geocoding_geom,
                sdo_cs.make_2d(gps_point.geometry) gps_geom,
                geocoded_point.address_id geocoded_address_id,
                gps_point.address_id gps_address_id
           FROM access_point geocoded_point, access_point gps_point
          WHERE geocoded_point.DESCRIPTION = 'Geocode: Suburban'
            AND geocoded_point.address_id = gps_point.address_id
            AND gps_point.DESCRIPTION LIKE 'GPS%'
            AND gps_point.priority_id = 1) pt;

DROP TABLE suburbandetection;

UPDATE access_point ap
  SET ap.address_id = (SELECT p.address_id
                         FROM detection_details dd, ottawa_parcels p
                        WHERE dd.parcel_rep = p.parcel_rep
                          AND ap.access_point_id = dd.access_point_id
                          AND run = 120
```

```
                          AND p.address_id is not null)
 WHERE address_id IS NULL
   AND description = 'Detection Process';

COMMIT;

CREATE TABLE suburbandetection AS
  SELECT rownum id,
         gps_address_id,
         gps_id,
         detected_id,
         pt.detected_geom.sdo_point.x x,
         pt.gps_geom.sdo_point.y y,
         SDO_CS.TRANSFORM(SDO_GEOMETRY(2002, 4617, NULL,
            SDO_ELEM_INFO_ARRAY(1,2,1),
            SDO_ORDINATE_ARRAY(
              pt.gps_geom.sdo_point.x, pt.gps_geom.sdo_point.y,
              pt.detected_geom.sdo_point.x,
              pt.detected_geom.sdo_point.y)),
         32189) offsetLine,
    FROM (SELECT detected_point.access_point_id detected_id,
                 gps_point.access_point_id gps_id,
                 sdo_cs.make_2d(detected_point.geometry) detected_geom,
                 sdo_cs.make_2d(gps_point.geometry) gps_geom,
                 gps_point.address_id gps_address_id
            FROM access_point detected_point,
                 access_point gps_point,
                 detection_details dd
           WHERE detected_point.address_id = gps_point.address_id
             AND detected_point.access_point_id = dd.access_point_id
             AND run = 120
             AND gps_point.description LIKE 'GPS%'
         ) pt;

CREATE INDEX sub_detect_dpnt_spx ON suburbandetection("DETECTEDPOINT")
       INDEXTYPE IS MDSYS.SPATIAL_INDEX;
CREATE INDEX sub_detect_gpnt_spx ON suburbandetection("GPSPOINT")
       INDEXTYPE IS MDSYS.SPATIAL_INDEX;
CREATE INDEX sub_detect_offset_spx ON suburbandetection("OFFSETLINE")
       INDEXTYPE IS MDSYS.SPATIAL_INDEX;

DELETE FROM suburbandetection i0
WHERE NOT id IN (
   SELECT id
     FROM suburbandetection i1
    WHERE SDO_GEOM.SDO_LENGTH(offsetLine, 0.05) = (
          SELECT MIN(SDO_GEOM.SDO_LENGTH(offsetLine, 0.05))
            FROM suburbandetection i2
           WHERE i1.gps_address_id = i2.gps_address_id));
COMMIT;
```

APPENDIX E – Bulk Geocoding

The following is code used to generate bulk geocoding. This version of the bulk geocoder was executed using the interim model.

```
CREATE PROCEDURE GEOCODE(v_sample_area VARCHAR2)
AUTHID CURRENT_USER AS
  v_hn                ottawa_parcels.house_number%TYPE;
  v_strt              ottawa_parcels.new_streetname%TYPE;
  v_fullstrt          VARCHAR2(100);
  v_typ               street_types.type_en%TYPE;
  v_street_id         ottawa_parcels.street_id%TYPE;
  v_address           VARCHAR2(100);
  v_acspt_id          access_point.access_point_id%TYPE;
  v_matchcode         NUMBER;
  v_dir               street_directions.dir_en%TYPE;
  v_lat               NUMBER(10,6);
  v_long              NUMBER(10,6);
  v_point             SDO_GEOMETRY;
  v_municipality_id   NUMBER := 90; /* placeholder */
  CURSOR c1 IS
    SELECT civic_number, streetname, streettype, streetdir,
           p.access_point_id, street_id
      FROM access_point p, sample_areas sa, gps_details d
     WHERE SDO_ANYINTERACT(p.geometry, sa.geometry) = 'TRUE'
       AND sa.name = v_sample_area
       AND priority_id = 1
       AND p.access_point_id = d.access_point_id;
BEGIN
   OPEN c1;
   LOOP
      FETCH c1 INTO v_hn, v_strt, v_typ, v_dir, v_acspt_id, v_street_id;
      EXIT WHEN c1%NOTFOUND;
      BEGIN
         v_fullstrt := TRIM(v_street || ' ' || v_type || ' ' || v_dir);
         v_address := TRIM(v_hn || ' ' || v_fullstreet);
         SELECT x.results.latitude, x.results.longitude,
                   x.results.matchcode INTO v_lat, v_long, v_matchcde FROM(
                   SELECT SDO_GCDR.GEOCODE('',SDO_KEYWORDARRAY(v_addr,
                      'OTTAWA'),'CA','DEFAULT')
                      AS results FROM dual) x;
         v_point := SDO_GEOMETRY(3001, 4955, SDO_POINT_TYPE(v_long,
                  v_lat, 100), NULL, NULL);
         INSERT INTO access_point(civic_number,street_id,geometry,
description)
              VALUES(v_hn,v_street_id,v_point,'Geocode: '||v_sample_area);
         INSERT INTO geocode_details(access_point_id, matchcode)
              VALUES(v_acspt_id, v_matchcde);
      END;
   END LOOP;
   CLOSE c1;
   COMMIT;
END geocode;
```

APPENDIX F – Interim Model

An examination of CityGML and Oracle Geocoder resulted in an interim model used to support service location issues. The interim model established relationships between objects; however the cardinality on relationships were very loosely defined.

Entities with a dark brown title bar represent CityGML objects, with some slight modifications. The entities with an orange title bar represent Oracle Spatial geocoding tables, with many represented as views. Entities with a sage title bar are additional objects that were added to support service location.

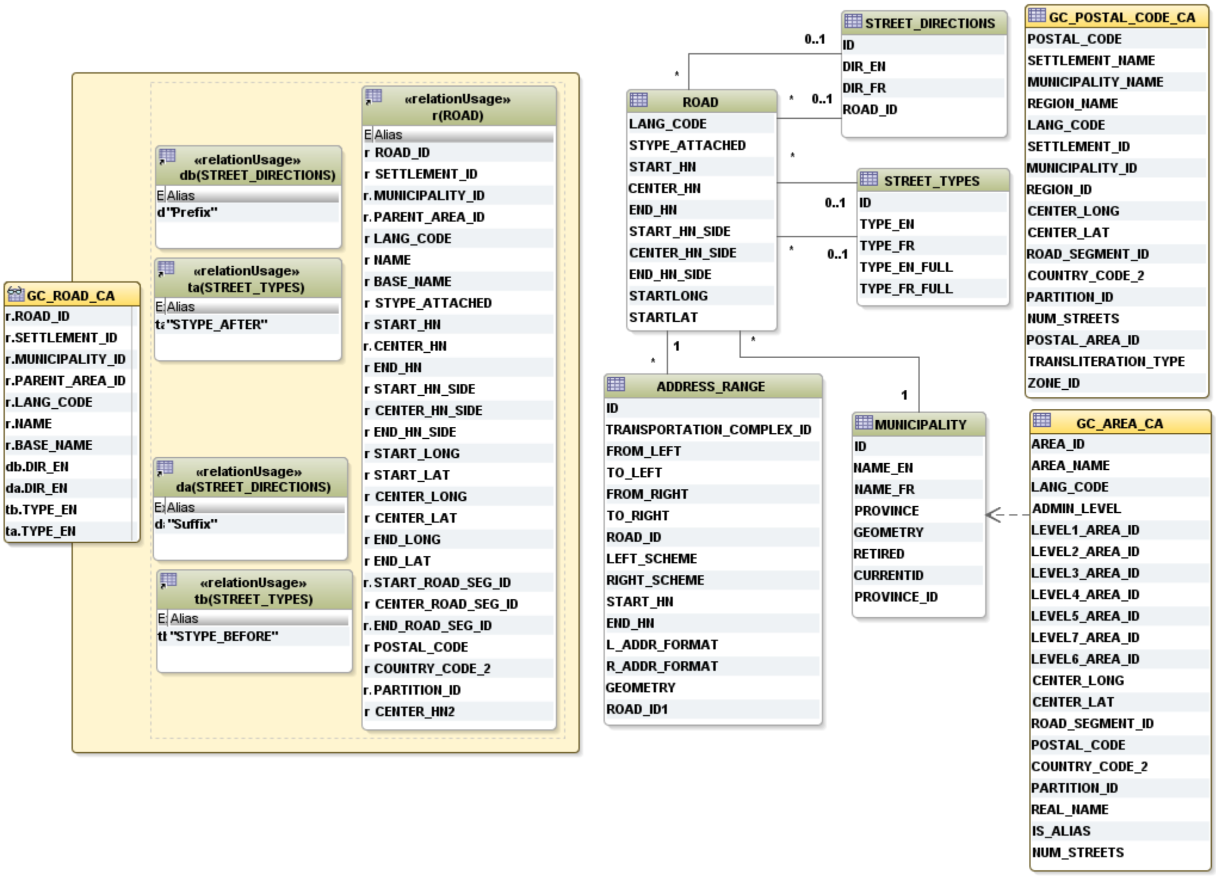

Figure 19 - Adding Road Components to Support Geocoding

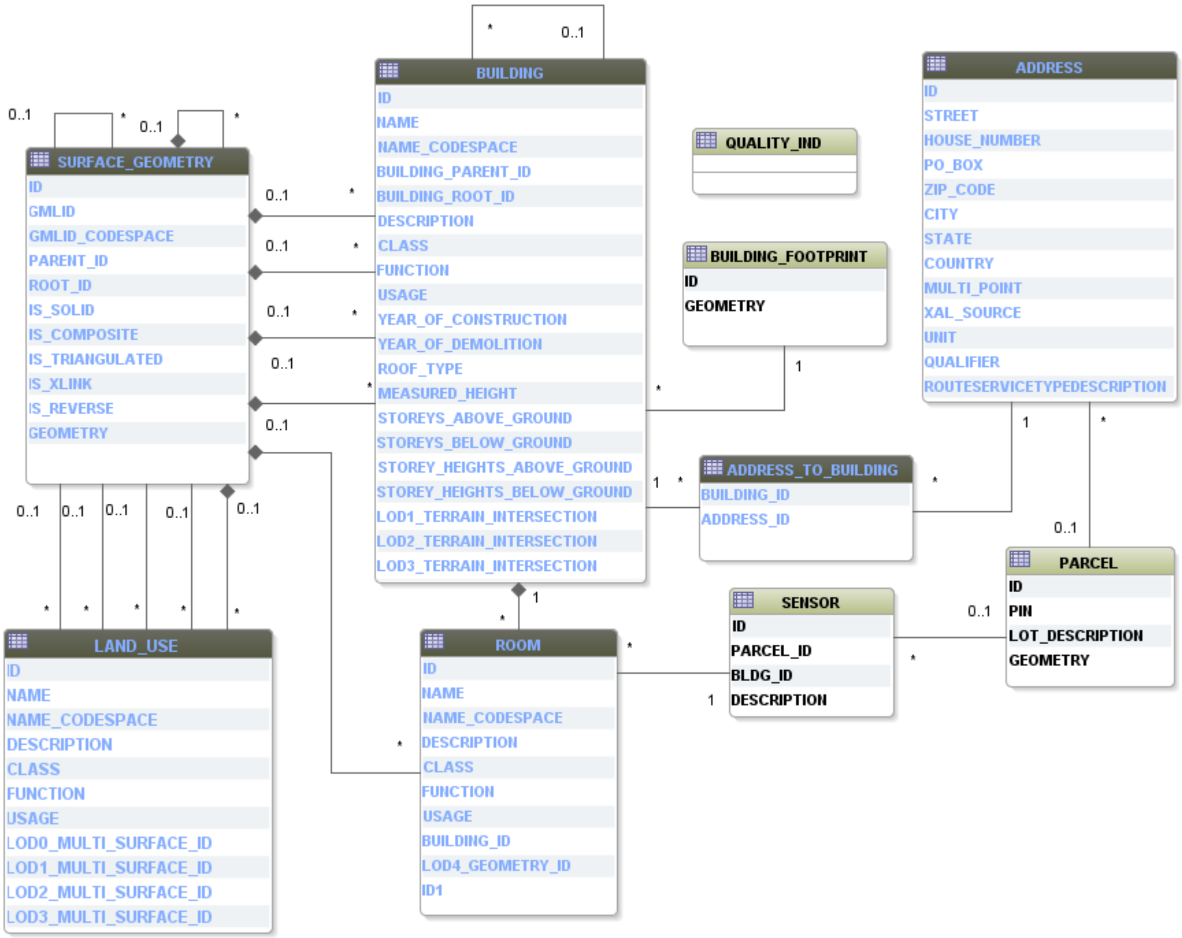

Figure 20 - Extensions to CityGML Buildings to support the Oracle Geocoder

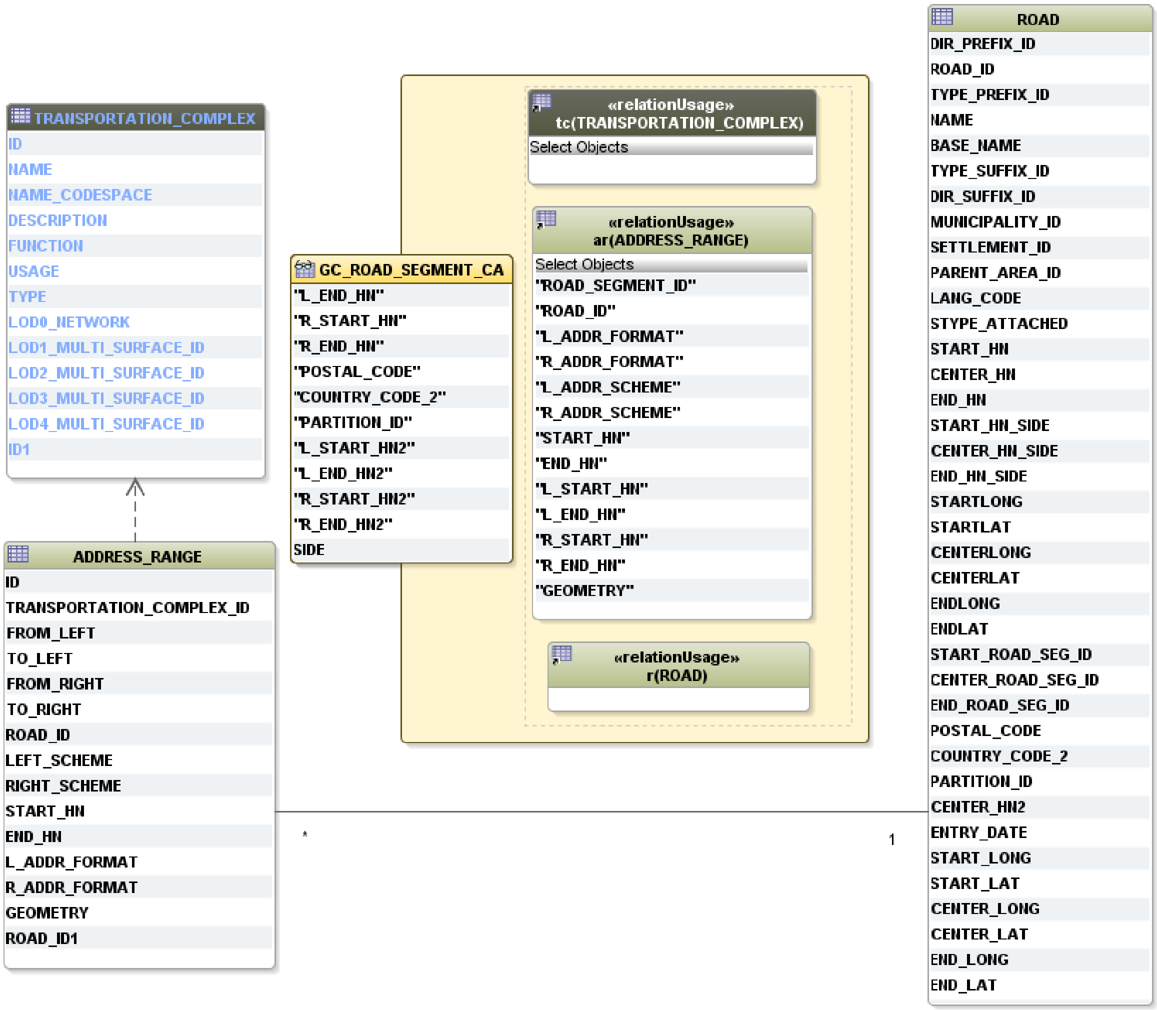

Figure 21 - Extensions to CityGML Transportation Complex to support the Oracle Geocoder

APPENDIX G – Insert Triggers

Insert triggers were placed on tables to generate unique keys using predefined sequences or to generate content based on content in other fields. The first is an example of a key generator using sequences, and the source for which is commonly found in Oracle documentation or through on-line forums.

```
create or replace TRIGGER ACCESS_POINT_TGR
before insert on ACCESS_POINT
for each row begin
   if inserting then
      if :NEW."ACCESS_POINT_ID" is null then
         select ACCESS_POINT_SEQ.nextval into :NEW."ACCESS_POINT_ID"
           from dual;
      end if;
   end if;
end;
```

The next trigger is an example of an insert trigger that derives content from other fields, in this case required content for geocoding. Null values for address range values could be accommodated using the `NVL` function in Oracle.

```
create or replace TRIGGER address_range_before_tgr
BEFORE INSERT ON address_range FOR EACH ROW
DECLARE
         v_leftscheme number;
         v_rightscheme number;
BEGIN
  v_leftscheme := mod(:new.from_left, 2) + mod(:new.to_left, 2);
  v_rightscheme := mod(:new.from_right, 2) + mod(:new.to_right, 2);
  :new.left_scheme := case v_leftscheme WHEN 2 THEN 'O' WHEN 1 THEN 'M'
  WHEN 0
     THEN 'E'
     ELSE null
  END;
  :new.right_scheme :=
     CASE v_rightscheme
        WHEN 2 THEN 'O'
        WHEN 1 THEN 'M'
        WHEN 0 THEN 'E'
        ELSE null
     END;
  IF NOT v_leftscheme IS NULL THEN
     :new.l_addr_format := 'N';
  end if;
  IF NOT v_rightscheme IS NULL THEN
     :new.r_addr_format := 'N';
  END IF;
  :new.start_hn :=
      least(:new.from_left, :new.to_left, :new.from_right, :new.to_right);
  :new.end_hn:=
      greatest(:new.from_left,:new.to_left,:new.from_right,:new.to_right);
END;
```