DATA INTEROPERABILITY ACROSS BORDERS: A CASE STUDY OF THE ABBOTSFORD-SUMAS AQUIFER (BRITISH COLUMBIA / WASHINGTON STATE)

by

Aparna Deshpande
M.Sc. (Geology) University of Pune, India

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE

In the Department

of
Geography

SIMON FRASER UNIVERSITY

March 2004

APPROVAL

Name: **Aparna Deshpande**

Degree: Master of Science

Title of Thesis: Data Interoperability Across Borders: A Case Study Of The Abbotsford-Sumas Aquifer (British Columbia/Washington State)

Examining Committee:

Chair: J. Hyndman, Associate Professor

Dr. N. Schuurman, Assistant Professor
Senior Supervisor

Dr. D. Allen, Associate Professor
Department of Earth Sciences
Committee Member

Dr. K.D. Rothley, Assistant Professor
School of Resource & Environmental Management, SFU
External Examiner

Date Approved: **March 31, 2004**

ABSTRACT

The ability to integrate data from multiple sources is central to geographic information science (GIS). Although data integration is an active field of research in the GIS community, a number of challenges remain unresolved. Interoperability research addressing data integration challenges experienced by institutions in an international setting also remains sparse. Groundwater is an example of an environmental phenomenon which does not respect political borders, and its management requires data from multiple jurisdictions. The Abbotsford-Sumas aquifer, straddling the Canada US border, is used as a case study to explore integration challenges in an international setting. Development of groundwater management practices to ensure a sustained source of good quality groundwater is dependent, on an understanding of the conceptual model of the aquifer. Due to a lack of geophysical studies, geological information contained in the water well reports, is the chief source of depth-specific lithological information. The use of this information in constructing the conceptual model is constrained by poor data quality and a lack of an integrated and standardized lithological database. To achieve the research goals of exploring integration challenges in an international setting, lithological datasets from BC and Washington State are integrated. The resultant lithological database is used to test the usability of water well reports for constructing the conceptual model. Numerous interoperability challenges such as data availability, lack of metadata, data quality and formats, database structure, semantics, policies and cooperation are identified as inhibitors of data integration. Despite the numerous challenges the lithological database is useful in constructing a generalized conceptual model. This research is important as it presents challenges to data integration that should be considered as a starting point for environmental management projects.

DEDICATION

This thesis is dedicated to the two individuals that influenced my life the most: Mom and Dad.

ACKNOWLEDGEMENTS

I take this opportunity to thank all the individuals that made this Master's program a success. First of all I would like to thank my supervisor Dr. Nadine Schuurman for her support and guidance during this thesis. Thank you for having confidence in me and always encouraging me to do my best. I would like to thank Dr. Diana Allen for her guidance on the geological aspect of this research and also for reviewing my thesis. I would also like to thank the external examiner, Dr. K.R. Rothley, for her excellent suggestions that helped make this a better thesis.

Thanks to all the organizations in Canada and United States of America for providing digital datasets. These organizations, in Canada, include Ministry of Sustainable Resource Management for TRIM 1 and cadastral datasets; Ministry of Water, Land and Air Protection for groundwater datasets; Ministry of Transportation (Canada) for bridge construction reports; City of Abbotsford and Township of Langley for the city of Abbotsford and Langley datasets; Valerie Cameroon for drill core records and Safe software Inc. for FME software. Thanks to Lynn Doremus from Nooksack Tribe Natural Resources and Peter Gill from Whatcom County for the groundwater databases. Thanks to Basil Hii of Environment Canada for funding me during the summer of 2002. I would also like to thank Denny and Cynthia for data entry.

Thanks to the Department of Geography and all the students who helped make this program a wonderful experience. Thanks to Marcia and Jasper for their unconditional help during this research. A big thanks to Kathy for editing my draft and for the moral support during the Master's program, thanks for everything and I appreciate all your help and support.

Finally, I would like to thank my husband, Sandeep for his technical guidance and for gracefully handling my stress. This thesis would not have been completed without your love and encouragement.

TABLE OF CONTENTS

APPROVAL ii
ABSTRACT iii
DEDICATION iv
ACKNOWLEDGEMENTS v
TABLE OF CONTENTS vi
LIST OF FIGURES viii
LIST OF TABLES ix
LIST OF ABBREVIATIONS AND ACRONYMS x

CHAPTER 1 INTRODUCTION 1
1.1 Data Integration 1
1.2 What is Interoperability? 1
1.3 National Groundwater Initiatives 4
1.4 Abbotsford–Sumas Aquifer 7
1.5 Purpose 8
1.6 Objectives 9
1.7 Scope of Work 10
1.8 Thesis Outline 10

CHAPTER 2 DATA INTEGRATION 12
2.1 Background 12
2.2 Interoperability Issues 12
2.3 Spatial Data Standards 15
2.3.1 National Standardization Efforts 16
2.3.2 International Efforts 19
2.4 Spatial Data Infrastructures 25
2.5 Database Integration 28
2.5.1 Data Heterogeneities 28
2.5.2 Semantic Interoperability 30
2.6 Organizational Interoperability 35
2.7 Summary 36

CHAPTER 3 DATA INTEROPERABILITY ACROSS BORDERS 40
3.1 Introduction 40
3.2 Abbotsford-Sumas Aquifer 42
3.3 Pre Processing 45
3.3.1 Data Acquisition 45
3.3.2 Metadata 50
3.3.3 Data Dissemination / Access 54
3.3.4 Data Quality 55

3.4 Data Conversion 64
3.5 Data Integration 66
3.5.1 Schematic Heterogeneities 67
3.5.2 Semantic Heterogeneities 72
3.6 Summary 77

CHAPTER 4 **THE HYDROSTRATIGRAPHIC MODEL 83**
4.1 Introduction 83
4.2 Past Research 83
4.3 Geological Setting 85
4.4 Hydrostratigraphic Model 88
4.5 GMS Output 93
4.6 Summary 101

CHPATER 5 **CONCLUSION AND RECOMMENDATIONS 104**
5.1 Data Constraints 107
5.2 Institutional Factors 109
5.3 Hydrostratigraphic Model and Software Issues 111
5.4 Future Research and Recommendations 113

APPENDIX A: DATASETS, INTEROPERABILITY AND STANDARDIZATION 116

APPENDIX B: VISUAL BASIC SCRIPTS 125

APPENDIX C: SEDIMENT CLASSIFICATION SYSTEMS AND HYDRAULIC PROPERTIES 130

Appendix D: CROSS-SECTIONS 134

REFERENCE LIST 137

LIST OF FIGURES

Figure 3-1	Abbotsford-Sumas aquifer	44
Figure 3-2	CGDI design	57
Figure 3-3	Missing elevations (wells composed of single record)	59
Figure 3-4	Missing elevation (wells composed of two layers)	60
Figure 3-5	Original Washington State lithology table lacking sequence number/ LayerOrderNum and layer depths	60
Figure 3-6	Washington State lithology table showing the sequence / layerOrderNum and layer depths	60
Figure 3-7	Surface elevation represented as a null value	61
Figure 3-8	Surface elevation represented by non-zero value	62
Figure 3-9	Introduction of a layer of unknown geology	62
Figure 3-10	Inconsistent records	63
Figure 3-11	Table structure conflicts	69
Figure 3-12	Attribute conflicts	71
Figure 3-13	Semantic heterogeneities in the hydrographic datasets	73
Figure 3-14	BC transportation dataset showing Highway No.1	74
Figure 4-1	Fraser-Whatcom Basin stratigraphic units and corresponding glaciation stages *(adapted with permission from Jones, 1999)*	86
Figure 4-2	3-D model of the study area (side view along a single cross-section)	91
Figure 4-3	3-D model of the study area (front view)	91
Figure 4-4	Cross-sectional plan	95
Figure 4-5	Cross-section A-A'(generalized hydrostratigraphic model)	96
Figure 4-6	Cross-section D-D' (generalized hydrostratigraphic model)	97
Figure 4-7	Cross-section M-M (generalized hydrostratigraphic model)	98
Figure 4-8	Cross-section O-O' (generalized hydrostratigraphic model)	99
Figure 4-9	Cross-section O-O' (lithologic units)	100

LIST OF TABLES

Table 3-1 Data acquisition challenges 50

Table 3-2 Inconsistent well records in the BC and Washington State databases 58

Table 3-3 Total reduction of well records 64

Table 3-4 Details of datasets consulted 65

Table 3-5 Table name conflicts 67

Table 3-6 Geological descriptions recorded in databases 75

Table 4-1 Generalized Hydrostratigraphic Model 93

LIST OF ABBREVIATIONS AND ACRONYMS

ATKIS	Authoritative Topographic-Cartographic Information System
BC	British Columbia
BMGS	Base Mapping and Geomatic Services
C#	C sharp
CCOG	Canadian Council on Geomatics
CCOGIF	Canadian Council on Geomatics Interchange Format
CCSM	Canadian Council on Surveying and Mapping
CEN	Comité European de Normalization
CGDI	Canadian Geospatial Data Infrastructure
CGDS	Computerized Groundwater Data System
DBMS	Database Management System
DEM	Digital Elevation Model
FGDC	Federal geographic Data Committee
6666yhFSSD	Flexible Standardization Software
GDF	Geographic data File
GML	Geography Markup Language
GMS	Groundwater Modelling Software
GSC	Geological Survey of Canada
ISO	International Standardization Organization
LDBuilder	Lithological Database Builder
MSRM	Ministry of Sustainable resource Management
MWLAP	Ministry of Water, Land and Air Protection
NAD	North American Datum
NCGIA	National Centre for Geographic information and Analysis
NRCan	Natural Resources Canada
NSDI	National Spatial Data Infrastructure
NTS	National Transfer Standard
OGC	OpenGIS Consortium
OGIS	OpenGIS Interoperability Specifications
OS	Operating System
RASA	Regional Aquifer System Analysis
SAIF	Spatial Archive Interchange Format
SDI	Spatial Data Infrastructure
SDTS	Spatial Database Transfer Standard
TC	Technical Committee
US	United States of America
USGS	United States Geological Survey
UTM	Universal Transverse Mercator
XML	Extensible Markup Language

1 INTRODUCTION

1.1 Data Integration

Data integration is at the core of Geographical Information Systems (GIS) and has been used as one of its defining properties (Vckovski, 1998). Due to its integration ability, GIS technology is becoming a key tool in resolving many socio-economic and environmental issues. Although data fuels the GIS industry, it is also the source of most data integration challenges. Resolving these issues requires an understanding of data formats, data models, data standards, data quality and semantics. Thus, data integration, the focus of this research, is not an easy task, and Laurini (1998) has rightly expressed that "Interoperability is a users' dream but a programmer's nightmare."

1.2 What is Interoperability?

Interoperability, which is the integration of datasets, is an active field of research in the GIS community. The term interoperability has a range of meanings (Sondhiem et al., 1999; Goodchild et al., 1997) and has been defined differently by various authors. For example, Bishr (1998) has defined interoperability at six levels: Network Protocols, Hardware and Operating System, Spatial Data Files, Data Model and Application Semantics, whereas Goodchild et al. (1997) have identified interoperability at three levels: technical, semantic and institutional level. Moreover, terms like data sharing, interoperability and integration have also been used synonymously. Despite these differing terminologies, interoperability research focuses on information exchange between systems in heterogeneous environments without loss of information.

Goodchild et al. (1997) argued that the problem of interoperability stems from the lack of an overarching theory or the lack of common terminology used during the early proliferation of GIS technology into diverse fields like forestry, planning and agriculture. Although considerable research has been conducted since 1970s in this regard, interoperability is far from being a reality (Vckovski, 1998). The many advantages of data integration, such as reduction in duplicative efforts, economic savings and informed decision making, however, provide impetus for interoperability research.

During the 1960s and 70s data were maintained by government organizations (Rhind, 1999) in proprietary formats, which made information exchange a difficult task. In response to these problems, data transfer formats were developed by various government organizations around the world (Sondheim et al., 1999; Salgé, 1999; Moellering, 1991). During the 1980s, government organizations had also recognized the importance of geospatial datasets as national assets (Barr and Masser, 1997). Advances in computer technology at that time also resulted in a proliferation of geospatial datasets into diverse fields. Government organization's recognition of geospatial datasets as a national asset and the need to bring coherence to the widespread proliferation of datasets resulted in the development of spatial data infrastructures (SDI's) (Rhind, 1999; Masser, 1998; Tosta, 1997).

Although standards and SDI's provided a foundation for exchanging information data could not be shared due to semantic issues (Bishr, 1998). As a result, data were shared but not the meaning (Kuhn, 1994). Semantics, which results from the different world visions of information communities, was identified as the source of most interoperability problems (Bishr et al., 1999).

Various methodologies have been proposed by researchers to resolve semantic interoperability issues (Sheth, 1999; Bishr, 1998; Visser et al., 2002b). A commonality among these various approaches is the use of data modelled on the object oriented paradigm. In reality, however, bulk data are modelled in a relational format and these methodologies do not resolve the immediate needs of the organization seeking interoperability, where data are maintained in a relational format (Schuurman, 2002).

Further advances in information technology, the advent of the Internet, and distributed computing architecture brought new ways of data exchange and dissemination. The OpenGIS Consortium (OGC) was born out of these new developments in Information Technology. OGC is a non profit organization comprising the academia, public and private sector seeking to achieve interoperability through the standardization of interfaces[1] (Buehler and McKee, 1998).

Although sharing of information between organizations was facilitated by technical advancements, organizational issues resulted in stalled collaborative initiatives (Onsrud and Rushton, 1995; Nedovic-Budic and Pinto, 1999; 2000; 2001). As a result, research has been conducted by various authors from an organizational perspective, drawing solutions from organizational theory, inter-group dynamic and exchange theory to decipher the complex inter-organizational relationships (Onsrud and Rushton, 1995; Nedovic-Budic and Pinto, 1999; Azad and Wiggins, 1995), which may hinder the interoperability process.

---

[1] Interfaces are defined as a boundary between two independent systems where the systems meet and communicate with each other. There are three types of interfaces:

1. User Interface: the mouse, keyboard and menus are example of user interfaces. User interfaces allow the user to communicate with the Operating System
2. Software Interfaces: software languages and code that applications use to communicate with each other and hardware
3. Hardware interfaces: wires and plugs that hardware devices use to communicate with each other (Webopedia, 2004)

Despite extensive research to unravel the technological and organizational intricacies of interoperability, initiatives are still in the prototyping stage and have not been incorporated into commercial software. Furthermore, there appears to be institutional reluctance among organizations to achieve interoperability. Schuurman, (2002) has cited a study in British Columbia (BC) where the Ministry of Water, Land and Air Protection (MWLAP) is reluctant to achieve interoperability unless incorporated into commercial software.

This lack of motivation for standardization is compounded by limited resources and lack of a definite interoperability solution roadmap. As a result, integration continues to be time consuming and involves large amount of resources prior to analysis and decision making. Complexities associated with data integration can be a daunting task and can lead to abandoned initiatives and recreation of datasets to suit particular needs, thus duplicating the effort. Moreover, interoperability research from a technical and organizational perspective has focused primarily on achieving interoperability within single jurisdictions, and sparse literature exists for cross-border interoperability issues.

1.3 National Groundwater Initiatives

Increasing urbanization, industrial growth and agricultural practices have put tremendous pressure on the quality, quantity and usage of natural resources. Groundwater is a natural renewable resource and a potential source of drinking water. Exploitation of groundwater can have adverse repercussions; ranging from direct effects, such as water table lowering to indirect effects, such as subsidence[2], saltwater intrusion

---

[2] Subsidence: is the lowering of the ground elevation due to compaction of sediments caused by excessive withdrawal of water, oil or gas from the subsurface. Excessive withdrawal of groundwater aquifers is one of the causes of subsidence. Subsidence is a permanent process and cannot be reversed. Damage from subsidence can cause sever damage to urban structures (Leake, 2004).

and contamination. Aquifers are porous media that store and transmit groundwater. They also do not follow political boundaries. Effects of uneven exploitation and contamination can lead to political and legal issues. As a result, groundwater management calls for a collaborative approach requiring data from the affected nations.

In Canada, the availability and quality of groundwater is under stress from increasing demand, contamination, and possibly, climate change. About 25% of Canadians depend on groundwater. As Canada has been blessed with an abundance of water resources the less visible portion which is groundwater, has received little public attention (Environment Canada, 2003). In Canada, up until 1993, knowledge of groundwater resources at a regional scale were limited (Sharpe, n.d.). In contrast, the groundwater programme RASA (Regional Aquifer System Analysis), in the United States of America (US), was initiated in 1978 with the aim of defining the regional hydrogeology and establishing a framework of background information on geology, hydrology, and geochemistry of the important aquifer systems (Sharpe, n.d.). The study was initiated in response to federal and state requirement for information to improve the management of groundwater resources. This research has developed an advanced understanding of the groundwater systems at a regional scale in the US.

In Canada, one of the first national syntheses of groundwater resources by the Geological Survey of Canada was published in 1967 (Sharpe, n.d.). A study in 1993 by the Canadian Geoscience Council identified that Canadian efforts in groundwater inventory, protection and research were fragmented and lacked a knowledge base at a national scale (Sharpe, n.d.). These recommendations were addressed by the Geological Survey of Canada (GSC), who in collaboration with provincial and municipal partners initiated a regional hydrogeology program. The aim of the program was to delineate and characterize aquifers in rapidly growing centres, in terms of hydrogeology, geophysics and chemistry (Ricketts, 1997). Two pilot projects were launched: "The Oak Ridges Moraine Project" in Ontario and "The Vancouver-Fraser Lowland Project" in British Columbia.

The principal objective of both projects was to develop a hydrogeological database and characterize the hydrogeological architecture of the aquifer(s) for use at all levels of management and research. In both projects, water well log information contained within the provincial governments databases and geophysical surveys were used to delineate the aquifers. The Oak Ridges Moraine project resulted in a standardized hydrogeological database and the development of a 3-dimensional regional hydrogeological model for the area. The Vancouver-Fraser Lowland project, however, resulted in a database for the Fraser lowland and a 3-dimensional hydrogeological model only for the Brookswood aquifer, located near Langley, BC. Development of detailed hydrogeological models was not extended to other aquifers in the Fraser Lowland area. In both studies geophysical surveys played a significant role in defining the hydrogeological architecture. Geophysical studies, an expensive method for determining subsurface geology, however, can be an economic constraint on under-funded government organizations that must contribute to groundwater management activities.

In Canada, water well drillers are either encouraged or required (by legislation) to submit well log information to the provincial governments. The task is mandatory in Ontario, Alberta and Newfoundland and Labrador. New amendments to the Water Act in BC require submission of well logs (Allen, 2004). These reports have developed groundwater data banks in various provinces and are an important source of depth-specific geological information. In the absence of geophysical field surveys, the water well lithology log information is invaluable, and may perhaps be the chief source of geological information. The use of the water well data, however, has limitations due to unreliable geological descriptions (Russell et al., 1998).

A more recent initiative for groundwater management in Canada is the 'Canadian Framework for Collaboration on Groundwater' initiated as part of the Clean Environment program of Natural Resources Canada. This initiative aims to develop a high standard groundwater knowledge database, promote inter-organizational collaboration, establish linkages of groundwater information systems, and provide a resource base for all levels of government for policy making (Rivera et al., 2003). Although this initiative has developed a framework for initiating collaborative efforts between government organizations and stakeholders in Canada, it does not call for collaborative initiatives at an international level.

1.4 Abbotsford–Sumas Aquifer

The Abbotsford-Sumas aquifer, which straddles the Canada-US Border (BC / Washington State), is the largest aquifer in the Fraser Lowland. It is located in the Fraser Valley and extends from the city of Abbotsford, BC (Canada) to Lynden, Washington (US) in northern Puget Willamette Lowland. Both Canada and the US share an equal proportion of the aquifer. This aquifer supports the activities of approximately 200,000 people who live in this area. Groundwater is used not only for drinking purposes, but also supports industrial, farming and agricultural activities (Cox and Kahle, 1999; Kohut, 1987). These activities have threatened the integrity of this aquifer (Ricketts, 1997). Agricultural practices and the poultry industry have lead to nitrate contamination of the aquifer (Cox and Kahle, 1999). The aquifer has also been exploited extensively on both sides of the border (Cox and Kahle, 1999). Canada is concerned with the excessive groundwater withdrawal south of the border (Kohut, 1987) and the US is concerned with groundwater contamination that may originate north of the border (Cox and Kahle, 1999). As groundwater is the primary source of water for many inhabitants of the study area there is a pressing need to develop groundwater management strategies for the area. Groundwater management strategies, however, call for a thorough understanding of the hydrogeological architecture, which is lacking for this area.

1.5 Purpose

The ability to integrate data is central to geographic information science. Data integration, an issue in the GIS community since the 1970s, is not an easy task to resolve. Data integration requires an understanding of data quality, data models, data structure, data formats, and data semantics. Despite extensive research, data interoperability is far from being a reality (Vckovski, 1998).

To resolve data integration issues, various authors have proposed methodologies based on the object oriented paradigm (Bishr, 1998; Sheth, 1999; Visser et al, 2002b). This is in contrast to the relational format, a format commonly used by most organizations (Schuurman, 2002). Moreover, as interoperability research is still in the prototyping stage, these methodologies do not resolve the immediate need of organizations seeking interoperability.

Interoperability research from technical and organization perspectives have concentrated on issues within a single jurisdiction. Environmental phenomena, however, do not respect political boundaries and their management often requires data from multiple jurisdictions. With the exception of OGC activities, sparse literature exists on interoperability research in an international setting. Also, little is known about the data integration challenges faced by organizations in an international setting.

The Abbotsford–Sumas aquifer provides a unique case study as interoperability research for a cross-border aquifer within a Canadian-US context has not been investigated. It is also an aquifer that is highly used and highly vulnerable to contamination (Berardinucci and Ronneseth, 2002). Development of groundwater management activities to ensure a sustainable source of good quality water is dependent on an understanding of the conceptual model of the aquifer. Construction of the conceptual model for this is further dependent on the depth-specific lithological information provided in the water well database. The use of the lithological information is, however, constrained due to the inconsistent geological descriptions (Russell et al., 1998), extremely poor data quality, and the lack of a standardized integrated database.

Thus, the purpose of this research is to explore data integration challenges in an international setting and to test the usability of the water well database for the construction of the hydrostratigraphic model for the aquifer.

1.6 Objectives

This research explores the multifaceted challenges encountered when integrating datasets needed for developing a hydrostratigraphic model of the Abbotsford-Sumas aquifer. This research is timely given that interoperability research is in the prototyping stage and has focused primarily on interoperability issues within a single jurisdiction.

The objectives of the research are:

1. To study the challenges encountered when integrating datasets for cross-border projects;
2. To create an integrated hydrogeological dataset for Abbotsford-Sumas aquifer; and
3. To develop a hydrostratigraphic model to evaluate the usability of the water well lithology database.

1.7 Scope of work

To achieve the above mentioned objectives of investigating data integration challenges and construction of the hydrostratigraphic model, the following tasks were undertaken:

- Acquiring spatial and attribute (lithological) information for the Abbotsford-Sumas aquifer from organizations in BC and Washington State.
- Acquiring lithological information from various geological reports, drill core reports and bridge construction reports.
- Converting data in paper format to digital formal. Conversion includes manual encoding or digitizing.
- Converting data from BC and Washington State to a common framework (Projection: Universal Transverse Mercator (UTM); Datum: North American Datum (NAD 83); Zone 10).
- Merging spatial data using ArcGIS® v 8.3.
- Integrating lithological datasets from BC and Washington State using MS Access.
- Documenting data integration challenges for cross-border projects.
- Standardizing lithological terms using Flexible Standardization Software (FSSD).
- Constructing the hydrostratigraphic model using Groundwater Modelling Software (GMS) v 4.0.
- Documenting the usability of the lithological database in constructing the hydrostratigraphic model.

1.8 Thesis Outline

This thesis is organized into five chapters. Chapter 1 provided an introduction to the subject of interoperability, described the motivation for the research and presented the objectives of the research. Chapter 2 provides a summary of the interoperability problem and the current status of interoperability research. It Chapter 3 describes the interoperability challenges of integrating dataset for developing the hydrostratigraphic model of the Abbotsford-Sumas aquifer. Chapter 4 analyzes the usability of the water well database in the development of the hydrostratigraphic model, and Chapter 5 provides a conclusion and recommendations for future research.

2 DATA INTEGRATION

2.1 Background

Data standardization and integration is an active field of research in the GIS community. Despite ongoing research, overarching solutions that can be incorporated into commercial software packages do not exist. Although most interoperability solutions require the use of object oriented paradigm, bulk data are maintained in a relational format (Schuurman, 2002). Economically constrained organizations face innumerable challenges when integrating datasets for decision-making, and are unlikely to make a transition from relational to object oriented data formats. Interoperability research from a technical and institutional perspective that addresses integration challenges faced by organizations in an international setting remains sparse with the exception of Open GIS Consortium (OGC) activities.

The primary objectives of this thesis are to investigate challenges encountered while integrating datasets for groundwater management activities of the Abbotsford-Sumas aquifer and to test the usability of groundwater datasets for constructing a hydrostratigraphic model of the aquifer. For this purpose literature will be drawn from interoperability studies.

2.2 Interoperability Issues

Interoperability issues in the GIS domain emerged in the late 1970s (Sondheim et al., 1999) when data were primarily maintained in proprietary formats that resulted in insular data (Harvey et al., 1999). What was initially termed data integration research (Shepherd, 1991; Flowerdew, 1991) is now being referred to as interoperability research.

Interoperability has a range of meanings (Goodchild et al., 1997; Sondhiem et al. 1999) and, as such, has been construed differently by various authors. Moreover, the tendency for terms like data sharing, integration, standardization and interoperability to be used interchangeably by the research community has compounded this confusion.

Bishr (1998) defined interoperability as "the ability of a system or components of a system to provide information portability and inter-application cooperative process control." He further segregated interoperability into six levels: network, hardware and operating systems (OS), spatial data files, database management systems (DBMS), data models and application semantics. Vckovski (1999) provides a simpler definition: interoperability is "the ability to exchange and integrate information". Unlike Albrecht (1999), who conceptualized interoperability as standardization protocols, Sheth (1999) examined interoperability in the context of syntactic, schematic, structural and semantic heterogeneities. Irrespective of these varied definitions and interpretations, the fundamental goal of interoperability is to facilitate seamless integration of information without *a priori* knowledge of data formats, data models and semantics in heterogeneous environments.

Goodchild et al. (1997) argued that the problem of interoperability stems from the lack of an overarching theory or common terminology used during the early proliferation of GIS into such diverse fields as forestry, agriculture, and planning. Vckovski (1999), called for a spatial information theory as a foundation for interoperability research, which would address the theoretical vacuum. A lack of consensus on underlying theoretical frameworks (Goodchild et al., 1997), compounded by complexity of spatial information, has increased the challenges associated with developing and implementing data interoperability. Today, interoperability solutions are in the prototyping stage and are far from being fully operational (Vckovski, 1998).

Prior to the mid-1990s, interoperability research in the GIS domain concentrated on static methods of integration, such as the development of ad hoc data translators (Sondheim et al., 1999; Guptill, 1991) and standardization of data formats (Moellering, 1991) for information exchange. Advances in technology during the 1990s caused an increase in the number of available spatial datasets and opened new avenues for achieving interoperability. The focus has changed to achieving interoperability in dynamic (run time access to datasets) environments, which has produced its own new set of interoperability problems.

This interoperability research achieved a level of coherence through the Varenius Project's research initiative I-20 'Interoperating Geographic Information Systems,' which sought to provide a framework for interoperability research. Three key areas of research were identified: technical, semantic and institutional (Goodchild et al., 1997). Since then, considerable research has been conducted on these various fronts. While this research has primarily attempted to resolve interoperability issues within a single jurisdiction, little has been documented about the challenges of working in an international setting where complex dynamics exist[3].

Interoperability research has progressed from static to dynamic approaches in which the focus has changed from developing transfer formats to building infrastructures, understanding organizational dynamics and achieving dynamic exchange of information and software services. Based on this changing focus, this chapter is further divided into four sections. Section 2.3 investigates data standardization activities as a method of facilitating data integration. This is followed by a summary of spatial data infrastructures in Section 2.4. Though standardization had facilitated data

---

[3] Complex dynamics include political culture and policies prevalent within the jurisdictions, or the relationship between the concerned jurisdictions.

exchange, semantic issues continue to inhibit use of spatial datasets (Sheth, 1999; Bishr, 1998); this is the focus of Section 2.5. In spite of advances on the technical front to facilitate exchange of information, organizational dynamics have resulted in stalled data sharing initiatives (Onsrud and Rushton, 1995). This organizational component of interoperability is discussed in Section 2.6. The chapter concludes with a summary in Section 2.7.

2.3 Spatial Data Standards

This section provides a summary of data standardization activities in the GIS community to facilitate data exchange. Studies on data standardization began about 25 years ago (Salgé, 1999) and was recognized as a key element of the data integration process (Guptill, 1991). The impetus for developing transfer standards emerged within the mapping organizations in the early 1980s as a means to distribute data between government organizations and spatial data users (Taylor, 1996; Salgé, 1997). Since then, many countries and international organizations have developed spatial data transfer standards (Moellering, 1996; 1991).

Standardization has inherited concepts and model procedures from Information Communication Technology (Salgé, 1997) and has advanced from a data centric approach (direct translator, common exchange formats) to a process centric approach used by Open GIS Consortium (OGC) (Strand et al., 1994). Most spatial database transfer standards have employed the data centric approach, which concentrates on achieving data portability and standardization of conceptual data models (Salgé, 1999). Despite the current standardization trend toward a process centric approach, as seen in ISO TC211 and OGC specifications (based on open operability standards that support distributed data management), efforts are still concentrated on data properties (Salgé, 1999).

The first approach to data standardization can be traced back to the 1970s when data were maintained in proprietary formats and exchange of data was a complicated task. Data exchange was achieved using direct translators, which provided an inefficient means of data exchange (Sondheim et al., 1999; Guptill, 1991). Such translations not only resulted in a loss of information due to dissimilar input and output data models, but also new translators had to be developed for new information sources which proved to be extremely expensive (Sondheim et al., 1999, Guptill, 1991). New methods, based on a vendor neutral common exchange file structure, were developed to overcome the inefficient means of data exchange afforded by direct translators, (Guptill, 1991). Such vendor neutral exchange formats are the basis for most national and international standardization efforts.

2.3.1 National Standardization Efforts

During 1970s and 1980s spatial datasets were the purview of government organizations. As increasing number of government organizations embraced GIS technology; however, redundant datasets were generated as each agency sought to fulfil its own data needs. Recognizing inefficiencies arising from the maintenance of redundant datasets and the prevailing inefficient methodologies of translating information, government organizations around the world initiated standardization activities intended to develop exchange standards based on non-proprietary formats having sophisticated data models, data catalogues and data encoding information (Sondheim et al., 1999; Moellering, 1991). This resulted in national data transfer standards around the world. The Spatial Data Transfer Standard (SDTS), for example, was developed by the US (SDTS, 2002), and later adopted as a standard in Australia (Clarke, 1991), while the Spatial Archive Interchange Format (SAIF) was developed by

Canada (Albrecht, 1999), and the National Transfer Format (NTS) was developed by the UK (Sowton, 1991).

The US was the first nation to develop the concept of a spatial data transfer standard (Salgé, 1999). In 1980, through a memorandum of understanding between the USGS and US National Bureau of Standards, now called the National Institute of Standards (NIST), USGS was assigned leadership in developing, defining and maintaining earth science data for use by US government agencies (Rossmeissl and Rugg, 1991). In 1983, the office of Management and Bureaus directed the USGS to eliminate duplication and waste in the development of digital cartographic databases and to coordinate digital cartographic activities (Rossmeissl and Rugg, 1991). The USGS established the Federal Interagency Coordinating Committee on Digital Cartographic Data (FICCDC), which undertook the task of developing a data exchange format called the "Federal Geographic Exchange Format" (Rossmeissl and Rugg, 1991), now called the Spatial Data Transfer Standard (SDTS). This standard achieved official status in 1994 after a rigorous development and testing phase of ten years, as a Federal Information Processing Standards (FIPS: 173-1) (Salgé, 1999). This standard has a rich conceptual model, is capable of representing any form of spatial data (Arctur, 1998), and is implemented through the use of profiles (SDTS, 2003). Compared to other national standards, this standard is unique in that its use is mandated by all federal organizations (Salgé, 1999; Wortman, 1992). Such mandatory distribution of datasets using the national standard has not been legislated in other nations.

In Canada five related standardization activities were in progress to develop standards for geographic data (Evangelatos and Allam, 1991). Of importance is the development of Digital Topographic Data Standards, which began in the late 1970s under the aegis of Canadian Council on Surveying and Mapping (CCSM), now, known as the Canadian Council on Geomatics (CCOG). These efforts resulted in development of a Digital Topographic Information Model (DTIM), which provides detailed classification and coding rules for structuring digital topographic features and a CCOG Interchange Format (CCOGIF) for the transfer of digital topographic data (Evangelatos and Allam, 1991). The CCOG standards, which serve as a basis for exchanging digital topographic data between federal, provincial and private surveying companies has been implemented successfully in Canada (Evangelatos and Allam, 1991). Though the CCOG classification provides standardization in terms of definition and encoding of topographic features and facilitates exchange of topographic information, it is truly not a spatial data transfer standard as it lacks a rich conceptual model that can go beyond modelling topographic data.

A more sophisticated standard, which formed the basis for other standards like OGC specifications and SQL3MM, was developed by Henry Kucera and Mark Sondhiem working at the Ministry of Environment, Land and Parks (MELP), now called Ministry of Water, Land and Air Protection (MWLAP) (Albrecht, 1999). Their efforts resulted in the Spatial Archive Interchange Format (SAIF) (Albrecht, 1999). This exchange format, which was developed initially for the Ministry of Forests (BC), became a Canadian National Standard in 1993 (Albrecht, 1999). Unlike the CCOG and other national standards, however, the SAIF standard is developed using the object oriented paradigm and supports data exchange (Albrecht, 1999). It also provides a solution for modelling spatio-temporal features that were not accounted for in previous data transfer standards (Albrecht, 1999). Although similar standardization activities may be observed in various countries around the world, and today more than twenty such standards exist (Moellering, 1996), these standards tackle data integration problems in a national context and do not provide a solution for transnational issues.

2.3.2 International Efforts

In response to data integration problems at transnational level, international communities developed application specific de-facto standards in the fields of hydrology, Defence and automobile industry to suit their specific needs (Salgé, 1999; Albrecht, 1999). During the 1990s, advances in information technology, such as the Internet and distributed computing architecture, introduced new paradigms for achieving interoperability. Distributed networking concepts were adopted by Open GIS Consortium (Buehler and McKee, 1998) to develop sophisticated solutions for exchange of information and geoprocessing software. Widespread use of geographic information prompted international standardization bodies such as the International Standards Organization (ISO) and the Comité European de Normalization (CEN) to develop templates for other functional standards (Albrecht, 1999). Following is a brief summary of the various international standardization activities.

- **Digital Geographic Information Exchange Standard (DIGEST)**: In 1983, the Digital Geographic Information Working Group (DGIWG) was formed in response to the growing demands for geographical defense datasets (Smith, 1991). The main aim of DGWIG was to ensure that different national geographic information systems could exchange data effectively and efficiently without loss of information (Cassettari, 1993). In 1991, DGIWG developed DIGEST, a family of standards that facilitates the exchange of defense data between NATO Organizations and makes use of existing ISO standards (Smith, 1991).
- **Geographic Data File (GDF):** Work on this standard was initiated in 1980s by the members of the European Automobile Industry to develop a European Digital Road Map (ERDM) for car navigation systems (Albrecht, 1999). Though originally a de-facto standard, it is now endorsed by Comité European de Normalization (CEN) and International Standards Organization (ISO) (Albrecht, 1999).

- **S-57:** The International Hydrographic Organization (IHO) developed the S-57 transfer standard for exchange of digital hydrographic data between national hydrographic offices and for its distribution to manufacturers, mariners and other data users (Albrecht, 1999). Currently, efforts are underway to harmonize S-57 and DIGEST (Salgé, 1999).

- **Open GIS Consortium (OGC):** Many national and international standards were developed at a time when computer technology was limited in scope. As information technology was evolving it introduced new concepts for GIS technology (Batty, 1999). Object-Oriented technology brought new ways of modeling spatial data; the Internet increased interconnectivity (Sheth, 1999); networked and distributed computing architecture facilitated dynamic access to various data repositories and the development of platform independent languages like JAVA, C# (Tsou and Buttenfield, 2002). Such advances introduced new paradigms for data exchange and interoperability and provided a foundation for organizations like OGC and ISO/TC 211.

  OCG is a non-profit organization composed of private, public sector and the academia (or the academy) that was founded in 1994 in response to the widespread interoperability problems faced by the industry (Buehler and McKee, 1998). Its vision is "the full integration of geospatial data and geoprocessing into mainstream computing, as well as the widespread use of interoperable geoprocessing software and geospatial data products throughout the information infrastructure" (Buehler and McKee, 1998). Its primary objective is to resolve integration problems for on-line GIS applications (Tsou and Buttenfield, 2002) through the standardization of interfaces based on object oriented technology. By standardizing the representation of spatial data at the interface level, its goal is to evade from the need to define a universal data model at the database level (Reed, 2003).

  To realize its vision of seamless integration of data and geoprocessing, OGC is active in developing the OpenGIS Interoperability Specifications (OGIS), which is a comprehensive set of specifications that address various aspects of interoperability. There are two types of specifications: the Abstract specifications[4] and the Implementation specifications[5]. These specifications provide software developers a template for writing interfaces that interoperate with OGIS compliant software developed by other vendors (Buehler and McKee, 1998). Interoperability is thus achieved at the interface level by writing Application Programming Interfaces (APIs) that plug and play with other applications adhering to these specifications, or at the encoding level by using Geography Markup Language (GML) (Reed, 2003) which is an extension of Extensible Markup Language (XML) technology. Although interoperability is achieved at the interface level it does not account for interoperability between information communities. Therefore, to ensure data interoperability, data need to comply with the abstract specification, which is the true obstacle at this stage.

XML technology was developed by World Wide Web Consortium (W3C) for the exchange of a wide variety of data on the web (W3C, 2003). The popularity of XML for exchange of data in various domains lies in its ability to separate information from presentation. This provides application developers the flexibility and a set of rules to build custom data structures for industry specific needs without being concerned about presentation of information (Pitts, 2001; Lake, 2001). GML is an XML encoding that has been developed specifically for the geospatial domain (Lake, 2001; Cox et al., 2003). This specification is based on the OGC feature specification and is used for modeling, storage, access and transfer of spatial and non spatial information on the web (Cox et al., 2003). It is organized into schemas that provide a set of standards and a restricted methodology for describing features and their geometries (Lake, 2001). Although GML provides a method of sharing data, interoperability is achieved by comparing the geometry of spatial objects (Lake, 2001). Methods to achieve semantic interoperability based on GML, at

---

[4] The Abstract Specification is a formal language expression of how real world features ought to be expressed. They are abstract, implementation independent and are modelled using modelling languages like Object Model Template (OMT) or Unified Modelling Language (UML) (Kottam, 1999). These specifications are identical to the ISO 211 standard (Reed, 2003).

[5] Implementation Specifications are software specifications for various distributed computing platforms.

present, are in the developmental phase (Lake, 2001) and are domain specific. Geospatial One Stop Transportation Pilot Initiative (GOS TP) of OGC attempts to achieve semantic interoperability within the transportation community (Reed, 2003) in US by mapping schemas of local transportation models to the national transportation model using the Web Feature Services[6] (WFS) and GML for transferring data. A client that implements a WFS compliant interface can access two or more repositories that have different conceptual models. An advantage of using GML is that it can avoid information loss by providing details of features that do not translate. Although this initiative provides interoperability for transportation industry within the US, it does not account for transportation models for transnational projects. Despite OGC's lack of support for semantic interoperability (at this time) OGC specifications are gaining quick recognition as compared to other standards (Albrecht, 1999).

OGC has contributed to achieving interoperability and geo-enabling the web through its various programs. OGC holds a unique position in the standardization arena as it is user centered; users and vendors work together to develop new standards (Kottam, 1999). In addition to its work with developing specifications, OCG offers data providers and organizations test beds, technical assistance and feasibility studies through its Interoperability Programs. It is also through these initiatives that new specifications are usually conceived and developed. For example, the first Interoperability Program Initiative 'Web Mapping Testbed' (WMT1), which allowed users to overlay maps from different sources without being concerned about the proprietary software managing these data sources, resulted in the Web Map Server and GML specifications (McKee, 2002) . Prototypes built through these initiatives also provide vendors an opportunity to test specifications in a real world setting (OGC, 2003b). Products built using OGC implementation specifications can be tested for OGC specification compliance. Vendors of OGC compliant products are then licensed to use the OGC Trademark. This

---

[6] Web Feature Service (WFS): An interface specification for describing data manipulation operations on features (OGC, 2003a).

process offers users the capability of the products (OGC, 2003c) and instills faith in the products by assuring interoperability.

These initiatives by OGC have provided a much needed framework for building the next generation of GIS. This is evident by the steadily increasing number of internet mapping services available today. Internet mapping applications connect various data servers and provide users with a seamless view of the available datasets without being concerned of proprietary format. Further, these applications provide downloading capabilities and benefit the casual user by providing minimal analysis capabilities such as visualization, adding new layers, and panning. Although these applications are a step closer to interoperability, they do not provide sophisticated functionality for the advanced GIS users. Lack of support for semantic interoperability at this stage, and slow migration of organizations to comply with OGC specifications leaves the user to resolve interoperability issues on an ad hoc basis. The Ministry of Water, Land and Air Protection in BC and the Department of Ecology in the United States, for example, have build internet mapping applications for water well information. These applications provide simple functionality, like adding layers, panning, locating water wells and downloading well information. Although they provide downloading capabilities, water well information can be downloaded as .PDF files. As such, analysis would first require converting these data into a digital format. The GIS community, however, is slowly adhering to OGC specifications. Achieving a truly interoperable world, however, will require widespread acceptance and a concerted effort by all concerned parties (Sondhiem et al., 1999).

- **ISO:** The International Standards Organization (ISO) established the ISO /TC211 technical committee in 1994 to oversee the development of standards in the field of digital geographic information. The ISO/TC211 standards are a structured set of standards covering all aspects for information concerning geographic objects or phenomenon. These standards are modeled using the unified modeling language and are abstract standards that do not provide implementation details (ISO, 2003). They emphasize a service-oriented view of geoprocessing, which balances data, task and systems (Tsou and

Buttenfield, 2002). Today, organizations like OGC and ISO have liaison status and are working together to develop new standards for the geospatial domain (Albrecht, 1999).

- **Metadata:** Although metadata standards have been developed by various standardization organizations it is included in this section as it a fundamental component of interoperability. Metadata, defined as data about data, is a key factor in enabling data sharing by facilitating data discovery and reuse of datasets. It includes information regarding data availability, fitness for use, access and transfer (Guptill, 1999), content, quality, producer, and lineage of the dataset. Metadata standards have been developed by national and international standardization organizations, such as FGDC (FGDC's metadata content standard), Australia and New Zealand Land Information Commission (ANZLIC metadata standard), OGC /ISO (OGC /ISO/TC-211 Metadata Standard 19115) to provide a consistent terminology and methodology for describing datasets. Although these standards provide information regarding identification, data quality, spatial reference, attributes information, metadata reference, and distribution, they differ in their mandatory and conditional information and lack information regarding its purpose of use which is now included in the ISO metadata standard (OGC, 2003d). A comparison of metadata standards, in terms of it content can be found in Kim (1999).

  Metadata records, in conjunction with data catalogs, form the basis for clearinghouses, which provide an online mechanism for accessing metadata records. As these metadata records are described in HTML format searches for relevant data and can result in useless information (Guptill, 1999). Emerging technologies like XML, which separate information from presentation, are now providing new methods for describing metadata for 2D as well as 3D geographic information (Houlding, 2001).

The development of standards is a complex task, involving a lengthy developmental phase and an even longer acceptance phase. The importance of standards can be summed up by Kleinrock's (1992) quote: "Standards efforts are almost always slow, laborious, political, petty, boring, ponderous, thankless, and of utmost criticality" cited in Tosta (1994). OGC activities are the most prominent standardization activities today (Albrecht, 1999) and organizations are making a move to adopt OGC specifications. It remains to be seen, however, whether OGC specifications will fulfil their high expectations (Albrecht, 1999). Despite these new concepts of standardization the importance of national standards is underscored by its role in achieving interoperability at a time when information technology was limited in scope. Although these standards provided a method for exchanging information, Kuhn (1994) argued that spatial data transfer standards only formalized the lexicon and syntax and lacked semantic translation, which enabled users to share information but not meaning (Kuhn, 1994). Despite these deficiencies, standards were regarded as key elements of the integration process (Guptill, 1999) and building blocks of Spatial Data Infrastructure (Taylor, 1996; Hogan and Sondheim, 1996).

2.4 Spatial Data Infrastructures

Until the early 1990s, geographic information was predominantly collected and maintained by government organizations. "GIS technology transformed geospatial data handling capabilities and made it necessary for government organizations to re-examine their role with respect to the supply and availability of geographic information" (Masser, 1998). Technological advances during this time resulted in a proliferation of GIS technology in both the public and the private sectors. Consequently, many organizations were involved in collecting and maintaining geospatial datasets. A lack of coordination between these organizations, however, resulted in a cacophony of spatial datasets. In recognition of geographic information as a national asset (Barr and Masser, 1997), efforts to minimize redundancy and efforts to provide coherence for the geospatial data activities (Rhind, 1999; Tosta, 1994) resulted in national initiatives called Spatial Data Infrastructures (SDIs). SDIs, defined as policies, technology, data standards and people, were being initiated around the world in parallel to the development of standards its role in facilitating interoperability is discussed in this section.

The concept of a SDI was formulated in the US in 1990 when the interagency Federal Geographic Data Committee (FGDC) was set up in response to Circular A-16[7], which was issued by the Office of Management and Budgets (OMB) to coordinate the use and dissemination of geospatial data at a national level (FGDC, 2003). In 1994, President Clinton's Executive Order No 12096 entitled "Coordinating Geographic Data Acquisition and Access: The National Spatial Data Infrastructure (NSDI)" set the stage for the development of the NSDI, and assigned FGDC as the coordinator of NSDI activities (NSDI, 2003). The Executive Order identified the development of standards, the National Geospatial Data Clearinghouse and Framework Data as the key elements for realizing NSDI goals. It also provided a political backing for standardization (Rhind, 1999) and mandated all geospatial data producing agencies to document their datasets and make them available electronically. This initiative, backed by the open policy for access to information (Tosta, 1997), has contributed to the development of a thriving GIS industry in the US by promoting value added businesses.

In Canada, SDI activities were initiated in 1995 (Masser, 1998), and the Canadian Geospatial Data Infrastructure (CGDI) was established in 1999. GeoConnections, a Natural Resources Canada program, was established by the Federal government and industry to oversee the development of CGDI. Data sharing is promoted through the development of partnerships, policies, standards, framework data and online

---

[7] Circular A-16 was originally issued in 1953 by Bureau of Budgets (now called Office of Management and Budgets) to reduce duplicative efforts of mapping and surveying agencies. This was revised in 1967 and late re-revised in 1990 to include digital geographically referenced data (OMB, 2004). The 1990 revision of Circular A-16 also called for the development of a 'national spatial digital information resource' (Maitra, n.d.).

geospatial data access (GeoConnections, 2003). Copyright and data dissemination policies, however, have prevented Canada from realizing the benefits of data sharing activities.

Similar activities to build national SDIs were also observed in other parts of the world, including the UK, Australia, Netherlands, Japan, and Korea (Masser, 1999; Mohammed, 1997). Although the common goals of these SDIs were to facilitate data sharing through collaborative activities, based on a firm foundation of standards and facilitated by networked access to geospatial data, a number of factors have prevented SDIs from realizing their goals (Nedovic–Budic and Pinto, 2001; Chan et al., 2000). National and state level policies that provide a framework and a guide for data sharing activities (Nedovic-Budic and Pinto, 2001) are one of the factors that either hinder or promote collaborative efforts. Some countries, such as Canada, England, Australia and New Zealand, follow a restrictive data policy (O'Donnell and Penton, 1997; Rhind, 1999; Mooney and Grant, 1997; Robertson and Gartner, 1997). In these countries, spatial datasets, which are created by the federal and state governments, are subject to crown copyright and a license fee. These data policies, complemented by the lack of uniform geospatial data pricing, have stifled the use and proliferation of GIS and created an inconsistent spatial data culture in Canada (Klinkenberg, 2003). In the US, however, the open data access policy provides a geospatial data culture that is in stark contrast to the situation existing in Canada. Lack of copyrights for federal datasets and dissemination for the cost of production, have created a thriving GIS industry founded on NSDI activities (Nedovic–Budic and Pinto, 2001).

The lack of a mandatory component to the dissemination of data through clearinghouses and the implementation of metadata standards (Nedovic–Budic and Pinto, 2001) are other factors that work against the very essence of curbing redundancies. In Canada, for example, organizations are not mandated to document or distribute their data through the Discovery Portal, a facility that provides online access to geospatial data. Further, Masser (1999) identified a lack of awareness of SDI activities as being most critical for the success of SDIs. Despite these challenges, the many benefits afforded by SDI activities have resulted in local (Harvey et al., 1999a), regional (Mohammed, 1997) and global (Coleman and McLaughlin, 1998) SDI. Thus, while SDI activities coalesced and brought organization to the GIS community (Tosta, 1997), standards provided a common format for exchange of data. Although these activities provided a mechanism of exchanging data they did not address issues of integrating information.

2.5 Database Integration

Integration of databases and information resources are central to achieving interoperability (Devogle, 1998; Sheth, 1999). SDI's provided a framework for data sharing activities, and while standards define an ideal target for conversion, they do not address the problems of integrating data from diverse sources (Devogele et al., 1998). Integration of data from heterogeneous sources is a challenging task that has been the focus of research within the computer science and GIS domains for many years. Despite extensive research no definite solution or consensus of approach has emerged (Widom, 1995). Moreover, semantics have remained the most challenging aspect of data integration.

2.5.1 Data Heterogeneities

Data heterogeneities inherent in the component database have been identified as the source of data integration problems (Stock and Pullar, 1999; Vckovski, 1999; Bishr, 1998; Sheth and Larson, 1990). Although various types of data heterogeneities have been identified by researchers, this research will adopt Bishr's (1998) data heterogeneity classification. Bishr identified three types of heterogeneities:

- **Syntactic heterogeneity:** Syntactic heterogeneity stems from the use of different data models, such as relation or object oriented models, to represent database elements or the use of raster and vector models (Bishr, 1998). Elevation information, for example, can be represented in raster format as Digital Elevation Models or as contours in vector format.

- **Schematic Heterogeneity:** Schematic heterogeneities result from different classification schemes employed in the component databases or structuring of database elements in component databases. For example, in this research the geological description of the well logs are represented by a single attribute in the BC database and with three attributes in the American database. Schematic heterogeneities also result from different definitions of semantically similar entities, missing attributes, and different representations for equivalent data. A detailed classification of schematic heterogeneities can be found in Kim and Seo (1991).

- **Semantic Heterogeneity:** Semantic heterogeneity occurs when there is a disagreement about the meaning, interpretation or intended use of the same or related data (Sheth and Larson, 1990). This heterogeneity results from the different categorizations employed by individuals when conceptualizing real world objects. Such categorizations differ between individuals depending on education, experience and theoretical assumptions (Stock and Pullar, 1999). For example, an environmentalist may perceive water bodies as areas that require protection whereas the tourism industry may perceive these water bodies as recreational areas. An example from my research is geological descriptions and subsurface lithologies. A driller's sand may not necessarily be another driller's sandy silt. Semantic heterogeneities have been identified as the main cause of data sharing problems and are the most difficult to reconcile (Bishr, 1998; Vckovski, 1998; Kottam, 1999).

2.5.2 Semantic Interoperability

Until the 1990s syntactic and schematic heterogeneities were the primary focus of interoperability research and concentrated on issues related to data models and database structures (Visser et al., 2002b; Sheth, 1999). By the mid-1990s increasing global interconnectivity had caused a proliferation in the number of databases from a few sources to millions of information resources (Sheth, 1999). As such, the focus changed to semantic issues (Sheth, 1999). It was soon recognized that resolution of syntactic and schematic heterogeneities first required semantic reconciliation (Bishr, 1998; Sheth, 1999). In the context of geographic information, semantic reconciliation poses a greater problem due to its complexity and diversity of spatial representation. Given that the origins of semantics lay in human conceptualizations of space, Harvey et al. (1999b) proposed the need to resolve semantic issues in a holistic fashion, drawing conclusions from domains such as computer science, social science, cognitive science and linguistics.

The dependence of GIS application on databases has meant semantic interoperability solutions are heavily reliant on techniques from computing science. Schuurman (2002) divided these techniques into the federated environment, mediator and the linguistic approaches. Based on the approaches described by Schuurman (2002) and the research framework proposed by Harvey et al. (1999b), this section explores the federated, mediator, context, ontological and cognitive approaches to resolving semantic interoperability issues.

- **Federated Approach:** The federated database environment offers one of the most common approaches to achieving semantic interoperability (Sheth and

Larson, 1990). In this approach export schemas[8] of component databases are integrated as a method to resolve semantic conflicts. Solutions based on the federated environment have addressed semantic issues from a data or systems perspective. Devogele et al. (1998) proposed techniques for developing inter-schema correspondences between data objects as a means to resolve semantic heterogeneities, while integrating databases of different scales. Similarly, Laurini (1998) explored the issues of integration of databases from a representational perspective, issuing solutions for maintaining geometry and topology of semantically similar elements. On the other hand, Abel et al. (1998) focused on achieving system interoperability and presented a design for a virtual GIS based on an object oriented data model for the seamless integration of heterogeneous repositories to facilitate transparent data access. Semantic interoperability solutions based on federated environment provide a static integration methodology (Leclercq et al., 1999) and do not provide a solution for dynamic resolution of semantic conflicts.

- **Mediator Approach:** Dynamic resolution of semantic conflicts can be achieved using the mediator approach, which consists of a mediator and wrappers (Voisard and Jurgens, 1999; Widom, 1995). In this approach mediators reconcile semantic difference between the client and component databases and wrappers provide communication between the mediator and component databases. This mediator-wrapper architecture was employed by Bishr et al. (1999) to resolve semantic differences of transportation data contained in GDF[9] and ATKIS[10] data files. Visser et al. (2002a), however, argued that

---

[8] Export schemas: is a subset of the component schema (component schema is derived by translating the local schema into the data model of the federated database system) a database participating in the federated database system (Sheth and Larson, 1990).

[9] Geographic Data File (GDF): is the European standard used for the transfer of road network and road related data (GDF, 2003). It is also briefly described in the Spatial Data Standards section.

[10] Authoritative Topographic-Cartographic Information System (ATKIS): is a topographic and cartographic standard of Federal republic of Germany (Harvey et al., 1999b).

mediators resolve schematic conflicts, while the resolution of semantic conflicts plays a subordinate role.

- **Context Approach:** Researchers have explored the use of context to resolve semantic issues (Kashyap and Sheth, 1996), as it provides a method for differentiating the meaning of terms. For example, based on context the term 'cricket' can be identified as a game or as an insect (Sheth, 1999). Bishr (1998) has used both the mediation and context approaches to resolve interoperability. In so doing, he has introduced the concept of 'proxy context,' which acts as a mediator between two or more contexts. Comparison between the contexts is then achieved using semantic translators. In another approach, Kashyap and Sheth (1996) used context to identify semantically similar data in a federated environment. They developed the concept of a 'semantic proximity' descriptor to capture the degree of similarity between semantically similar elements. In this approach the context of comparison plays an important role. When a query is posed to the federated environment the context of the query is compared with the context of a database to which it is directed. The context derived from ontologies (discussed below) is represented as a set of attributes and the role[11] played by the objects. Two objects are considered semantically similar when the roles of the objects in the databases are equivalent. Kashyap and Sheth (1996) have provided the example of employee databases where the *role* of employee object in database1 is to identify employees by their name and the *role* of the object employee in database2 is to identify employees by their number. The employee objects are considered semantically similar as they share the same role, which is to identify employees.

- **Ontologies:** The use of context in mediators/wrapper or the federated approach for achieving semantic translations has often been based on the use of ontologies. Today there is no consensus on the definition of ontology, instead it is based on the context in which it is used (Winter, 2001). Ontologies have traditionally been studied in philosophy; in that context ontology means

[11] Role: refers to the relationship between an object and its semantic context (Sheth and Larson, 1996).

existence of entities or the nature of being (Merriam Webster Dictionary; (Guarino, 1998). Ontologies in the computer science community are defined as 'an explicit specification of a conceptualization' (Gruber, 1995) or a 'shared understanding of some domain of interest' (Unschold and Gruninger, 1996). Ontologies were first introduced in Artificial Intelligence studies and their use has since been explored by researchers for a variety of applications, such as communication, information integration, system engineering, and database theory (Unschold and Gruninger, 1996).

The use of ontologies for resolving interoperability issues in the GIS community has recently received considerable attention (Visser et al., 2002b; Kokla and Kavouras, 2001; Smith and Mark, 2001; Bishr et al., 1999; Fonseca et al., 2000). As ontologies provide a shared understanding in the form of vocabulary of terms (Kashyap and Sheth, 1996) they have been used to mediate semantic differences and provide the building blocks for semantic translators (Bishr et al., 1999; Visser et al., 2002b). The development of a stratigraphic ontology for the description of stratigraphic layers, for example, can help to mediate differences between the transportation (databases storing information on bridge and overpass construction sites), stratigraphic and hydrogeological databases. Use of ontologies is not only restricted to the development of semantic translators. Their use was explored by Fonseca et al. (2000) for the development of software components to facilitate data integration. However, a prerequisite for use of ontologies for data integration is that there should be an agreement on the terminology used, either through the definition of ontologies or an ontological commitment among domains (Fonseca et al., 2000). Although these approaches provide an advanced method for resolving semantic issues, they are still confined to the scientific community and have relied on object oriented technology, whereas, bulk data are maintained in relational format (Schuurman, 2002).

- **Cognitive Approaches:** Resolving semantic issues has not only been the purview of computer scientists. Today, cognitive, linguistic and social scientists are also investigating methods to provide a better understanding of semantic issues (Harvey et al., 1999b). The cognitive approach to resolving semantic interoperability involves understanding the process of developing categorizations of space. Semantic conflicts originate from the varied categorizations employed by humans', thus; cognitive science theories have been explored by researchers to resolve semantic issues. In one such approach, Stock and Pullar (1999) used the theory of concept attainment to develop a methodology for identifying semantically similar elements. In this approach semantics of database elements are represented as predicates of different database elements, which can then be compared to determine the similarity between objects. In another approach, Frank and Raubal (1999) explored the concept of image schemata[12] to facilitate interoperability by focusing on the use of image schemata as a method to provide formal definition of spatial relations. Such spatial relations are important for standardization and interoperability of GIS (Frank and Raubal, 1999).

Despite extensive research, semantic interoperability solutions are in the prototyping stage. Although bulk data are maintained in relational format, semantic interoperability solutions call for data to be modelled using object oriented techniques (Schuurman, 2002). As no definite solutions exist, there is a general inertia among organizations to achieve interoperability. A study of interoperability at the Ministry of Environment, Land and Air Protection (BC, Canada), for example, has revealed that the Ministry is not prepared to embrace interoperability unless it is incorporated into ESRI software (Schuurman, 2002). This institutional or societal component of interoperability is discussed in the next section.

---

[12] Image Schemata: Are defined as recurring, imaginative patterns to comprehend and structure their experiences while moving through and interacting with the environment (Frank and Raubal, 1999). Although the concept of image schemata is not well defined in the cognitive and linguistic literature (Frank and Raubal, 1999) researchers have used a working definition that image schemata describe spatial relations between objects (Frank and Raubal, 1999). Matching, merging, splitting, contact, link, centre – periphery are few examples of image schemata (Frank and Raubal, 1999). The theory of image schemata however has not been universally accepted (Frank and Raubal, 1999).

2.6 Organizational Interoperability

The ability and willingness to share information is affected by the behaviour and needs of individuals, organizations and institutions (Onsrud and Rushton, 1995). This societal or institutional component, which has received little attention in interoperability research (Onsrud and Rushton, 1995; Nedovic-Budic and Pinto, 1999), is the focus of this section.

Data-sharing offers a number of benefits, including cost saving, productivity, improved policy making and service (Nedovic–Budic and Pinto, 1993; 2001) and the reduction of redundant datasets. Despite these benefits, there has been a general inability and unwillingness to share data and information across organizations, which has been compounded by low levels of coordination (Nedovic–Budic and Pinto, 1999; 2000; 2001). NCGIA's research initiative I-9, Sharing Geographic Information between Organizations, set the stage for this research by identifying three key areas of research to promote data sharing activities: theories of individual and organizational behaviour, arenas among which data sharing occurred, and observations of existing geospatial data sharing activities (Onsrud and Rushton, 1995). Since then extensive research has been conducted on resolving inter-organizational issues, drawing solutions from diverse arenas, including organizational theory, exchange theory, inter-group dynamics, legal and political policies. Despite these research initiatives, study of information sharing within the context of a GIS environment is still in its infancy (Pinto and Rushton, 1995). Even today little is known about the detractors and facilitator of successful data sharing activities (Nedovic–Budic and Pinto, 2001).

Achieving inter-organizational interoperability is dependent on organizational dynamics (Azad and Wiggins, 1995). Inter-organizational cooperation occurs within complex settings and is usually associated with changes that may involve loss of autonomy, increased interdependence and redefinition of existing tasks, and redistribution of powers (Evans and Ferreira, 1995; Azad and Wiggins, 1995). The prospect of such changes may ultimately affect the success or failure of collaborative relationships existing between organizations.

To understand these complex and dynamic inter-organizational relationship Pinto and Rushton (1995) and Azad and Wiggins (1995) proposed several antecedents for information sharing, such as organizational rules, incentives[13], goals, exchange relationships (characterized by trust and satisfaction, dispersion of power among stakeholders, external mandate and redistribution of power). Kevany (1995) studied environments conducive for sharing information and developed a quantitative scale for measuring the likelihood of success in sharing information, while Nedovic-Budic and Pinto (2000) studied the mechanisms and factors affecting the coordination and use of geospatial databases.

While these approaches have primarily concentrated on organizational or technical issues, Evans and Ferreira (1995) argued that as we are living in a world where technology is rapidly evolving interoperability research should concentrate on the overlap of technical and organizational issues.

2.7 Summary

This chapter has provided a summary of a wide spectrum of approaches to achieve interoperability. Although a lack of common understanding and different interpretations of interoperability prevail, there is a consensus that interoperability facilitates information exchange from heterogonous sources. This process is dependent on data models, structures, formats and semantics.

[13] Incentives include funding, access to information and payoffs (Pinto & Rushton, 1995).

Interoperability issues can be traced back to the 1970s and 1980s when data were maintained in proprietary formats (Sondheim et al., 1999; Bishr, 1999; Guptill, 1991). Spatial datasets were the purview of government organization due to high hardware, software and associated collection and maintenance costs. Data collected and maintained to meet specific departmental needs resulted in redundant datasets. Realizing the wastage of economic resources, government organizations initiated standardization activities around the world and resulted in national transfer standards (Salgé, 1999; Moellering, 1996; 1991). The limitations of national standards for transnational issues prompted application specific international groups to develop de-facto standards (Salgé, 1999). In addition to the development of standards, the changing attitude of the government organizations with respect to spatial information (Masser, 1998) and the need to provide organization for the multitude spatial datasets resulted in the development of national SDIs. This infrastructure for spatial data provided a framework for data sharing activities through the development of standards, policies and electronic access of spatial datasets. Access to information and data dissemination policies, in countries like Canada, however, has stifled the growth of GIS industry (Klinkenberg, 2003) and prevented the SDIs from realizing their true potential. Despite data sharing activities promoted by SDIs, institutional conflicts resulted in stalled initiatives (Onsrud and Rushton, 1995). This institutional aspect of interoperability has been researched by authors like Pinto and Ruston (1995), Azad and Wiggins (1995), and Nedovic-Budic and Pinto (1999; 2000; 2001).

Advances in technology like the Internet, object management and distributed computing architecture opened new avenues for achieving interoperability. These concepts were adopted by OGC to develop specifications using object oriented technology. Although standards and OCG specifications promoted information exchange, data semantics prevented its use for analysis and decision making (Kuhn, 1994). Semantic data heterogeneities were identified as the main cause of all data sharing problems (Sheth, 1999; Bishr, 1998). Various approaches to resolve semantic heterogeneity have been explored by authors like Frank and Raubal (2001), Bishr et al. (1999), Laurini (1999), Bishr (1998), and Kashyap and Sheth (1996).

To date, semantic research is in the prototyping stage and calls for data to be modelled using the object oriented paradigm (Schuurman, 2002). Although this research provides a futuristic solution for interoperability, it does not resolve the immediate problems faced by organization where bulk data are modelled using relational format (Schuurman, 2002). In addition, interoperability research from a technical and institutional perspective has tackled issues within a single jurisdiction (except OCG activities). Little research has been undertaken on the technical and institutional challenges faced by organizations from an interoperability perspective in an international setting. Sieber (2003) pointed out the lack of research in Public Participation GIS (PPGIS[14]) arena for cross-border and multinational applications, and although literature may exist for cross-border projects (Sieber, 2003) they address issues other than the challenges of data integration. Organizations like OGC have made significant contributions to resolve interoperability issues for an international market, but present day solutions cater to online applications for the layman GIS user. As such, they provide minimal analysis capabilities, including visualization, panning, adding layers and downloading data. Further, a lack of support for semantic issues at this stage leaves the advanced user to resolve integration on an ad hoc basis. Thus, information exchange for cross-border projects complicates an already complex issue.

---

[14] PPGIS: A research field that explores the use and value of GIS technology on marginalized people and communities interested in social change (Sieber, 2003)

Finally this literature has explored the activities of standardization activities in the Canadian, US and international contexts, and provides a summary of the multifaceted research initiatives to understand and resolve semantic interoperability. Furthermore, it has considered the institutional aspects and the spatial data policies prevalent in the two countries, which have developed strikingly different spatial data cultures that affect data integration in an international setting.

3 DATA INTEROPERABILITY ACROSS BORDERS

3.1 Introduction

Data that fuels the burgeoning GIS industry are based on different conceptual models, formats and structures and are unknown to most users. Inherent complexities and differences in the nature of geographic data are the source of most interoperability problems (Evans and Ferreira, 1995). Problems related to interoperability are further exacerbated for environmental phenomena that do not respect political boundaries.

Anthropogenic activities are progressively stressing the environment. Increasing incidences of flooding, drought, pollution, hurricanes, tornadoes, wild fires and earthquakes are of common occurrence today. Such phenomena do not follow political boundaries and are adversely affecting humans, property and the environment. Anthropogenic activities are also having a direct impact on ecosystems and wildlife. Research to protect the environment has necessitated cross-border investigations; however, these studies are dependent on the availability of integrated datasets. Although, integrated data are a prerequisite for such studies, little is know about the challenges and experiences of organizations sharing and integrating information across borders. My research integrates groundwater datasets from BC (Canada) and Washington State (US) to explore challenges of integrating cross-border datasets.

Groundwater is an environmental issue that does not follow jurisdictional boundaries: it is expensive to manage and requires a multidisciplinary approach that involves the development of collaborations and information-sharing amongst a number of agencies (Rivera et al., 2003). In Canada, approximately ten million people rely on groundwater, which is under stress due to increasing urbanization, climate change and contamination from anthropogenic activities (Rivera et al., 2003; Ricketts, 1993). Despite the increasing stresses upon groundwater in Canada, twenty-six years elapsed between the publication of the Brown report on groundwater conditions in Canada by the Geological Survey of Canada (GSC) in 1967, and the British Columbia Geoscience Council Report, 'A Groundwater Synthesis at a National Level' in 1993 (Sharpe, n.d.). The 1993 report concluded that Canada lacked a groundwater knowledge base at a national scale and characterized groundwater research in Canada as being fragmented and inadequate (Sharpe, n.d.). In response, the GSC launched two pilot projects: The Oak Ridges Moraine Hydrogeology Project in Ontario and the Brookswood Aquifer Project in BC. The primary objective of these projects was to develop a 3D hydrostratigraphic model to facilitate groundwater management activities.[15] Unlike the Oak Ridges Moraine Project, which has resulted in a standardized database that can be used by consultants and researchers for groundwater management activities the Brookswood Aquifer Project only resulted in a 3D hydrostratigraphic model for the one aquifer. These activities were not extended to other aquifers in BC.

The Abbotsford-Sumas aquifer is one of the largest aquifers in the Greater Vancouver area. Its location, straddling the Canada-US border, has resulted in the implementation of collaborative efforts on both sides of the border for the management of this aquifer. The Abbotsford-Sumas International Task Force, for example, is a coordinated initiative by Canada and the US to develop aquifer management strategies and educate the public about groundwater resources (MWALP, 2002). The decision by

---

[15] Water well information provided by the drillers to the provincial government was used as a primary source of information for constructing the hydrostratigraphic model. Geophysical studies were also conducted to aid in the development of the model. Although geophysical exploration provides an accurate method for identification of the subsurface, these methods are expensive and may be inaccessible for organizations with small budgets.

the task force to adopt a managerial approach, however, has meant that it does not address issues at a data level. The absence of interoperability solutions within an international setting, as well as the lack of a standardized database and geophysical studies, makes the Abbotsford-Sumas Aquifer a unique case study for research. To explore these issues, three objectives were considered:

1. To investigate the challenges encountered while integrating datasets for cross-border projects;
2. To create an integrated dataset for groundwater management activities; and
3. To develop a hydrostratigraphic model to evaluate the usability of the water well database.

Based on these primary objectives, this chapter addresses the challenges of integrating datasets for groundwater management activities. In conducting this research, datasets from Canada and US were integrated to build the hydrostratigraphic model, a process which can be divided into Pre-Processing, Data Conversion and Data Integration[16].

After a brief summary of the Abbotsford-Sumas aquifer (Section 3.2), this chapter discusses Pre-Processing in Section 3.3, Data Conversion in Section 3.4, Data Integration in Section 3.5, followed by concluding comments (Section 3.6).

3.2 Abbotsford-Sumas Aquifer

The Abbotsford-Sumas aquifer (Figure 3.1) is a shallow unconfined aquifer located partially in BC (Canada) and partially in Washington State (US). It is situated

[16] For the purposes of this research, Pre-Processing includes issues related to data acquisition and data quality, Data Conversion refers to the process required to convert datasets to a common format, and Data Integration includes issues related to data heterogeneity.

within the Fraser-Whatcom Lowland and is approximately 200 square km in area; shared equally by Canada and U.S (Cox and Kahle, 1999). This aquifer caters to the needs of approximately 200,000 people living in the area. The aquifer supports farming, industrial and domestic activities (Kohut, 1987). The aquifer is one of the largest aquifers in the area and is highly vulnerable to contamination.

There is extensive agricultural activity in the area, primarily in the form of raspberry and poultry farming, which has resulted in nitrate concentration measured in groundwater that exceed levels permissible by the Environmental Protection Agency (EPA) and Health Canada (Cox and Kahle, 1999). Consequently, there is a need for research that studies and monitors nitrate pollution within the aquifer. Groundwater management activities are complicated by the complexity of the aquifer. Specifically, the geological architecture is complex owing to the complex glacial history (Armstrong, 1981).

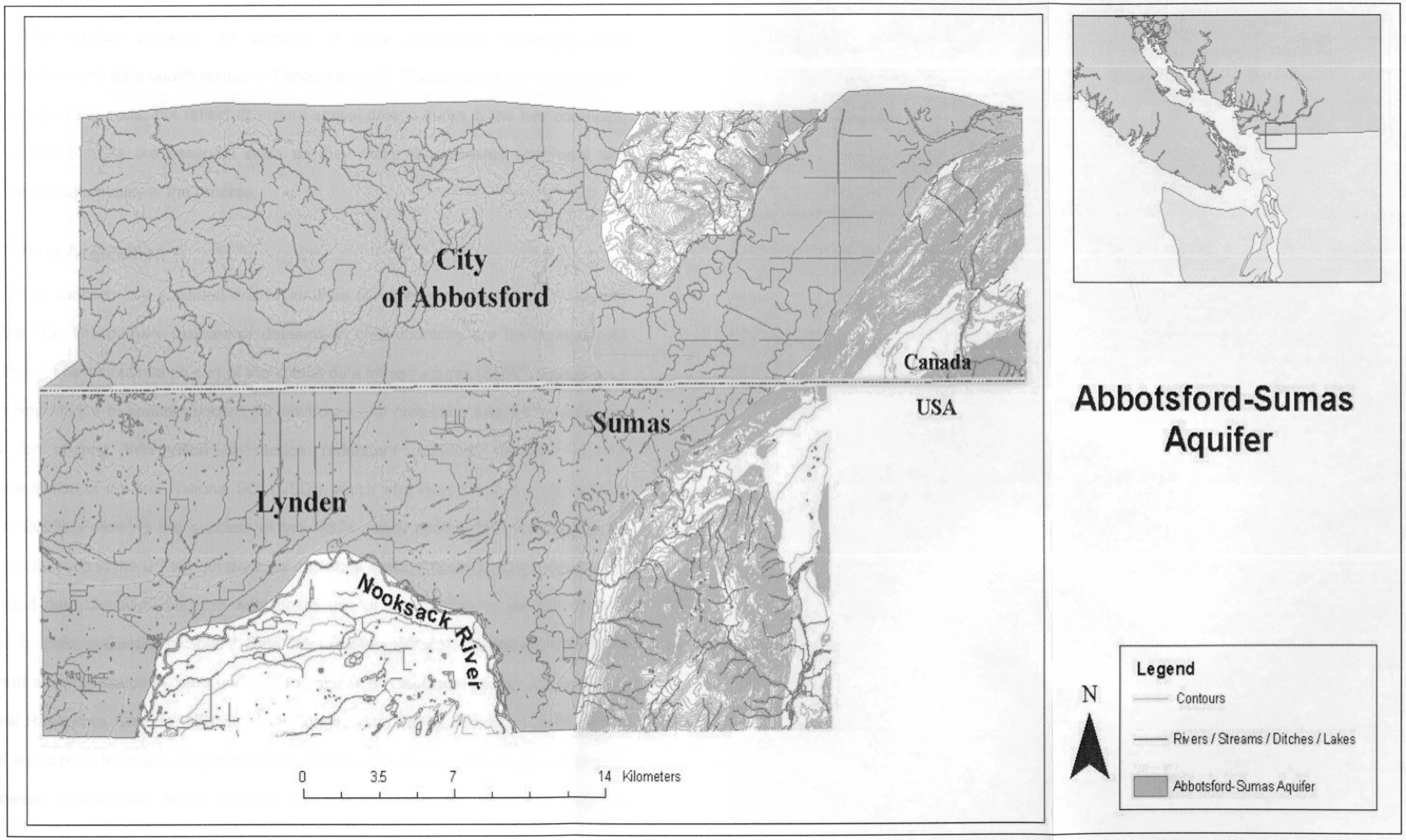


Figure 3-1 Abbotsford-Sumas aquifer

3.3 Pre Processing

This section provides an account of data acquisition, metadata, data dissemination, and data quality issues in Canada and US. These issues are fundamental to any project life cycle, but reflect divergent spatial data cultures in the two countries. Each of these tasks are essential steps as they make the datasets congruent and improve the data quality in the process.

3.3.1 Data Acquisition

GIS datasets are collected and maintained by various organizations in Canada and the US. The primary sources of datasets in both countries are the spatial data clearinghouses maintained as part of the spatial data infrastructures (SDI)[17]. Recognition of the importance of spatial datasets for environmental protection and the subsequent need for sharing information to reduce redundant activities, resulted in the implementation of the first National SDI (NSDI), which was instituted in the US in 1994 (NSDI, 2003). Following the success of the NSDI, many spatial data clearinghouses were established in the US at both the local and state levels. These clearinghouses have promoted data discovery and the widespread use of spatial data in the US. Similar efforts to build a Canadian Spatial Data Infrastructure (CGDI) initiated in 1995, and reached a formal status in 1999 under the auspices of GeoConnections, a program of Natural Resources Canada (NRCan) (GeoConnections, 2003). The main goals of the CGDI are to provide access to geospatial information, develop a national geospatial data framework, develop and adopt common international geospatial standards, improve

---

[17] In Canada, although a national spatial data clearinghouse does exist, very few organizations have submitted information about their products. As a result local knowledge is required for acquiring GIS datatsets.

partnerships between federal and provincial governments, and develop geospatial data policies (GeoConnections, 2003). Despite its implementation in 1999, there has been a general inertia among organizations to submit information regarding their spatial data products and services: In Canada, a total of 54 provincial ministries have submitted information regarding their spatial data products; only two entries are from BC. Despite the establishment of a CGDI, local knowledge is still required for acquiring datasets in Canada.

Data for this project were acquired from a number of sources and include spatial as well as attribute information[18] (see Appendix A). When datasets were unavailable for a region, maps were digitized using ArcGIS® v 8.3. Topographic datasets for the British Columbia portion of the study area were obtained from LandData BC; a service provided by Base Mapping and Geomatic Services (BMGS), which is a branch of the Ministry of Sustainable Resource Management (MSRM). These include digital topographic maps at a scale of 1:20,000 and Digital Elevation Models (DEM) at 25m interval. Data were obtained for map sheets 92G008, 92G008 and 92G010. Topographic datasets for Washington State were obtained from the United States Geological Survey (USGS) at a scale of 1:24,000. Digital maps were obtained for Sumas, Lynden, Kendall and Bertrand Creek 7.5 topographic quadrangles. DEM's were also obtained for the same map sheets (see Appendix A).

As few organizations submit information to the CGDI, lithological information for the study area was obtained directly from the ministries responsible for groundwater management in Canada and the US. In Canada, groundwater and its management is a

---

[18] Attribute information in this case refers to the lithological databases.

provincial, as opposed to a federal, responsibility. The lack of a national groundwater policy has resulted in provincial governments developing widespread groundwater policies (Piteau Associates and Turner Groundwater Consultants, 1993). In Ontario, Alberta and Newfoundland & Labrador, for example, it is mandatory for drillers to submit water well information to the provincial government, while in BC submission of water well information is currently a voluntary task[19]. Unlike Canada, the federal government in the US has some rights to groundwater, and thus, influence over its management is delegated to various federal government organizations (Piteau Associates and Turner Groundwater Consultants, 1993). In Washington State, groundwater information is maintained by the Department of Ecology and the USGS. Lithological information for this project was obtained from the following sources:

- **Ministry of Water, Land and Air Protection (BC):** The Ministry of Water, Land and Air Protection is the custodian of groundwater activities and undertakes groundwater related management activities in BC. When a water well is constructed, the driller is expected to submit information to the provincial government. Although voluntary, this process has resulted in the production of a large volume of water well records containing information on geology (i.e. lithology), water quality and water level information within the province of British Columbia. These records are maintained in the Computerized Groundwater Data System (CGDS) and are available directly from the ministry's website or through the internet mapping service developed by the Ministry. This database is maintained primarily to provide groundwater information to the public, researchers, consultants and water well drillers, and can be an invaluable source of geological information. Although these data are maintained in a computerized system, it is structured and distributed in a format that cannot be directly used for analysis. Consequently, the

[19] New Legislature under the Drinking Water Protection Act requires mandatory submission of water well drilling records; however, regulations have not yet been developed (Allen, 2004)

information can either be requested from the Ministry or may require laborious data entry.

Lithological data for the study area were downloaded into a Microsoft® Access v 2002 database using LDBuilder (Lithological Database Builder), a Graphical User Interface (GUI) developed by Dr. Schuurman (2002). The software downloads the data into the following three tables:

- **Lithology** (stores information on the geology)
- **Location** (stores locational information in terms of Universal Transverse Mercator (UTM) Coordinates)
- **General** (stores general information, including the address, owner, lot, concession, parcel, bedrock depth, and water table depth)

- **Whatcom County / Department of Ecology (Washington State):** This ArcGIS dataset, which includes scanned well log reports from the USGS, Whatcom County Health and Human Services, and the State of Washington's Department of Ecology, was compiled for the Watershed Management project (WRIA) instituted by the State of Washington's Department of Ecology in 1998 (Gill, 2002). The WRIA study area includes the drainage area of the Nooksack River and its tributaries, including parts of Skagit County, which are drained by the south fork of Nooksack River, the Sumas River drainage and the US portion of the Abbotsford-Sumas aquifer (WRIA, 2002). The objective of the WRIA was to develop plans for allocation of water, protection of water quality and restoration of the fish habitat in the WRIA study area. These scanned images can now be downloaded from the WRIA website (WRIA, 2002). Although this database has been assembled for WRIA activities, it has not been error checked or verified for overlap between the different sources of well logs (Gill, 2002).
- **USGS (Washington State):** This MS Access database was compiled by USGS to provide a comprehensive resource of the available geological data for survey investigators (Doremus, 2002). Although this database was complied from the original reports filed with Washington State's Department of Ecology, it only includes those wells that were in digital format or those

converted to digital format by the USGS staff (Doremus, 2002). The database stores four main tables:

- **Tblmaterial** (geology)
- **Tblwelldata** (location and general information like address, parcel)
- **Tblrecovery** (recovery test data)
- **Tblwelltest** (pumping test data)

The USGS database was used for this research as it was already in a digital format. A few well logs from the WRIA dataset were manually entered into the USGS database, because this database did not store all the well logs for the study area. One of the major challenges was cross-referencing the well logs in the Washington State Department of Ecology and the USGS databases, as they lacked common identification.

Various challenges were encountered while acquiring datasets for this research. Table 3.1 describes the data acquisition challenges for BC and Washington State datasets. Following is a brief explanation of the categories used in table 3.1.

- **Data format:** Data format refers to format of geospatial datasets. For example, thematic datasets, such as, surficial geology and soil maps were available only in paper format for the Canadian portion of the study area.
- **Data acquisition:** Refers to the knowledge required for accessing datasets.
- **Lack of cross referencing:** This refers to the lithological databases from BC and Washington State. Multiple groundwater datasets are maintained in Washington State and lack cross referencing between the water well records.
- **Institutional reluctance:** Institutional reluctance refers to the lack of awareness of activities within the same department or levels of governance.
- **Institutional priorities:** Institutional priorities of organization can result in maintenance of duplicate datasets. Varying institutional priorities of the Department of Ecology and the USGS resulted in separate groundwater databases for Washington State.

Interoperability Challenge	British Columbia	Washington State
Data format	Surficial geology and soil maps available in paper format	Maps available in digital format
Data acquisition	Local knowledge required	NSDI
Institutional reluctance	Not encountered	USGS
Lack of cross referencing	Only one database: maintained by MWLAP	Lack of cross referencing between the USGS and Department of Ecology datasets
Institutional priorities	Not encountered	USGS and Department of Ecology

Table 3-1 Data acquisition challenges

Thus, data for this research were acquired from a number of sources in Canada and US. As compared to the US where data are easily available via the clearinghouses, local knowledge is required for acquiring digital datasets in Canada. Data format is another important parameter for data integration. Soil and surficial geology datasets for BC were available only in paper format and, therefore necessitated digitizing. Finally the lack of awareness of activities in the same department or levels of governance, and institutional priorities resulted in duplicate lithological databases maintained by USGS and Department of Ecology, which lack cross referencing.

3.3.2 Metadata

The use of metadata records is an important component of the data discovery process via the spatial data clearinghouses. Metadata, or data about data, provides information about the content of the data set; its quality, the producer, its lineage, data availability, fitness for use, access and transfer and is essential for data sharing. Lack of metadata, and consequently information about the datasets, renders the data ineffective for analysis (Goodchild and Longley, 1999). Today, a few metadata standards exist for geospatial data including Open GIS Consortium / International Standards Organization (OGC/ISO TC 211) and the US Federal Geographic Data Committee's Content Standards for Geospatial Data (FGDC). However, as the OCG / ISO TC 211 metadata standard has yet to be ratified and most countries have adopted the FGDC metadata standard due to its high standards and popularity (Goodchild and Longley, 1999).

In the US, President Clinton's Executive Order 12906, which established the NSDI, also mandated all federal agencies to document their spatial datasets using the FGDC metadata standard and make it electronically available via the clearinghouses. Non-government organizations were also urged to do the same. Following this mandate, most organizations now distribute metadata in FGDC format. The mandatory components of this metadata standard provide necessary information to users and enable them to make an informed decision regarding the dataset. Although this standard does not provide information regarding the purpose or fitness for use (Guptill, 1999), the user is at least aware of the existence of the datasets. These clearinghouses have improved access to spatial datasets by providing private and public organizations with a method for disseminating their information over the internet, thus developing a flourishing geospatial industry in the US. Today several spatial data clearinghouses exist, which are advertising their spatial data services and products and promoting the growth of the GIS industry.

In Canada, the Discovery Portal of GeoConnections develops and maintains the geospatial data clearinghouse. The Discovery Portal is responsible for providing clients a search engine for the discovery of Canadian and International datasets. It also allows Canadian stakeholders to advertise and distribute their products and services to a national and an international market. In the absence of indigenously developed metadata standards, the CGDI has adopted FGDC and OGC/ISO metadata standards. These standards are the basis for data discovery via the Discovery Portal.

The Discovery Portal disseminates information about organizations and the services and data they provide. Organizations can submit information about themselves or the services they provide by listing basic information such as organizational profile, address and service information. Organizations wishing to advertise data products are only required to submit information in FGDC compliant format. Consequently, organizations wishing to advertise their services are not required to submit metadata information. LandData BC, for example, an Internet product ordering utility maintained by Base Mapping and Geomatic Services (BMGS)[20] has advertised information about their services through the Discovery Portal. Metadata information is available only on their website, and provides information regarding the synopsis, schema, spatial, sample, usage, constraint, methodology, lineage, contacts and related data. The metadata provided on the LandData BC website, however, lacks detailed information about the spatial attributes. This attribute information, which forms the basis of GIS analysis, can be browsed through a maze of links on the MSRM website. Direct links to these documents are conspicuously absent from the metadata. Metadata standards based on ISO TC 211 19115 metadata standard are being developed for MSRM data products. At this stage, however, the MSRM is concentrating on data discovery rather than fitness for use (Fulton, 2004).

Similarly, metadata for the cadastral datasets are available on the MSRM website in PDF format. The metadata contains information about the spatial files and the attributes, but it lacks accuracy information. The cadastral data does not contain

---

[20] BMGS is a branch of MSRM

information on private parcels, which is available from local municipalities. Until recently, separate cadastral datasets were maintained by various levels of government organizations and points to the challenges of working across different levels of governments. An initiative called the Integrated Cadastral Information Society has been established, which that is attempting to combine the cadastral datasets available from provincial and local governments. This information is currently available to members for a fee (ICIS, 2004).

Although the provincial government provides some information regarding their products, metadata may be completely lacking for other organizations. In the Greater Vancouver Regional District (GVRD)[21] metadata may be in the form of a text file, an ArcGIS format that is compliant with FGDC format, a self explanatory format or by word of mouth (Regier, 2002). Absence of metadata may have serious repercussions, including the unnecessary duplication of datasets and the misinterpretation of data and analyses.

Although GeoConnections is creating awareness among the Canadian GIS community through conferences and workshops, only 523[22] organizations have submitted information to the Discovery Portal (Discover Portal, 2003). To date, only two ministries in BC and seven municipalities in the entire country have submitted metadata information about their products. This may be attributed to a lack of knowledge about CGDI, as well as to economic factors. Short-staffed organizations, for example, may not have the resources to complete this labour intensive task. Creating and maintaining metadata have been identified as an economic constraint by Guptill (1999). The absence

---

[21] GVRD consists of twenty one municipalities and one electoral area.

[22] Organizations include municipal, provincial, federal, and commercial institutes.

of metadata records, which comprise the basis for data catalogues, is a major obstacle for data sharing because first, the user is unaware of the dataset, and secondly, is unaware of the purpose or semantics of the dataset. Guptill (1999) has argued that although the purpose of metadata is to provide information availability, fitness for use, access and transfer, metadata seldom address the issues of fitness for use. In contrast to the situation in the US, a lack of metadata for Canadian datasets and a lack of political backing to mandate electronic documentation of spatial datasets backed by data dissemination policies have stifled the growth of the GIS industry (Klinkenberg, 2003).

3.3.3 Data Dissemination / Access

The data obtained from the two countries differ in many respects with data prices and copyright issues. The spatial data dissemination policies have developed very diverse spatial data cultures in the two countries (Klinkenberg, 2003). These policies directly affect data access and, in turn, the economy of a country (Sears, 2001). Although the US follows an open free data access policy, where federal datasets are available at no cost to the user and do not require copyright permission, this free data access policy is not adhered at the county and local level where a fee may be charged (Nedovic--Budic and Pinto, 1999). Despite this unevenness in data prices, for the most part data are freely accessible to the public.

In contrast to the data dissemination policies in the US, Canada's data cost recovery policy and copyright issues have prohibited the widespread use of Geospatial data (Klinkenberg, 2003). This cost recovery policy has had negative consequences for the consumers and the economy, a lower level of research and development, and a higher cost for services and products (Sears, 2001). These cost recovery policies are further governed by jurisdictional constraints, which have developed an uneven balance in the country (Klinkenberg, 2003). The BC datasets used in this research cost approximately Cdn. $400.00 per map sheet whereas, the Washington State datasets can be obtained free of charge.

The cost recovery policy advocated in Canada has had an ancillary effect on data access. High data costs have limited data access for low budget organizations (Sears, 2001; Klinkenberg, 2003). Copyright issues and data redistribution policies have also had an economic impact and have limited the use of spatial data in Canada (Sears, 2001). These policies can directly affect cross-border projects where different spatial data cultures exist. Such policies have created problems in building the virtual database and floodplain management activities for the Red River Basin (Sieber, 2003).

Although metadata, data dissemination, copyright issues and data costs play a fundamental role in integrating datasets, data properties like data quality and formats are also problematic while integrating datasets.

3.3.4 Data Quality

Poor data quality is one of the many reasons that hinder the integration process: unusable formats and a lack of detailed information may result in misuse of dataset and can result in erroneous conclusions. Data quality is usually measured in terms of how data satisfies the needs of the users (Frank, 1998; Strong et al., 1997), and is defined by its fitness for use, which differs from person to person. Data quality is generally considered in terms of accuracy, precision, completeness, and consistency. This section does not follow the conventional analysis of data quality components; rather it addresses data errors that have prohibited the use of databases by groundwater consultants and users alike.

The use of the lithological databases has primarily been constrained by inconsistencies in the geological descriptions (Russell et al., 1998) and extremely poor data quality. Despite these problems, databases are maintained to provide groundwater information to researchers, public, consultants and drillers. These records have been primarily used by consultants and drillers to identify unsuccessful wells in a region (Symington, 2002; Livingstone, 2002; Dickin, 2002).

Following is a description of the data errors that were identified and resolved prior to the integration of the cross-border dataset. Such errors result in data inconsistencies, which that create problems during querying and analysis. Repeated data quality checks were necessary due to the extremely poor quality of the datasets.

3.3.4.1 CGDS Design

The design of the online water well entry interface used by the MWLAP is the source of most data quality problems; an issue that is unique to the BC lithological database. Water well information may be submitted to the provincial government using the online data entry form available on the Ministry's website (MWLAP, 2001) but is more often submitted as a hard copy record, from which data are subsequently entered by Ministry staff (Allen, 2004). In either format the field for geological description has a maximum length of 30 characters. If the driller's description exceeds the 30 character limit a new line entry must be created in the well record for the geological layer and the description is continued in the next line. In Figure 3.2 the black boxes highlight the continuation of a layer description. This style of representing the data is not consistent in all the records. In some cases, the depths are represented as null values in the new record (Figure 3.2, lower box) while in other cases (Figure 3.2, upper box) duplicate layer top and bottom depths are recorded. In either case, this method of data entry violates the rules of a relational database structure where one record is allowed per entity. Such errors are automatically corrected in the LDBuilder software, but without this software manual corrections can be very time consuming.

BCGS 092G010214 # 1 From 0 To 2 Ft. Top soil Lith Seq.# 1 WTN 000000019428
BCGS 092G010214 # 1 From 2 To 6 Ft. Boulder Lith Seq.# 2 WTN 000000019428
BCGS 092G010214 # 1 From 6 To 20 Ft. Gravel and boulders Lith Seq.# 3 WTN 000000019428
BCGS 092G010214 # 1 From 20 To 22 Ft. Gravel min. clay Lith Seq.# 4 WTN 000000019428
BCGS 092G010214 # 1 From 22 To 30 Ft. Glacial till (gravel & broken shale) Lith Seq.# 5 WTN 000000019428
BCGS 092G010214 # 1 From 30 To 34 Ft. Till - gravel in blue clay Lith Seq.# 6 WTN 000000019428
BCGS 092G010214 # 1 From 34 To 35 Ft. Till - gravel, shale in brown clay Lith Seq.# 7 WTN 000000019428
BCGS 092G010214 # 1 From 35 To 36 Ft. ~8" w-b gravel with clay interbeds Lith Seq.# 8 WTN 000000019428
BCGS 092G010214 # 1 From 36 To 38 Ft. Tight clay and gravel water sealed off Lith Seq.# 9 WTN 000000019428
BCGS 092G010214 # 1 From 38 To 39 Ft. Fine sand and gravel with clay interbeds Lith Seq.# 10 WTN 000000019428
BCGS 092G010214 # 1 From 38 To 39 Ft. (no water) Lith Seq.# 11 WTN 000000019428
BCGS 092G010214 # 1 From 39 To 43 Ft. Tight sand and gravel with clay inter- Lith Seq.# 12 WTN 000000019428
BCGS 092G010214 # 1 From 39 To 43 Ft. beds - W.B. .25 to .5 gpm Lith Seq.# 13 WTN 000000019428
BCGS 092G010214 # 1 From 43 To 45 Ft. Material appears freer Lith Seq.# 14 WTN 000000019428
BCGS 092G010214 # 1 From 45 To 50 Ft. Gravel in clay, water cut off Lith Seq.# 15 WTN 000000019428

BCGS 092G008111 # 19 From 0 To 14 Ft. Brown clay Lith Seq.# 69 WTN 000000019854
BCGS 092G008111 # 19 From 14 To 107 Ft. Blue silty clay with pebbles Lith Seq.# 70 WTN 000000019854
BCGS 092G008111 # 19 From 107 To 126 Ft. Till Lith Seq.# 71 WTN 000000019854
BCGS 092G008111 # 19 From 126 To 155 Ft. Silty sand and gravel (wet) Lith Seq.# 72 WTN 000000019854
BCGS 092G008111 # 19 From 155 To 176.5 Ft. Tight silty sand and gravel (dry) Lith Seq.# 73 WTN 000000019854
BCGS 092G008111 # 19 From 176.5 To 184 Ft. Fine to medium sand, some gravel, Lith Seq.# 74 WTN 000000019854
BCGS 092G008111 # 19 From 0 To 0 Ft. streaks of silt (W.B.) Lith Seq.# 75 WTN 000000019854
BCGS 092G008111 # 19 From 184 To 0 Ft. - ? Tight sand and gravel Lith Seq.# 76 WTN 000000019854

Figure 3-2 CGDI design[23]

3.3.4.2 Number of Records

An inconsistent number of wells in the various tables associated with a well record (i.e. lithology, location, general) were a problem common to both BC and Washington State water well record databases. In both databases, information on wells is distributed in different tables. Logically all tables should store equal number of well records; however, the tables stored inconsistent number of well records (Table 3.2). This issue was resolved after all the other errors were rectified.

---

[23] CGDI Design (**BCGS 092G010214:** British Columbia Geological Survey mapsheet coordinates; **From 38 To 39 Ft.:** Upper and Lower Bounds of Geologic Unit; **Geologic Description**; **Lith Seq#**: Sequence Number for each line entry; **WTN:** Unique Well Tag Number)

Mapsheet	Lithology	General	Location
92G008 (BC)	2310	2347	2302
92G009 (BC)	1418	1517	1517
92G010 (BC)	309	320	314
Washington State	937	1261	

Table 3-2 Inconsistent well records in the BC and Washington State databases

3.3.4.3 Missing Well Locations

Locational information in the form of coordinates is central to any GIS: without it spatial data are useless for analysis in a GIS environment. Although the BC database contains a table for the locational information, this crucial information for the well logs is often lacking[24]. Approximately ninety-nine percent of the records examined in this project were missing locational information. This information was obtained separately from a shapefile[25] (digitized well locations) provided by the MWLAP. The locational information was provided in Albers Equal Area projection (Datum NAD 83), which was then re-projected to UTM (Universal Transverse Mercator, NAD 83) projection. Although this dataset stores well locations, the associated attribute table for this dataset lacked coordinate information. Coordinate information for well locations were extracted using a Script (Add X Y Centroid) developed by Trent Hare (USGS) and can be downloaded from the Environmental Systems Research Institute (ESRI) website. Upon executing the script, coordinate information is automatically populated into the attribute table. Approximately 250 well locations were not digitized in the original shapefile provided by the Ministry. Locational information for these wells was obtained from DMTI Street Network file. The locations for only those records for which an exact match was possible

[24] MWLAP has an ongoing effort to provide locational data to the water well database (Allen, 2004).

[25] Shapefile is an ArcView format.

were obtained, and then inserted into the location table. In all, the locations for 116 wells were obtained from the DMTI Street Network Files[26].

Although the Washington State database stored locational information, it was expressed in UTM (NAD 27) coordinates. Before analyzing these data the information was updated to store UTM (NAD 83) coordinates.

3.3.4.4 Missing Elevations

The layer depth information (recorded as a depth below ground surface for each of the top and bottom of a geologic layer) contained null values, in approximately 20 to 30 percent of the wells records (Figure 3.3, 3.4). Such wells were deleted as they do not provide useful information in terms of the geologic layering i.e. only a general (ambiguous) representation of geologic layering is represented. A similar problem was the lack of geologic descriptions for the layers recorded (Figure 3.4). These wells were also deleted. The occurrence of missing depth information significantly reduced the usability of the database.

WellTagNum	UpperBound	LowerBound	LayerOrderNum	LayerDepth	Description
000000005629	0	0	1	0	Glacial clay
000000005620	0	0	1	0	Glacial
000000005625	0	0	1	0	Springs in clay - Whatcom
000000005714	0	0	1	0	Whatcom or Newton over Surrey ?

Figure 3-3 Missing elevations (wells composed of single record)[27]

---

[26] DMTI Spatial is Canada's leading provider of geospatial data products and services (DMTI Spatial, 2003).

[27] This figure is an output from LDBuilder. The BCGS number is omitted in the LDBuilder output. **Welltagnum**: corresponds to WTN; **UpperBound**: corresponds to From # (depth); **LowerBound**: correspond to To # (depth); **LayerOrderNum**: is a new field that stores the sequence number for the geologic layers. Although the lith Seq # serves the same purpose, it was not used due to

WellTagNum	UpperBound	LowerBound	LayerOrderNu	LayerDepth	Description
000000006590	0	0	1	0	null
000000006590	0	0	2	0	null

Figure 3-4 Missing elevation (wells composed of two layers)

3.3.4.5 Missing Sequence Numbers / Layer Depths

A lack of Sequence numbers and layer depths was a problem unique to the Washington State database (Figure 3.5). This information is crucial for querying the databases and for creating the hydrostratigraphic model. This information was inserted into the database by writing Visual Basic Scripts (Figure 3.6) (see Appendix B).

WellId	MaterialFrom	MaterialTo	Material1	material2	Material3
1000	59.4000015259	57.4000015259	Topsoil		
1000	57.4000015259	49.4000015259	Sand		
1000	49.4000015259	34.4000015259	Sand		
1000	34.4000015259	30.4000015259	Sand		
1000	30.4000015259	30.4000015259	Clay		

Figure 3-5 Original Washington State lithology table lacking sequence number/ LayerOrderNum and layer depths[28]

WellId	MaterialFrom	MaterialTo	LayerOrderNum	Material1	material2	Material3	layerdepth
1000	59.40000153	57.4000015	1	Topsoil			2
1000	57.40000153	49.4000015	2	Sand			8
1000	49.40000153	34.4000015	3	Sand			15
1000	34.40000153	30.4000015	4	Sand			4
1000	30.40000153	30.4000015	5	Clay			0

Figure 3-6 Washington State lithology table showing the sequence / layerOrderNum and layer depths

---

inconsistent Lith Seq numbering. **Layerdepth:** is a new field which stores the depth of the layers and **Description:** stores geologic description.

[28] **WellID**: is equivalent to Welltagnum; **Material, Material2 and Material3**; is equivalent to Description; **MaterialFrom:** is equivalent to UpperBound and **MaterialTo:** is equivalent to LoweBound.

3.3.4.6 Elevation Information

Depth information of each layer encountered during drilling is recorded from the youngest layer (top) to the oldest layer (bottom). Surface elevation is not taken into consideration. The surface elevations for well logs were always represented by a null value (zero elevation) in both databases (Figure 3.7). This information is critical for calculating the actual depth of the well layers, and subsequently, in the construction of the hydrostratigraphic model. The surface elevation for each well was extracted from the Digital Elevation Model (DEM) using a script called Sp3dPntzVal.ave developed by ESRI, which can be downloaded from the ESRI website. This script uses a DEM and a point theme (well location) as input, and automatically extracts the elevation information and populates the attribute table. As the DEM was projected in UTM coordinates, the unit of measurement was in meters. This information was converted to feet (conversion factor = 3.3) and inserted into the lithology table.

WellTagNum	UpperBound	LowerBound	LayerOrderNum	LayerDepth	Description
000000026719	0	15	1	15	Clay (brown)
000000026719	15	75	2	60	Blue clay
000000026719	75	89	3	14	Till
000000026719	89	93	4	4	Water bearing sand and gravel
000000026719	93	94	5	1	Clay

Figure 3-7 Surface elevation represented as a null value

After obtaining the surface elevations (reported in metres above sea level) for the wells, the top and bottom elevations for each unit were calculated. This was achieved by writing a Visual Basic script that automatically calculates elevation. The Visual Basic script first calculates the total number of layers in each well, which acts as a counter for the number of iterations that will be performed on each well. Starting from the topmost layer and making elevation equal to the well elevation, the script uses the layer thickness to calculate the actual depths of the layers. This automatic calculation of elevations, however, was further hampered due to data inconsistencies. Several inconsistencies in the surface elevation representations were observed, thus creating problems during the automatic calculation of the elevation using the Visual Basic Script. Further data quality checks had to be undertaken. Examples of these inconsistencies are described below:

In most cases the surface elevations for the wells were represented by a null (zero) value with a corresponding LayerOrderNum equivalent to '1'. Although in some cases the LayerOrderNum indicated that it was the topmost layer, the UpperBound stored a non-zero number (Figure 3.8). Such records were identified by comparing the surface elevation of the topmost layer to the surface elevation of the topmost layer in the original well record. Well logs that showed similar values were manually corrected by verifying the surface elevation obtained from the DEM. When the difference between the values was greater than or equal to 2 feet an imaginary layer of unknown geology was introduced (Figure 3.9). If the elevation difference was equal to 1 ft. the surface elevation obtained from the DEM was used.

WellTagNum	UpperBound	LowerBound	LayerOrderNum	LayerDepth	Description
000000005669	80	96	1	16	Sand boulders some clay Good water at 80 S/L 36'

Figure 3-8 Surface elevation represented by non-zero value

WellTagNum	UpperBound	LowerBound	LayerOrderNum	LayerDepth	Description
000000005669	219.8136	139.8136	1	80	unknown
000000005669	139.8136	123.8136	2	16	Sand boulders some clay Good water at 80 S/L 36'

Figure 3-9 Introduction of a layer of unknown geology

Inconsistent descriptions of well layers also caused concerns during the automatic calculation of elevations (Figure 3.10). For example, in wells '000000074587' and '000000019427' the UpperBound and LowerBound in the first layer are represented by '0' values, and in well 00000032441 the UpperBound for the fourth layer is represented by '0'. Such inconsistencies are problematic while querying the database and for building the hydrostratigraphic model. Such errors were manually corrected before updating the elevations using the Visual Basic Script (see Appendix B).

WellTagNum	UpperBoun	LowerBounc	LayerOrderNum	LayerDepth	Description
000000019427	0	0	1	0	Boulders and broken rock at surface
000000019427	0	0.8	2	0.8	Top soil (.8=10")
000000019427	0.8	10	3	9.2	Brown clay and boulders
000000019427	10	12	4	2	Broken rock and clay
000000019427	12	16	5	4	Shale and black rock formation with clay
000000019427	0	0	6	0	interbeds 16' water seeping in .25 gpm
000000019427	16	19.5	7	3.5	Glacial till in broken rock formation
000000019427	19.5	22	8	2.5	Boulder
000000019427	22	23	9	1	Glacial till (more like rock)
000000019427	23	24	10	1	Rock .75 gpm Static L. 12'
000000019427	24	27	11	3	Rock with some clay
000000019427	27	28.5	12	1.5	Rock - W.B.
000000032441	0	28	1	28	Overburden slide mat
000000032441	28	80	2	52	Shale
000000032441	80	143	3	63	Very badly broken and caving - redrilled
000000032441	0	8	4	0	Overburden
000000032441	8	65.5	5	57.5	Broken rock

Figure 3-10 Inconsistent records

3.3.4.7 Final Well Records

The last step in data cleansing was to ensure that all the tables stored the same number of well records. Well logs not present in the lithology table were deleted from the other tables. Table 3.3 shows the total number of records that could be used for developing the hydrostratigraphic model. A forty percent reduction in the original lithological databases highlights the poor data quality of these dataset, which ironically was the chief source of depth-specific lithological information.

Map Sheet	Original	Final	% reduced
92G008	2319	1882	19%
92G009	1505	878	42%
92G010	314	127	41%
Washington State	1261	937	26%

Table 3-3 Total reduction of well records

Data quality parameters are an important consideration and an important component of the interoperability process. Apart from conventional data quality parameters database errors should be considered for any environmental management project. Database errors are more likely to occur in datasets that cater to the needs of a smaller user community. Data quality issues were one of the reasons cited by water well consultants, to have limited the use of the lithological database maintained by MWLAP. Such data quality issues not only create problems during database querying and integration, but also have a direct effect on the accuracy of the models.

3.4 Data Conversion

The Pre Processing steps were followed by data conversion. Data from the two countries were distributed based on the level of government, jurisdictional mandates and organizational requirements. These datasets, which were based on different datums, projections, and formats, had to be converted to a common format to facilitate data integration.

The Washington State spatial datasets were obtained from the USGS, which is a federal organization, mandated to distribute data in SDTS format. Canada follows a more liberal approach where the format for disseminating spatial datasets is based on jurisdictional constraints, although sophisticated standards like SAIF exist (Quakenbash, 2002). The BC spatial datasets were obtained from MSRM and are distributed in

MOEP[29] format, ESRI[30] Shape format or in SAIF format. These datasets not only differed with respect to the transfer format, but also in the Spatial Reference System. All spatial datasets were converted to ESRI Shape format having a reference system of Universal Transfer Mercator (Zone 10) using the Feature Manipulation Software version 2002. Table 3.4 describes the inconsistencies in the spatial datasets.

Improving the quality of the lithological databases and the conversion of the spatial datasets to a common format was followed by integration of the datasets. Although the use of standards such as SDTS facilitates data transfer between non-compliant platforms, they do not address issues of data integration (Devogele et al., 1998) or semantics (Kuhn, 1994). Converting datasets to a common format does not necessarily mean that data can be integrated and used for analysis. Data heterogeneities further inhibit the data integration process and will be discussed in the next section.

INTEROPERABILITY ISSUES	**USA**	**CANADA**
Data Interchange format	SDTS (mandatory by federal agencies)	SAIF / ESRI
Projection	UTM / SPCS	BC Albers / UTM
Datum	NAD 27	NAD 83
Metadata	FGDC Compliant (Mandatory)	FGDC Compliant (Voluntary)
Scale	1:24000	1:20000
Vertical datum	NGDV	CGDV 28

Table 3-4 Details of datasets consulted

29 MOEP: stands for Ministry of Environment and Parks, however, does not have bearing to the current ministry name (Ministry of Sustainable Resource Management) (Poire, 2004).

30 ESRI: Environmental Systems Research Institute

3.5 Data Integration

Academic inquiry into resolving data integration problems has been an ongoing concern of researchers yet no consensus of approach has emerged (Widom, 1995). Data integration has been a long-standing problem in the GIS community and has been attributed to data heterogeneities inherent in the databases. Such heterogeneities arise as different organizations develop datasets based on their organizational business model, and have been classified by various researchers as Syntactic, Schematic and Semantic heterogeneities (Sheth, 1999; Bishr, 1998; Stock and Pullar, 1999).

Syntactic heterogeneities arise when different data models are used, such as relational vs. object model or vector vs. raster models (Bishr, 1998). Schematic heterogeneities originate due to differences in database structure (Bishr, 1998; Kim and Seo, 1991), and semantic heterogeneity occurs when there is a disagreement about the meaning, interpretation or intended use of the same or related data (Sheth and Larson, 1990). This heterogeneity results due to the different categorizations employed by individuals when conceptualizing real world objects, and has been identified as the most difficult to resolve (Sheth, 1999; Bishr, 1998). As research continues to resolve data heterogeneities authors have advocated that semantic issues be resolved prior to resolving syntactic and schematic heterogeneities (Kashyap and Sheth, 1996; Bishr, 1998). These approaches have been based on first identifying semantically similar elements and then resolving schematic heterogeneities between semantically similar elements. These approaches have also concentrated on using the object model (Schuurman, 2002), are restricted to the academic domain and are still in the prototyping stage. Although these studies may provide a futuristic solution, organizations still maintain datasets in relational format (Schuurman, 2002). Consequently, organizations continue to face the rigors of data integration. The complexity associated with data integration has further developed reluctance among organizations to achieve interoperability unless incorporated into Off-the-Shelf-Software (Schuurman, 2002). Although these object oriented methodologies of resolving data integration problems may resolve issues that require data from multiple jurisdictions, as of now it is not known when these solutions will be available in Off-the-Shelf-Software. Further, due to the lack of interoperability solutions in a relations environment, data were integrated on an ad hoc basis where the Washington State data was mapped to the BC data.

3.5.1 Schematic Heterogeneities

Schematic heterogeneities or conflicts originate due to the structure and representation of the databases. These heterogeneities pose a great problem while formulating queries and during the actual integration process. This section describes the schematic heterogeneities observed in the two databases, using the classifications forwarded by Kim and Seo (1991).

3.5.1.1 Database Structure

Although the Washington State and the BC databases store lithological information, they differ in their structure. The Washington State lithological database stores data in four tables, whereas the BC lithological database stores information in three tables. Table name conflicts (semantically similar tables are assigned different names or semantically different table are assigned similar names) were also observed in the databases (Table 3.5).

Washington State Database	British Columbia Database
Tblmaterial (geology)	Lithology (geology)
TblWellData (location + general information	Location
TblRecovery	General (general information)
TblWellTest	

Table 3-5 Table name conflicts

3.5.1.2 Table Structure Conflicts

Table structure conflicts arise when the number of attributes in the tables differs. Although the tables TblwellData in the Washington State database and the Location table in the BC database are semantically similar, they store a different number of attributes. TblWellData stores 109 attributes while the UTM table stores 19 attributes. Table conflicts also arise when similar information is stored in different numbers of tables. For example, the table TblWellMaterial in the Washington State database stores information that is distributed among two tables in the BC database (Figure 3.11)

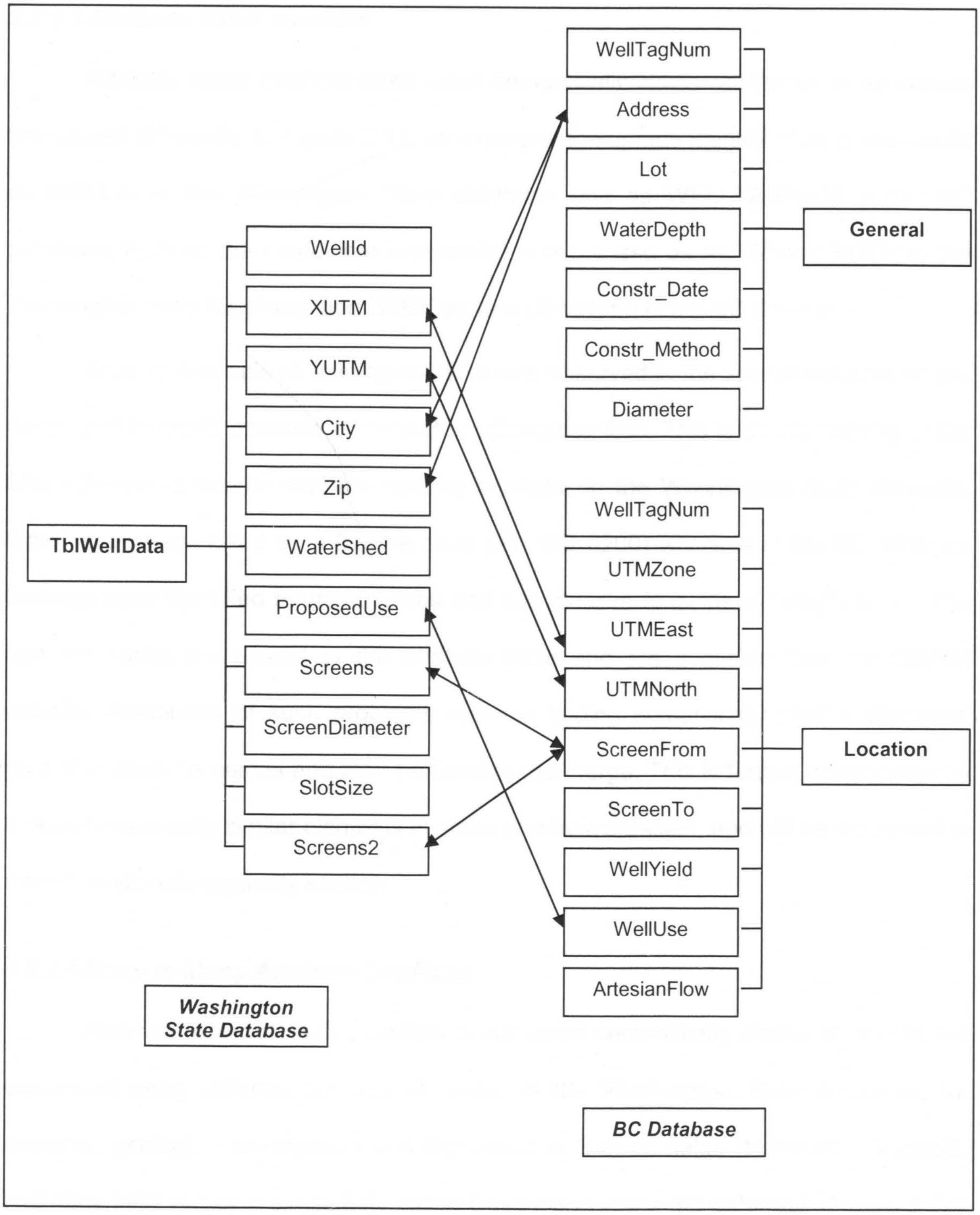


Figure 3-11 Table structure conflicts

3.5.1.3 Attribute name conflicts

Attribute name conflicts arise when semantically similar attributes in databases are named differently. In Figure 3.11, for example, the unique identification is expressed as WELLID in the Washington State database and as WELLTAGNUM in the BC database. Further, the coordinate information is expressed as XUTM and YUTM in the Washington State Database and UTMEast and UTMNorth in the BC Database.

Another example of naming conflicts was observed in the spatial datasets where merging of thematic datasets can result in information loss. This is due to inability of the GIS software to handle attribute naming conflicts. In the Washington State datasets, features were identified by a feature code (ex, 170 0200) whereas in the BC datasets features were identified by their names and sub classes (e.g., paved single lane). The features codes are absent in the attribute table and are available from the MSRM website. Resolution of such problems requires finding semantically similar elements, renaming attribute names and then performing the merge. This is further complicated by lack of semantically similar elements in same thematic datasets, and will be discussed in the semantic heterogeneity section.

3.5.1.4 Many to Many Attribute Conflicts

Many to many attribute conflicts occur when semantically similar attributes are expressed using different number of fields. In the Washington State database, for example, geological descriptions are expressed in 3 three fields (Material1, Material2, and Material3) versus a single field called Description in the BC database (Figure 3.12). This was resolved by concatenating Material1 and Material2 fields. Material3 was not concatenated as only the first two terms were used for reclassifying the geological terms in FSSD (Flexible Standardization for Spatial Data) a non proprietary GUI developed by Dr. Schuurman and Dr. Allen (Simon Fraser University, 2002).

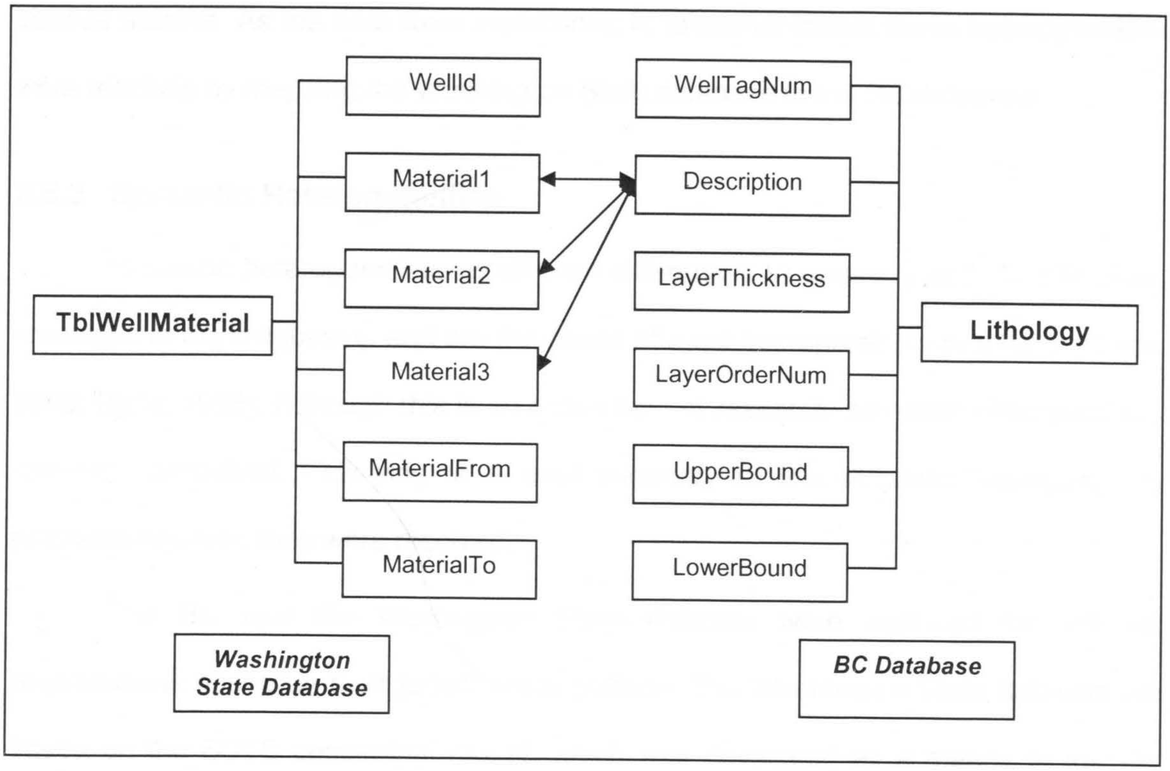


Figure 3-12 Attribute conflicts

3.5.1.5 Data conflict

Data conflicts may arise for any number of reasons. The data conflicts encountered in this research are summarized below:

- Unique identification was represented by a numeric data type in the Washington State database, and as a string type in the BC database.
- There were multiple representations of the same data. Sand may be expressed as sd, sand, snad, while clay may be expressed as cl, or caly.

Although schematic heterogeneities are not difficult to resolve, it is a time consuming process. To date, no general framework has been created for the comprehensive enumeration and systematic classification of resolution techniques for schematic conflicts (Bishr, 1998). Possible resolutions, such as a unified schema (Bishr, 1998) or use of object-oriented data models which include concepts like generalization, aggregation, inheritance and methods (Kim and Seo, 1991) have been suggested by various authors. As the data were maintained in relational format these heterogeneities were resolved by mapping the Washington State database to the BC database

3.5.2 Semantic Heterogeneities

Semantic heterogeneities result from differences in meaning and classifications employed in the databases, and are the cause of most interoperability problems (Sheth, 1999; Bishr, 1998). Although this is an active field of research, semantic interoperability remains unresolved. Following is a brief summary of the semantic interoperability problems and how they were resolved.

The BC and the Washington State datasets were captured for different organizations based on their jurisdictional policies. The Washington State datasets are based on the SDTS conceptual model, which was developed as a means to transfer data between the federal organization and its users. The standard is modular, flexible and conceptually rich in order to accommodate any kind of data model (Arctur, 1998). On the other hand the BC datasets are based on the CCSM classification, which was accepted as a standard for the distribution of digital topographic data among the federal, provincial and private surveying companies in Canada (Evangelatos and Allam, 1991). The different classification schemes employed in the two countries have created semantic interoperability issues. In Canada, for example, hydrographic information is distributed in two separate thematic layers [i.e. Lakes and Rivers (streams, ditches, canals and rivers)], while in Washington State streams, rivers, lakes, ponds and ditches are included in a single hydrographic layer. In Figure 3.13, lakes in Canada are represented in red colour and are captured as a separate thematic layer.

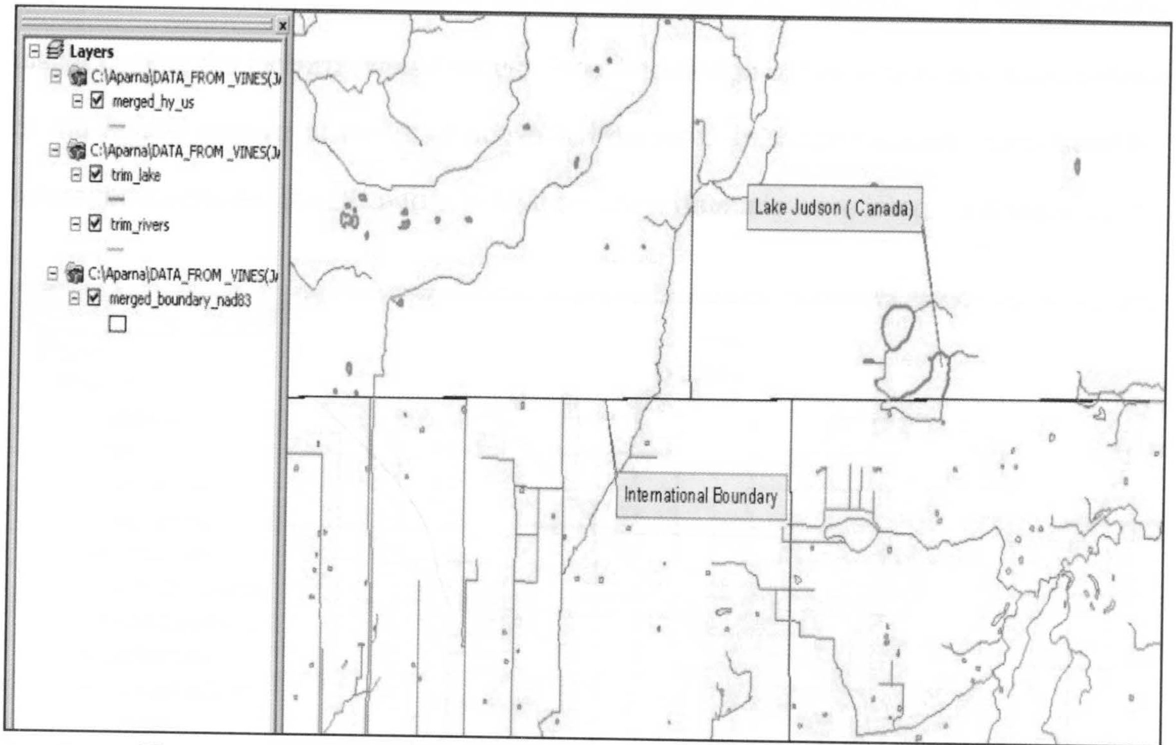


Figure 3-13 Semantic heterogeneities in the hydrographic datasets

Similarly the transportation network in US is captured in three main categories: Roads and Trails; Railways and Pipelines and Transmission Lines. In Canada, however, the BC transportation network is captured as one category, and includes both roads and railways. Transmission Lines, Pipelines and Trials are captured as part of Cultural features. The different classifications have also resulted in the different representation of spatial features, and can further hinder integration and analysis. In the Washington State datasets, for example, major highways are usually represented by a double line, while in BC a single line represents a road unless it is a major highway, where it is represented by a double line irrespective of its classification. This representation of the roads in the BC dataset can cause problems while integrating datasets based on automatic methods. Figure 3.14 shows Highway No.1 which has been classified differently in the study area. On the left (black) it is classified as 'paved 2 lane, two way, undivided), at the centre (red) it is classified as 'paved 2 lane one-way, one-way, undivided) and on the right (blue) it is classified as 'paved 4 lane divided, two way, divided). In this scenario distinguishing the highway would require local knowledge of the area or a visual analysis of the spatial pattern of the road network. However, as these datasets were used for reference purposes, the BC and Washington State datasets were integrated separately.

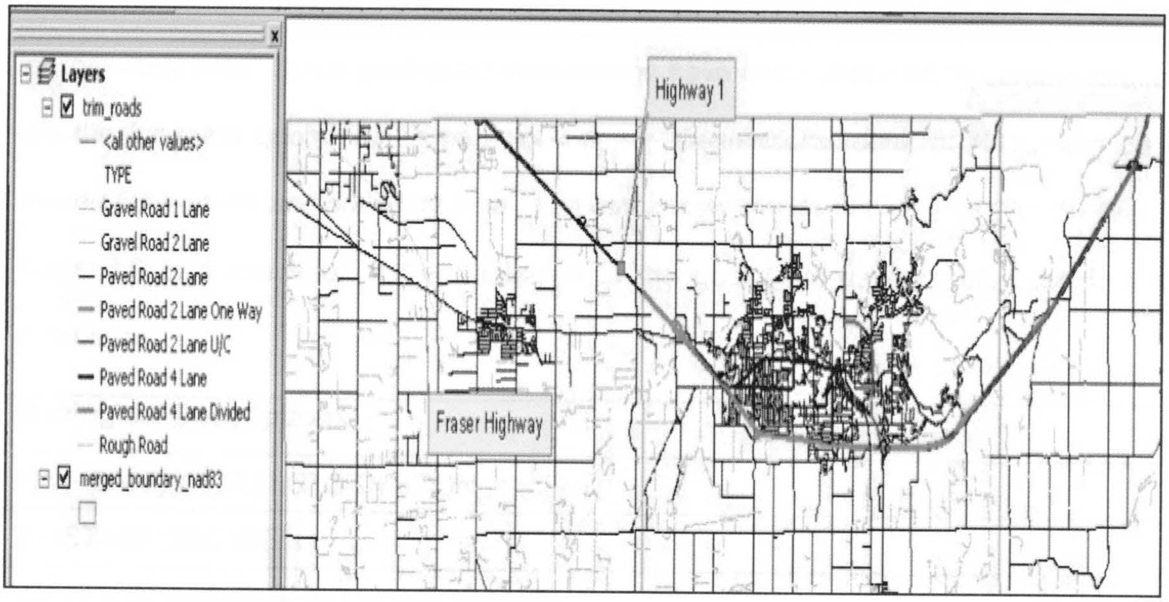


Figure 3-14 BC transportation dataset showing Highway No.1

Another source of semantic heterogeneity was the geological descriptions used in the various geological databases. Referential geological information for the study area was obtained from Ministry of Transportations Bridge construction report, drill core records from a study conducted at Simon Fraser University by Valerie Cameron and Dr. M.C. Roberts, stratigraphic information from Geological Survey of Canada (GSC) reports and well drillers logs. The geological classification for the bridge construction reports were based on the Unified Soil Classification System[31] which is used for engineering

[31] Unified Soil Classification System is a de-facto standard developed by American Society for Testing and Materials (ASTM). ASTM is an international a not-for-profit organization developing consensus standards for material products and services. This standard describes a classification system for minerals and organo-mineral soils for engineering purposes (ASTM, 2003).

purposes and is based on the particle size, liquid limit and plasticity index. The drill core record descriptions were based on the Wentworth Scale[32] and GSC geological descriptions were based on the stratigraphy and environment of deposition. The drillers' descriptions were based on experience or education. These semantic differences and lack of standardized descriptions for the study area resulted in 6000 unique categories for the study area. Some geological descriptions have been observed to contain jargon like the following quote "*Up to here look at Mr. Oesteruicks room*" or "*Whatcom clay moved 50' from #1 and drilled #2 hole at a point chosen by diviner - drilled 2nd dry hole.*" Table 3.6 is a snapshot of the variations in the geological descriptions used in the databases.

Geological Descriptions
FINE SAND (W.B.) LOTS OF IRON
EXISTING DUG WELL
WATER BEARING SAND SOME GRAVEL
SILTY GREY SAND (HARD PACKED)
SILTY GREY W.B. AND SAND PULLED BACK TO 106 FEET

Table 3-6 Geological descriptions recorded in databases

These discrepancies in the terminology may be attributed to:

- the level of education or experience of the driller;
- incorrect spellings cause a new category to be generated by the computer;

[32] The Wentworth scale is a widely accepted international standard for the measurement of sedimentary particles ranging from clay particles (less than 1/256 mm) to boulders (over 256 mm) (Whitten and Brooks, 1985).

- partial terminology used from published GSC maps[33]; and
- the semantics or the geology may be very complex.

Such differences in geological descriptions were standardized using FSSD software. Although a hydrogeological classification developed by Halstead (1986) already exists for the GVRD this classification was not used as the categories are difficult to distinguish without laboratory analysis. The FSSD program allows standardization using a flexible approach using any one of the three different classification schemes: Geological Survey of Canada (GSC), BC and National rules. For this research geological descriptions were standardized using the 'BC Rules' developed by Dr. Allen and Dr. Schuurman. This classification selects the first two geological descriptors and interprets the grammar of terms to assign priority to the descriptor, with the first term being weighted more heavily than the second term (Schuurman, 2002). The standardization using the BC Rules resulted in 66 categories (see Appendix A). Certain discrepancies were observed where geological formation names (such as Newton, Whatcom, and Surrey) used in GSC surficial geology maps was used to describe the geology. Terms like Newton or N.S.C have been used to describe stony clay deposits, whereas Whatcom or Surrey has been used to describe till layers. To maintain the originality of the descriptions such occurrences were manually corrected, resulting in 116 categories. These categories were further reduced using a rule-based approach (a detailed list of the rule-based approach is presented in Appendix A). For example,

---

[33] For example, Whatcom and Surrey which are described as till formations were used instead of 'TILL' or Newton or N.S.C. were used to describe 'STONEY CLAY'

IF

Geology = overburden + material or material + overburden (material could be sand or gravel or clay)

AND

Layer order = 1

THEN

Geology = Overburden

ELSE

Geology = material

This rule-based classification reduced the dataset to thirty one categories. These categories were further reduced to thirteen categories (see Appendix A, Table A-6). This dataset could then be imported into Groundwater Modelling Software (GMS v 4.0) for the creation of cross-sections and the development of the hydrostratigraphic model[34] and will be discussed in the following chapter.

3.6 Summary

Groundwater is a natural resource that does not respect jurisdictional boundaries. This chapter has explored the challenges of integrating datasets for a groundwater management project in an international setting. Research and maintenance activity with respect to these data are expensive and usually call for a collaborative approach (Rivera, 2003). In Canada, The Oak Ridges Moraine project in Ontario and the Brookswood Aquifer project in BC were launched in response to a lack of research at the regional scale. Both projects resulted in a 3D hydrostratigraphic models and involved geophysical field surveys to assist in the development of the hydrostratigraphic model.

---

[34] A Hydrostratigraphic model defines the hydrostratigraphic units of the study area. Hydrostratigraphic units are defined based on the hydraulic properties (porosity and permeability) of subsurface layers. Clay, for example, has a high porosity but low permeability, and hence, cannot transmit water and is considered as an aquitard (unit that does not transmit water). In contrast, a sand or gravel unit with high permeability is considered an aquifer (a unit that transmits water readily).

The Oak Ridges Moraine project has provided a framework for groundwater management activities in the Toronto Greater area. Although a 3-D hydrostratigraphic model was developed for the Brookswood aquifer it did not result in a standardized groundwater database for the region, and similar studies were not extended to other aquifers in the region.

To achieve the first objective of this thesis, which is to study the interoperability challenges in an international setting, groundwater related data from Canada and the US were integrated. Due to the lack of interoperability protocols data were integrated on an ad hoc basis, in which the US dataset schema was mapped to the BC schema.

Interoperability should be considered as a starting point for all cross-border projects. Interoperability research has concentrated on methods to resolve either technical or institutional interoperability issues and have rarely addressed all the factors or challenges of integrating datasets. Data integration for cross-border projects, however, calls for an investigation of technical as well as organizational aspects of interoperability. Following is a discussion of the technical and institutional challenges of integrating cross-border datasets.

Technical interoperability problems result due to the inherent nature of spatial data. Data are usually in an unusable format and their conversion is labour intensive. For example, in this research the water well information distributed by MWLAP on their website[35] is in a format that cannot be directly used for analysis. As technology is advancing, new methodologies are being developed for distributing information. The water well information from MWLAP and the Department of Ecology are now being

---

[35] Groundwater related data be downloaded directly (in bulk) from the Ministry's website or from the internet mapping application, in which case data can be downloaded in a PDF format one at a time.

distributed by Internet Mapping application developed by the organizations. Although the current trend in GIS is to develop internet mapping application to facilitate data download, the water well information can still only be downloaded only in a PDF format. In today's digital world downloading well log information in PDF format does not resolve the issue of distributing information digitally. Additional manual data entry is still required to convert these data into a digital format. Conversion of these PDF files will result in multiple databases, further augmenting the data integration in the event of value addition.

Although these datasets are maintained to provide groundwater information to the public, consultants and drillers, the information is rarely used due to its extremely poor quality. Data quality is one of the many reasons that affect data integration. Although a source of invaluable geological information, data quality issues have compromised the use of these lithological databases. Several data inconsistencies were observed which required repeated data quality checks. Some of the data errors can be attributed to data entry by the Ministry staff.[36] Changes to the database structure can significantly reduce data errors and improve the quality of the data. For example, creating a database adhering to database rules would resolve most database errors. Setting data entry integrity rules[37] would resolve the issue of inconsistent records in the three tables (section 3.3.4.2) or the issue of missing geological descriptions (section 3.3.4.4). Increasing the size of the 'description' field would resolve the issue of creating a new record for storing geological descriptions, if the description exceeded the 30 character limit.

---

[36] Discussed in section 3.3.4.1 (CGDS Design).

[37] Setting integrity rules that prevent data entry in certain tables only or setting integrity rules that prevent blank entries in the 'Description' field.

The most significant of these data quality issues was the lack of locational coordinates: the basis for discontinuing any GIS project. Other data quality issues included the absence of surface elevation for well records, a lack of layer orders in the US database, a lack of geological information, an inconsistent number well records in the various tables, and inconsistencies in the geological descriptions which resulted in 6000 unique descriptions for the study area. The geological descriptions in the BC databases also included important information regarding the specific yield and flow rates. Although this information is not important for geological descriptions is important from a groundwater modelling perspective.

Schematic and semantic heterogeneities, which have been identified as the most difficult to resolve, were also sources of interoperability problems. Schematic heterogeneities, such as differing attribute names and table name conflicts were observed. Such heterogeneities were resolved by mapping the US schema to the BC schema. Semantic heterogeneities, which develop due to differences in meaning of data, were the most difficult to resolve. Semantic heterogeneities in the spatial datasets were resolved on an ad hoc basis. As the spatial datasets were used for reference purposes only, the BC and Washington State datasets were merged separately. The semantic differences in the geological descriptions were resolved by using the FSSD software.

Apart from data issues, institutional factors also affected the data integration process. The issues of data dissemination policies, copyright issues and data costs have often been cited as the facilitators or detractors of a booming GIS industry. Although research has shown that no profits are generated by adopting a cost recovery policy (Sears, 2001), Canada continues to charge a price for its datasets. Academic inquiry into such issues in Canada has also shown that such policies have stifled the GIS industry (Klinkenberg, 2003; Sears, 2001). Although research has addressed these issues from a policy perspective the effects of metadata on data discovery, access and dissemination are rarely addressed.

The use of spatial datasets in Canada is constrained by lack of metadata. Metadata, defined as data about data, is not only useful for determining the fitness for use, but also forms the basis for data catalogues and data clearinghouses. Absence of metadata means that the user is unaware of the existence of the dataset but, when available, is unaware of the semantics of the dataset. Recent initiatives in Canada, such as the development of the Canadian Geospatial Data Infrastructure has paved the way for enhancing the geospatial community by providing the necessary framework for sharing information. The Data Discovery Portal of GeoConnection helps users find the necessary dataset and services available in Canada. Organizations can thus advertise their services and provide metadata for their products. Although such efforts are commendable and organizations have advertised their products they have been slow in submitting metadata information in FGDC compliant format. As of today there are a total of 523 organizations that have submitted information to the Discovery Portal, out of which only two provincial ministries have submitted information regarding their services and products. Although local government organizations maintain spatial databases, only seven municipalities in the country have submitted information to the Discovery Portal. Lack of metadata mandates for organizations in Canada has also resulted in metadata in the form of text documents, self explanatory or by word of mouth (Regier, 2002). On the contrary in the US most organizations have advertised their products and services via NSDI using the FGDC metadata standard. Following the success of national spatial data clearinghouse many organizations are developing their own data clearinghouses, thus making their data accessible to a wider user community. The success of the NSDI, the wide spread availability of spatial data and associated metadata can be attributed to political backing, which mandated federal organizations and urged organizations to document their spatial datasets using FGDC metadata standards and make them electronically available.

Lack of such legislations in Canada has also resulted in organizational reluctance to adhere to standards. Although SAIF is an indigenously developed Canadian standard very few organizations distribute their data in SAIF format. In Canada, adherence to spatial data standards depends on jurisdictional constraints (Quekenbash, 2002). The same is observed with metadata where organizations distribute metadata in any form they deem appropriate, which may be a text file, self explanatory or by word of mouth.

Institutional inertia and reluctance to adhere to standards can also directly affect the integration process. Lack of awareness or coordination of activities within the same department or level of governance can result in duplicated efforts. This was observed for the Washington State, lithological databases, which were are maintained by the Department of Ecology and USGS and which lack cross referencing.

Thus, interoperability research calls for an integrated approach that investigates the technical and institutional aspects of interoperability. Anthropogenic activities are stressing our environment and the increasing incidences of the adverse effects of climate change and natural hazards are taking a toll on humans and property. These effects do not follow political boundaries and there a pressing need to develop mitigation strategies and develop models to predict future events. The availability of integrated datasets, however, is the foundation of these activities. Data integration, an issue for a single jurisdiction, is further complicated for cross-border projects and there is a necessity to understand the challenges for integrating datasets across borders so that mitigation strategies can be developed for the benefit of mankind.

4 THE HYDROSTRATIGRAPHIC MODEL

4.1 Introduction

This chapter discusses the usability of the lithological database in building a hydrostratigraphic model for the Abbotsford-Sumas aquifer. Studies have revealed that the use of water well data to build hydrostratigraphic models is constrained, in part due to inconsistent geological descriptions (Russell at al., 1998). In the absence of geophysical studies, however, water well data may be the only source of subsurface geological information (apart from interpretations based on geologic mapping). Following a brief discussion of past geological investigations, this chapter will discuss the geological setting and describe the hydrostratigraphic model[38] of the study area.

4.2 Past Research

The Fraser-Whatcom lowland[39], in which the Abbotsford-Sumas aquifer is located, is host to numerous aquifers in the region. This region is composed predominantly of material from glaciations in the Quaternary Period[40]. Although geologic

---

[38] A Hydrostratigraphic model defines the hydrostratigraphic units of the study area. Hydrostratigraphic units are defined based on the hydraulic properties (porosity and permeability) of subsurface layers. Clay, for example, has a high porosity but low permeability, and hence, cannot transmit water and is considered as an aquitard (unit that does not transmit water). In contrast, a sand or a gravel unit with high permeability is considered an aquifer (a unit that transmits water readily).

[39] The Fraser-Whatcom Lowland is a triangular shaped area consisting of Quaternary deposits. It is bounded on the north by Coast Mountains, on the southeast by Cascade Mountains and on the west by the Strait of Georgia (Armstrong, 1981).

[40] Quaternary Period: The Quaternary Period ranging from 2 million year ago to present experienced extensive glaciation. The various stages of glaciation includes (from the youngest to the oldest): Fraser Glaciation (20ka - 10ka); Olympia Interglaciation (60ka - 20ka); Possession

investigations, to decipher the stratigraphy and surficial geology have been conducted on both sides of the Canada / US border, these studies often explore a large area, thus prohibiting the development of a detailed model. Subsurface geology, in these studies, is based on the lithology, depositional environment and chronologic history of the area. Further, the majority of studies do not investigate subsurface geology from a hydrostratigraphic perspective.

Although research from a hydrostratigraphic perspective has been conducted in both the Canadian and US portions of the study region, it varies in objective and study area. Research by Cox and Kahle (1999), for example, covered parts of Canada and US, without encompassing the entire study area. The most comprehensive research to date was conducted by Halstead (1986), and exists only for the Canadian portion of the Lowland. His study area extends from the Fraser Delta (west) to the Sumas Mountain (east) and from the international border (south) to the Fraser River (north). Halstead has divided the region into six hydrostratigraphic units based on the similarity of sediments and the environment of deposition (see Appendix C). The term hydrostratigraphic unit, however, has been incorrectly used in the article. Although Halstead's classification has been used by drillers to identify lithology (Symington, 2002; Livingstone, 2002), it is difficult to distinguish between hydrostratigraphic units A and B, which can be identified by the stone, clay content and the location (see Appendix C). Further, the research was conducted at a time when computer technology was not well-developed. Consequently, the hydrostratigraphic model exists in paper format and cannot be directly imported into groundwater modelling software.

Glaciation (80ka – 60ka); Whidbey Interglaciaition (100ka – 80 ka) and Double Buff Glaciation (> 100ka) (Jones, 1999).

Cameron (1989) explored the geologic history and provenance of the Sumas Valley sediments based on drill core and lithologs available from the Ministry of Water, Land and Air Protection (MWLAP) and from Washington State. Although similar hydrostratigraphic studies have concentrated on hydrogeology, the main objective has been to study water quality, flow rates and nitrate pollution levels in parts of the aquifer. Kahle (1990), for example, explored hydrostratigraphic and groundwater flow for a ten square mile area west of Sumas, Washington State. Although this study explored the hydrogeology it included a small portion of the aquifer south of the border. Later investigations showed the study to have provided incorrect interpretation of a clay unit (Stasney, 2000). Stasney's (2000) study investigated the causes of elevated nitrate levels in a localized two square mile area. A more comprehensive study, which explored the hydrostratigraphy of the entire aquifer, is provided by Cox and Kahle (1999). This study encompassed a 220 square mile area in Whatcom County and a portion in the Abbotsford-Sumas area. Although the study covers a major part of the aquifer the classification is based on surficial geologic units and lacks detailed classification. Further, the data cannot be imported for groundwater modelling as it is not in a digital format. Although the above studies provided information on parts of the aquifer they lacked hydrogeologic analysis for the entire aquifer. The studies also lacked data for the entire aquifer that could be imported or integrated with available groundwater modelling and GIS software.

4.3 Geological Setting

The study area has been subjected to extensive glaciations during the Quaternary Period (the most recent glaciations occurred during the Pleistocene), which has resulted in complex geological architecture (Figure 4.1). Glacial advances, followed by de-glaciations, deposited a thick sequence of sediments of diverse origins

(Armstrong, 1981). The thickness of these glacial deposits is so great that the bedrock is believed to be 1000-2000 feet below the Pleistocene deposits throughout the Fraser-Whatcom lowland (Cox and Kahle, 1999). This vast thickness of sediments means that much is known about the deposits from the most recent glaciation, the Fraser Glaciation, which deposited the Vashon Drift, the Fort Langley Formation / Everson Glaciomarine drift[41], the Capilano sediments and the Sumas drift.

Geologic Framework

Period	Age Ka	Geologic/Climate Units[1]		Stratigraphy	
Quaternary		Interglacial		alluvium marine deposits[2]	
	10	Fraser Glaciation	Sumas Stade	Sumas Drift	
	12		Everson Interstade	Everson 'glaciaomarine' Drift[3]	Bellingham Drift; Deming Sand; Kulshan Drift
				Fort Langley Formation[4]	
			Vashon Stade	recessional outwash	
				Vashon Till	
				advance outwash	Quadra Sand[5]
				Esperance Sand Member	
	20		Evans Creek Stade[6]		
		Olympia Interglaciaition		Quadra Sand	Cowichan Head Formation[7]
	60	Possession Glaciation		Possession Drift-----	Semiamoo Drift[8]
	80	Whidbey Interglaciation		Whidbey Formation	
	>100	Double Buff Glaciation		Double Buff Drift	
pre-Quaternary rock units					

Figure 4-1 Fraser-Whatcom Basin stratigraphic units and corresponding glaciation stages ***(adapted with permission from Jones, 1999)***[42]

[41] Fort Langley Formation is identified as Everson glaciomarine drift in the US.

[42] ([1]:Drift sequences are generally separated by unconformities; [2]: Marine sediments are considered part of a aquifer system where saturated with freshwater; [3]: Also includes glaciofluvial

The study area is dominantly composed of the Fort Langley Formation / Everson Glaciomarine Drift and the Sumas drift. The Fort Langley Formation was deposited during the Everson interstade, and consists primarily of a thick succession of interbedded marine and glaciomarine sediments of clays, silty clay, stony and silty clays (Armstrong, 1981; 1984). This formation is exposed in the western part of the study area, and is overlain by Sumas Drift (Armstrong, 1984).

The Sumas drift was deposited during the Sumas Stade, which represents the final stage of glaciation in the area, and occupied a small portion of the eastern Fraser lowland, extending a few kilometres south of the border (Armstrong, 1984). The Sumas Drift consists of outwash sand and gravels which comprise the Abbotsford-Sumas aquifer, till and glacio-lacustrine sediments (Armstrong, 1981). The Sumas Drift has been radiocarbon dated to approximately 11.5 to 11.1 ka BP (Clague, 1994). The Sumas Drift occupies approximately the central part of the study area, and underlies the Salish sediments in the eastern portion of the study area (Armstrong, 1981). The Salish sediments are observed only in the Sumas Valley, have been deposited by rivers, and consist of clay, sand, silt and gravel (Cameron, 1989).

4.4 Hydrostratigraphic Model

The integrated database (discussed in chapter 3) created as part of this research for Canada and US was used to construct the hydrostratigraphic model. In addition to

---

sediments-Everson Sand (early Everson) and Everson gravel (late Everson) [4]: Canadian name for Everson Glaciomarine Drift; [5]: Canadian name for Vashon deposits older than till, although in many locations the unit does not include advance outwash; [6]:Deposists of similar age and older than Evans Creek Stade generally not exposed in the basin, inferred from well-log information and from some exposures in Canada; [7]: Canadian name for Olympia Interglaciation deposits; [8]: Canadian name for pre-Olympia Integlaciation deposits) the water well data, data from Geological Survey of Canada reports, drill core records from a study conducted at Simon Fraser University and the Ministry of Transportation (Canada) bridge construction reports were also used. These records are available only in paper format and were subsequently converted to digital format by manually encoding the data.

These data were then reclassified using the rules employed for classifying the driller's records. The geologic material classifications used in the above mentioned reports were, however, based on different paradigms: for example, the Geological Survey of Canada reports and the drill core reports used the Wentworth scale[43] for measuring the size of the sediments, the Transportation Ministry used the Unified Soil Classification System (USCS)[44] (see Appendix C) and drillers used experience for identifying the sediments. Sediments have different hydraulic properties and merging databases based on different classifications can have an impact on the accuracy of the models generated. The geological term 'cobbles' in USCS, for example, is equivalent to 'boulders' in the Wentworth scale. Similarly, clay and silts are grouped as a single category (fines) in the USCS classification system (see Appendix C). Despite these differences, the data were reclassified using the rules employed for classifying the driller's records. These data were then imported into Groundwater Modelling Software (GMS v 4.0) for constructing cross-sections and developing the hydrostratigraphic model.

---

[43] Wentworth Scale: The Wentworth scale is a widely accepted international standard for the measurement of sedimentary particles ranging from clay particles (less than 1/256 mm) to boulders (over 256 mm) (Whitten and Brooks, 1985).

[44] The Unified Soil Classification System: This standard describes a classification system for minerals and organo-mineral soils for engineering purposes (ASTM, 2003).

Building cross-sections is crucial for any geological study. Although the manual construction of cross-sections is a tedious task, Off-the-Shelf Software are available that partially automate this process. The automatic generation of cross-sections helps speed up the interpretation and analysis of the geological architecture of an area. Today, commercial software packages are also available; these are either loosely or tightly coupled with GIS software. The software packages, however, are subject to interoperability problems based on different data structures and models; a problem that is further aggravated by the lack of standards for the earth science domain (Houlding, 2002). This results in syntactic, schematic or semantic interoperability issues, similar to those observed for geographic information (discussed in chapter 3).

Although many software packages enable the generation of cross-sections, interpretation of layers requires *a priori* knowledge of the geological architecture. Such software packages work efficiently when the user is familiar with the geological architecture, when good quality data are available, or when geophysical studies assist subsurface interpretation. Such software packages, however, pose a problem for complex geological areas where extremely poor quality data are available. This was the case for this study, where lithologic units in adjacent wells do not co-relate. Further, such software lack visualization capabilities that make interpretations even more problematic. In addition, to the above mentioned issues, software cannot handle lenses[45]. Lenses can be important to incorporate in a hydrostratigraphic model, depending on the scale of interest. At a local scale, they result in localized groundwater conditions and need to be identified and modelled in analysis. Despite these disadvantages, however, GMS

---

[45] A lens is a laterally elongated unit that is lithologically distinct from the surrounding geological materials (Allen, 2004).

software facilitates the development of the hydrostratigraphic model, which can subsequently be used for modelling groundwater.

In this research, a lack of visualization capabilities and correlation between adjacent lithologs necessitated that the dataset first be further categorized into seven basic material types (Sand & Gravel, Clay & Silt, Till, Topsoil, Bedrock, Unknown, and Organic) (see Appendix A, Table A-7). The wells for constructing the hydrostratigraphic model were then selected as a function of their distribution and depth. The depth of the Canadian wells ranged from 4ft. to 852 ft. These wells were classified into six categories (less than 50 ft., 50-100 ft., 100-150 ft., 150-200 ft., 200-400 ft. and greater than 400 ft. and above (Appendix A)). The US wells ranged from 2ft. to 371ft and were classified into 3 depth categories (less than 50 ft., 50-100 ft. and 100-400 ft. (Appendix A)). The wells for cross-section generation were selected by performing a visual analysis of the spatial distribution of the wells, categorized by depths, in ArcGIS® v 8.3. Spatial distribution of the deep wells was first observed and then cross-section lines were generated so as to incorporate the deepest wells. Wells in regions of low distribution were then selected from the next categories. This process ensured that the deepest wells were selected in the cross-section generation.

Due to the lack of visualization capabilities a 3-D physical model was constructed to interpret the hydrostratigraphy of the study area (figure 4.2 and 4.3). A total of twenty eight cross-sections were represented in the model (figure 4.4, and Appendix D). Where lithological layers could not be identified or where anomalous layers were present, lithologs from surrounding areas were verified. The lithological layers identified were the basis for constructing the hydrostratigraphic layers in GMS. This dataset was then imported into GMS v 4.0 to produce the final output for the model (Figures 4.5, 4.6, 4.7, and 4.8 and see Appendix D).

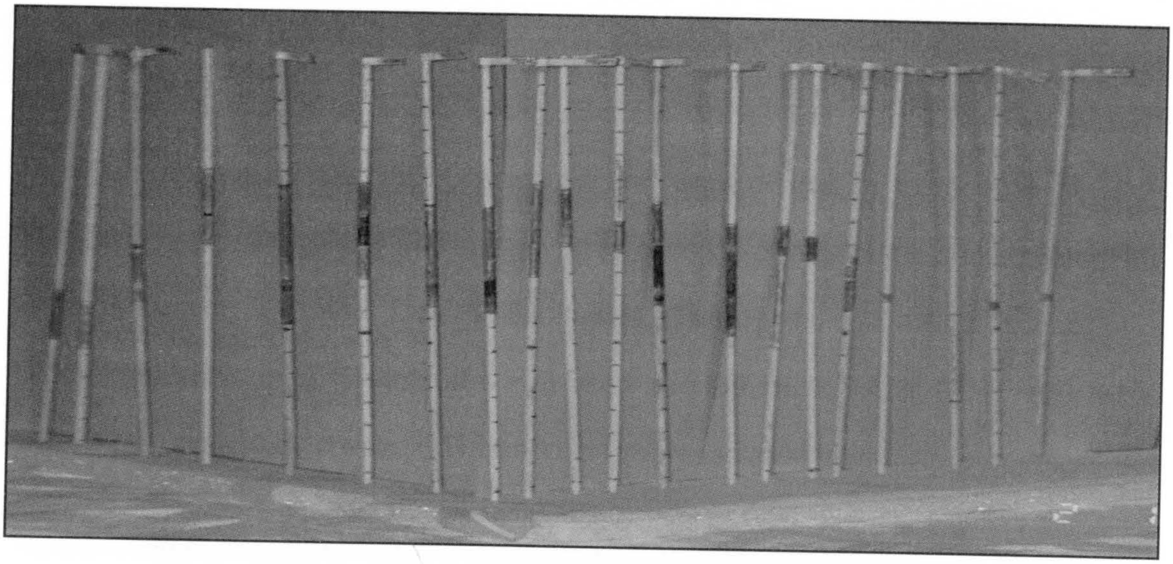


Figure 4-2 3-D model of the study area (side view along a single cross-section)

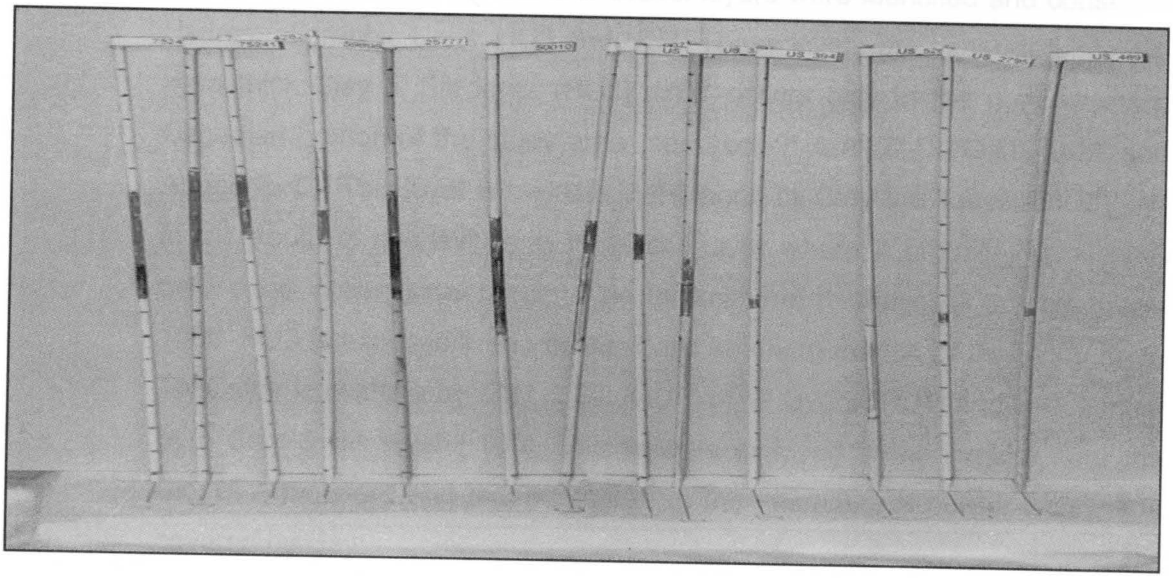
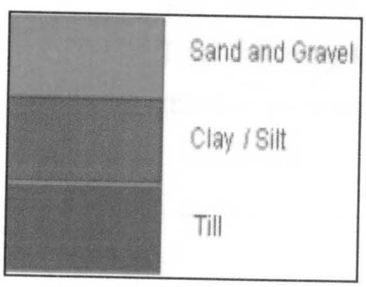


Figure 4-3 3-D model of the study area (front view)



Using the simplified classification (see Appendix A) three dominant lithology types were identified: Clay & Silt (green), Sand & Gravel (red), and Till layers (brown). Till layers appear as discontinuous lenses and were difficult to extrapolate between wells. The hydrostratigraphy of the study area consists of alternating sequences of the Sand & Gravel, Till, and Clay & Silt layers. These lithologic layers constitute 9 hydrostratigraphic units[46], which are classified as HU1, HU2, HU3, HU4, HU5, HU6, HU7, HU8, and HU9 (Table 4.1) based on representative hydraulic properties(see Appendix C) of the material. The three lithologic units are described below:

- **Sand & Gravel lithologic unit:** This unit consists of sand and gravel, sand and gravel. Two discrete Sand & Gravel layers were identified and constitute Hydrostratigraphic Units: HU3 and HU7. These layers are separated laterally by a thick Clay & Silt layer (HU5). HU7 occurs only in the north western Canadian portion of the study area (see section A-A', D-D', O-O', M-M' and Appendix D). This layer is overlain in the north by Clay & Silt layer (HU5) and in the south is overlain by a till layer (HU6) where it probably terminates before the international border. The thickness of this layer is approximately 100ft. HU3 is exposed in the eastern and southern portion of the study area. This layer is overlain by Clay & Silt (HU2) or till layers (HU1) and is underlain by a Clay & Silt layer (HU5). This layer is exposed at the surface near the City of Abbotsford and extends south to the international border where it is covered by clay.

- **Clay & Silt lithologic unit:** This unit consists of clay, silt and clay, stony clay, silty clays and constitutes Hydrostratigraphic Units (HU2, HU5, and HU8). This unit may form as the confining or semi confining unit based on its low relative permeability. A thick unit of Clay & Silt layer separates the two

---

[46] The hydrostratigraphic units are interpreted based on the relative depths as observed in the cross-sections. Some hydrostratigraphic units may be of the same age as ones above or below. Given the complexity of the study area and the uncertainty associated with the dataset, age relationships could not be determined.

discrete Sand & Gravel layers (HU3 and HU7). In some logs the clay extends unto 170ft. below sea level.

- **Till lithologic unit:** This unit consist of till, and till and sand/silt/clay. Although Halstead (1986) has identified continuous till layers in certain areas, the continuity of the till layers could not be verified. A continuous till layer (HU6) overlying hydrostratigraphic unit HU7 was observed in the western Canadian portion of the study area. This till layer, however, terminates near the international border (see section D-D', E-E', O-O'). Very few till occurrences are observed in the US portion of the study area.

Hydrostratigraphic Units	**Lithologic Unit**	**Halstead's Units**
HU1	Till	Hydrostratigraphic Unit D
HU2	Clay & Silt	Hydrostratigraphic unit A / B / E
HU3	Sand & Gravel	Hydrostratigraphic unit C
HU4	Till	Hydrostratigraphic Unit D
HU5	Clay & Silt	Hydrostratigraphic unit A / B / E
HU6	Till	Hydrostratigraphic Unit D
HU7	Sand & Gravel	Hydrostratigraphic unit C
HU8	Clay & Silt	Hydrostratigraphic unit A / B / E
HU9	Bedrock	Hydrostratigraphic unit F

Table 4-1 Generalized Hydrostratigraphic Model[47]

4.5 GMS Output

Two GMS outputs were generated for each cross-section: the generalized hydrostratigraphic model (Figures 4.5, 4.6, 4.7, and 4.8 and see Appendix D) was generated using the simplified classification (see Appendix A, Table A-7); whereas the lithologic units (Figure 4.9 and Appendix D) were generated using the final classification

[47] Table 4.1 is based on my interpretation of the layers based on relative depths as observed in the cross-sections. Given the complexity of the study area and the uncertainty associated with the dataset, a hydrostratigraphic unit of the same age may be observed higher of lower in the hydrostratigraphic column.

(see Appendix A). Figure 4.4 shows the cross-sectional plan for cross-section A-A', D-D', M-M' and O-O'.

The final hydrostratigraphic model is consistent with Halstead (1986) Hydrogeologic Fence diagrams except in areas where lateral continuity of the units could not be identified. The cross-sections are also fairly consistent with the cross-sections generated by of Cox and Kahle (1999) (see Appendix C for a comparison of the hydrostratigraphic units).

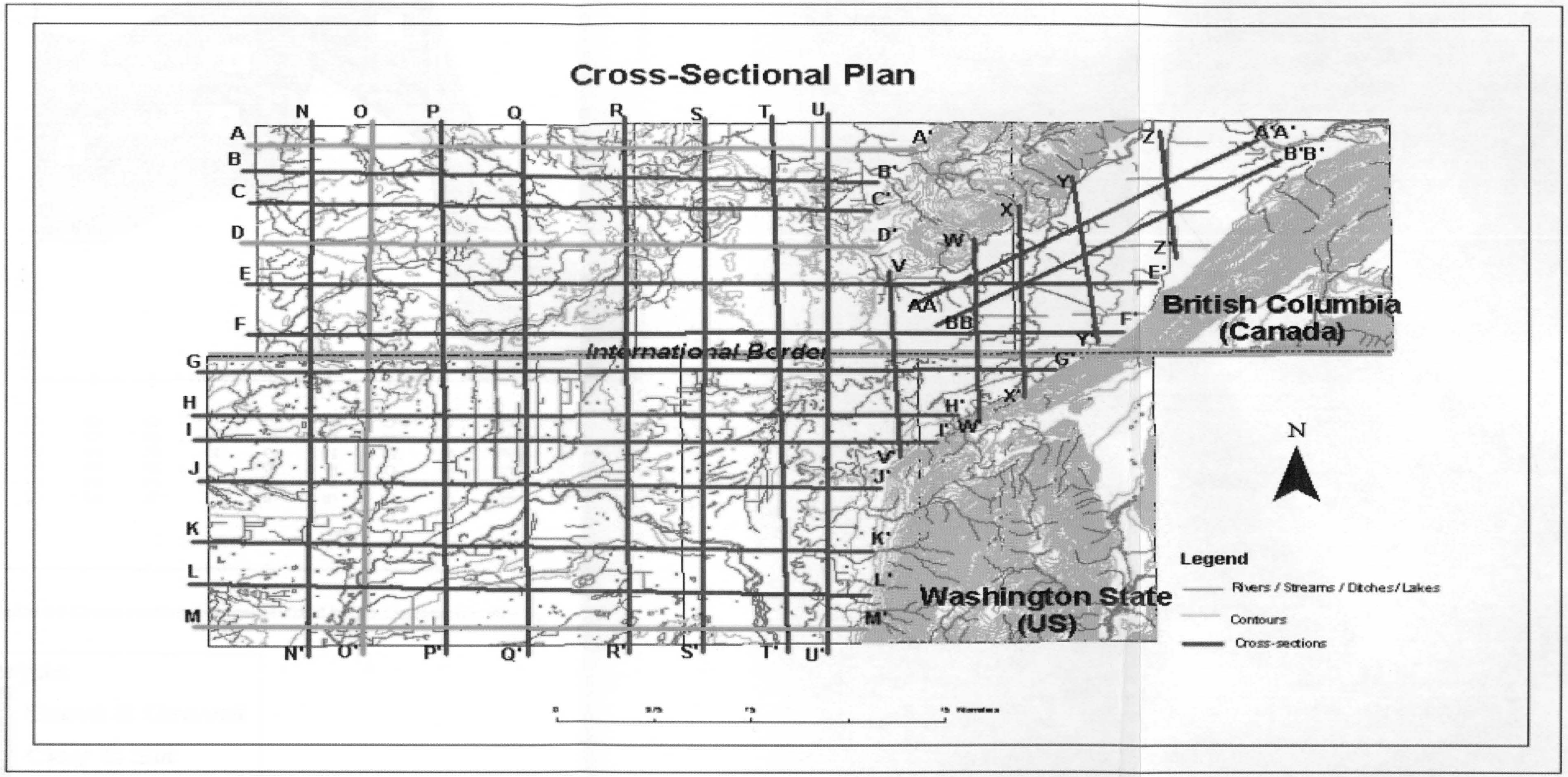


Figure 4-4 Cross-sectional plan

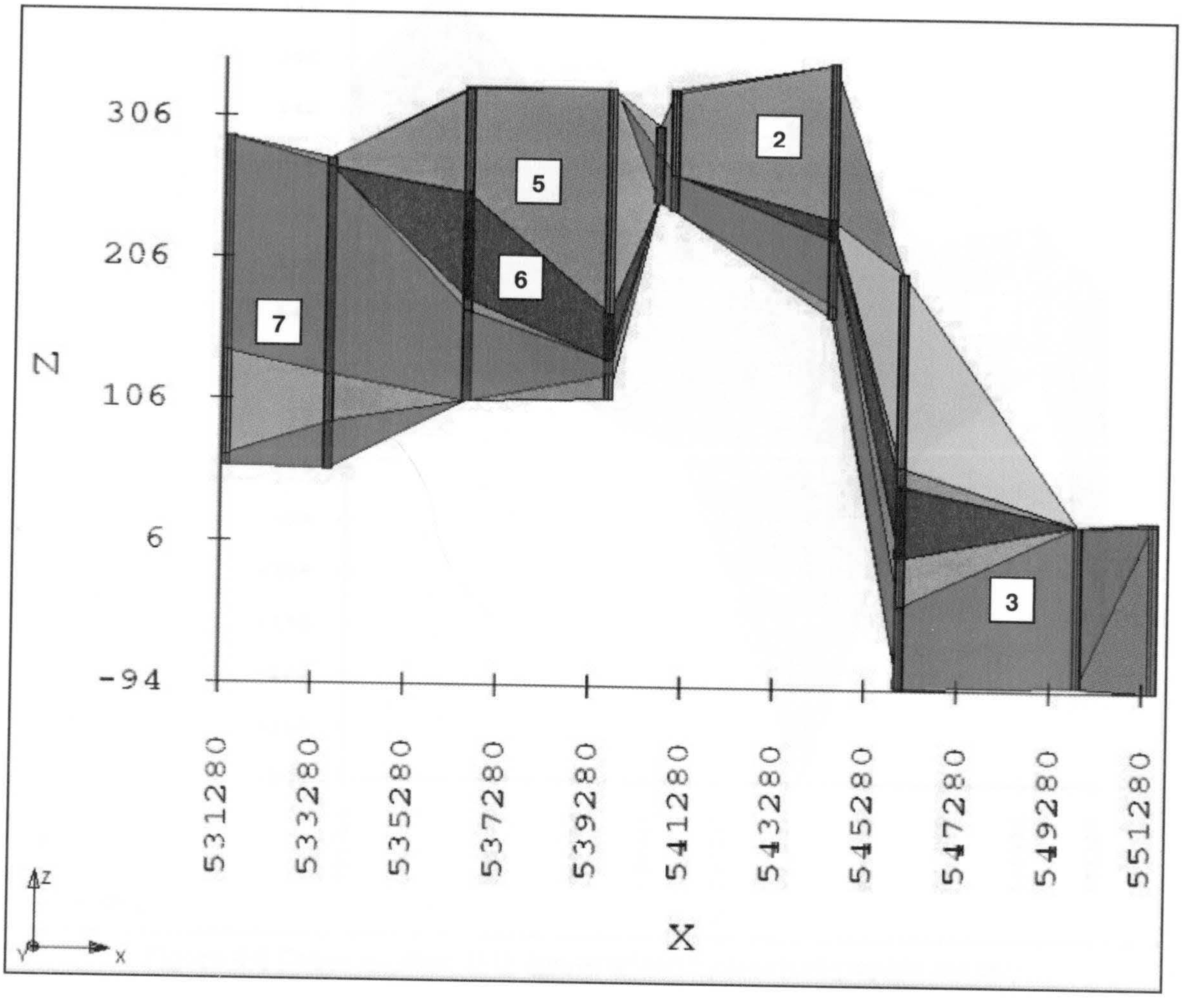
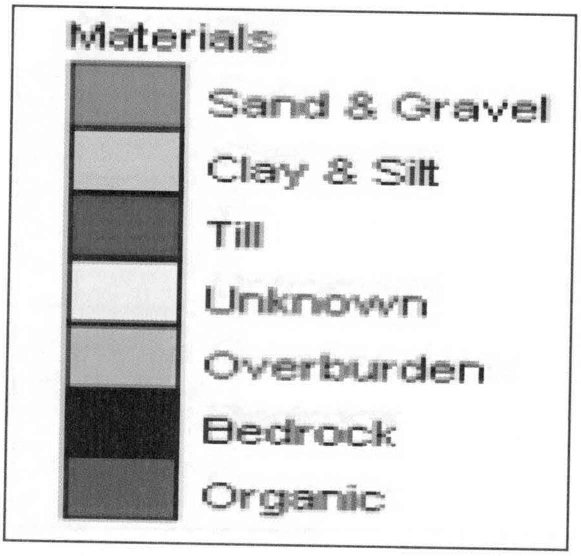


Figure 4-5 Cross-section A-A'(generalized hydrostratigraphic model)



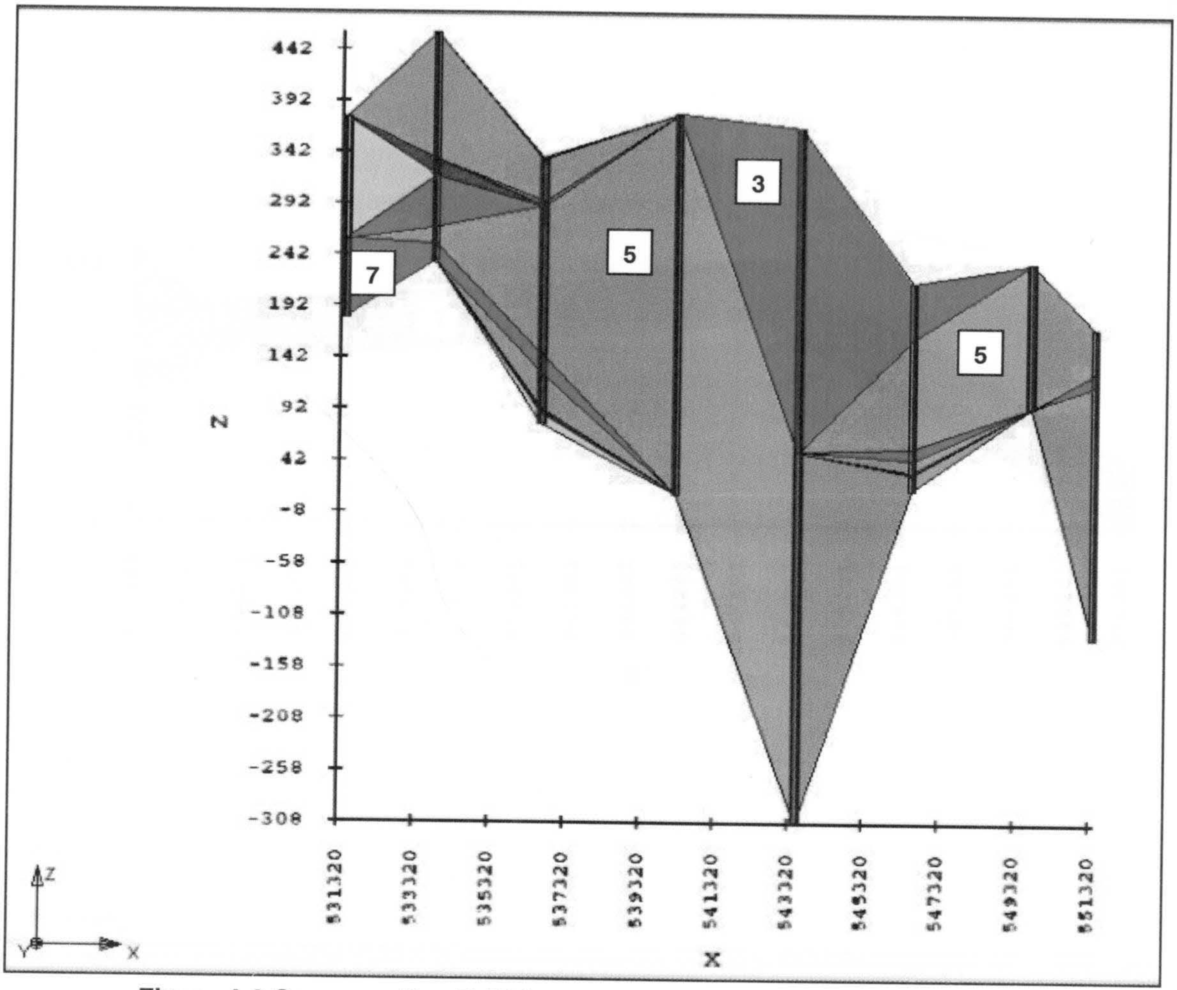
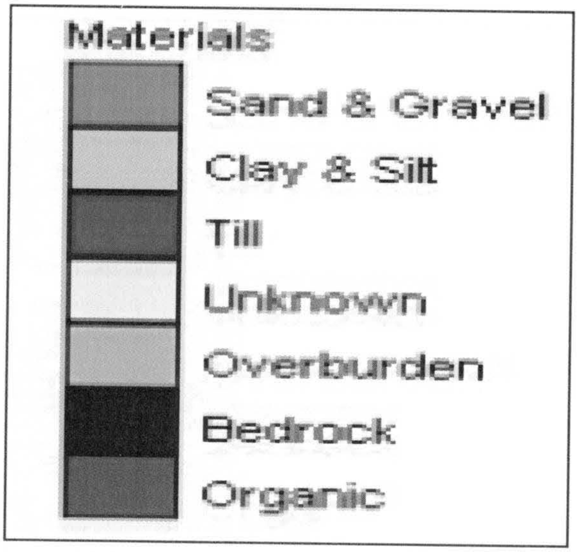


Figure 4-6 Cross-section D-D' (generalized hydrostratigraphic model)



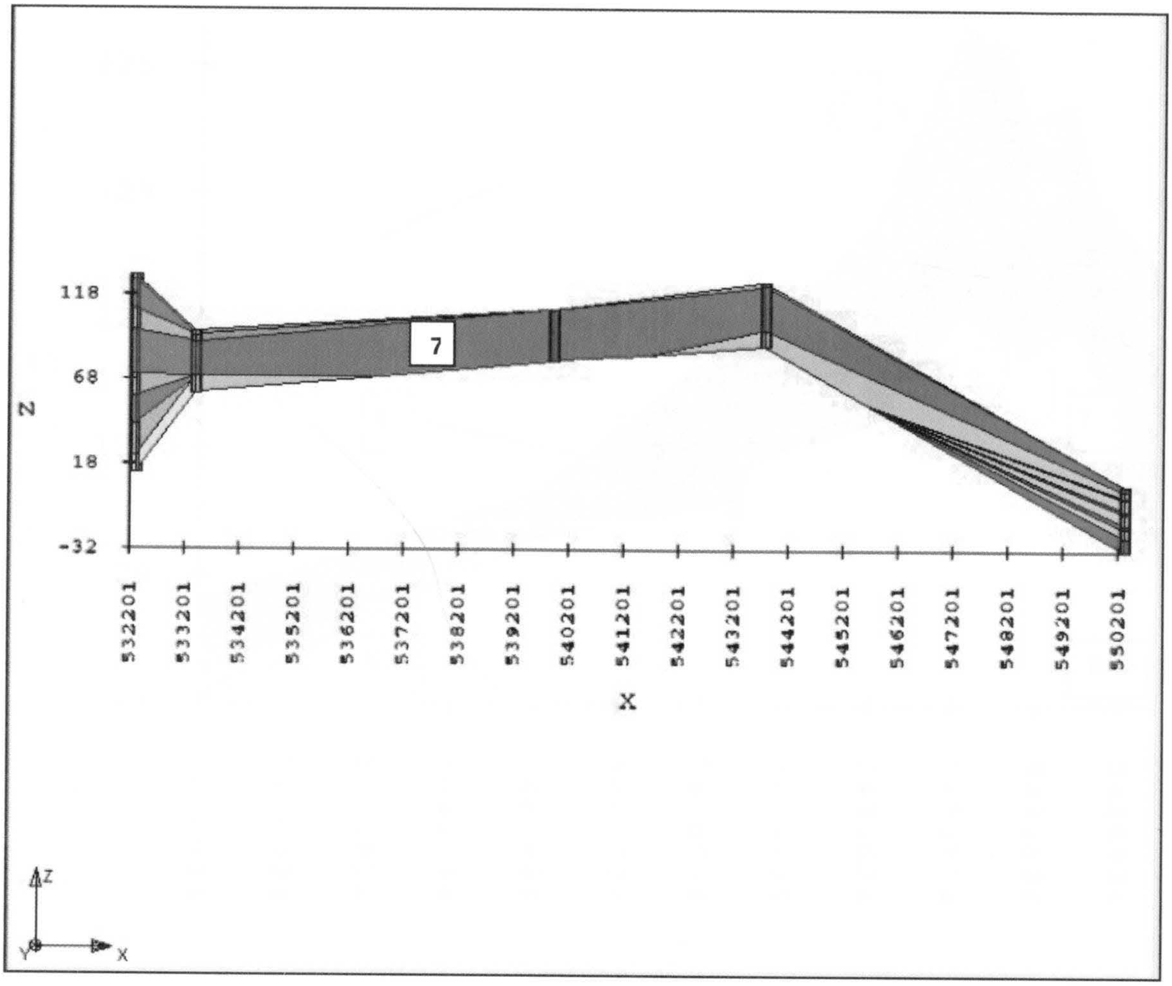
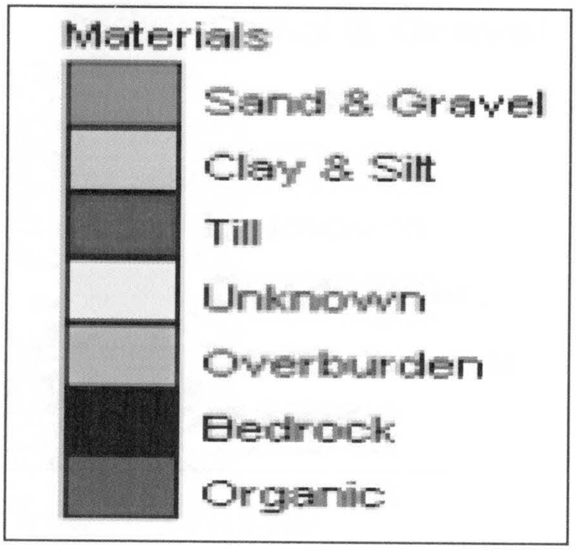


Figure 4-7 Cross-section M-M (generalized hydrostratigraphic model)



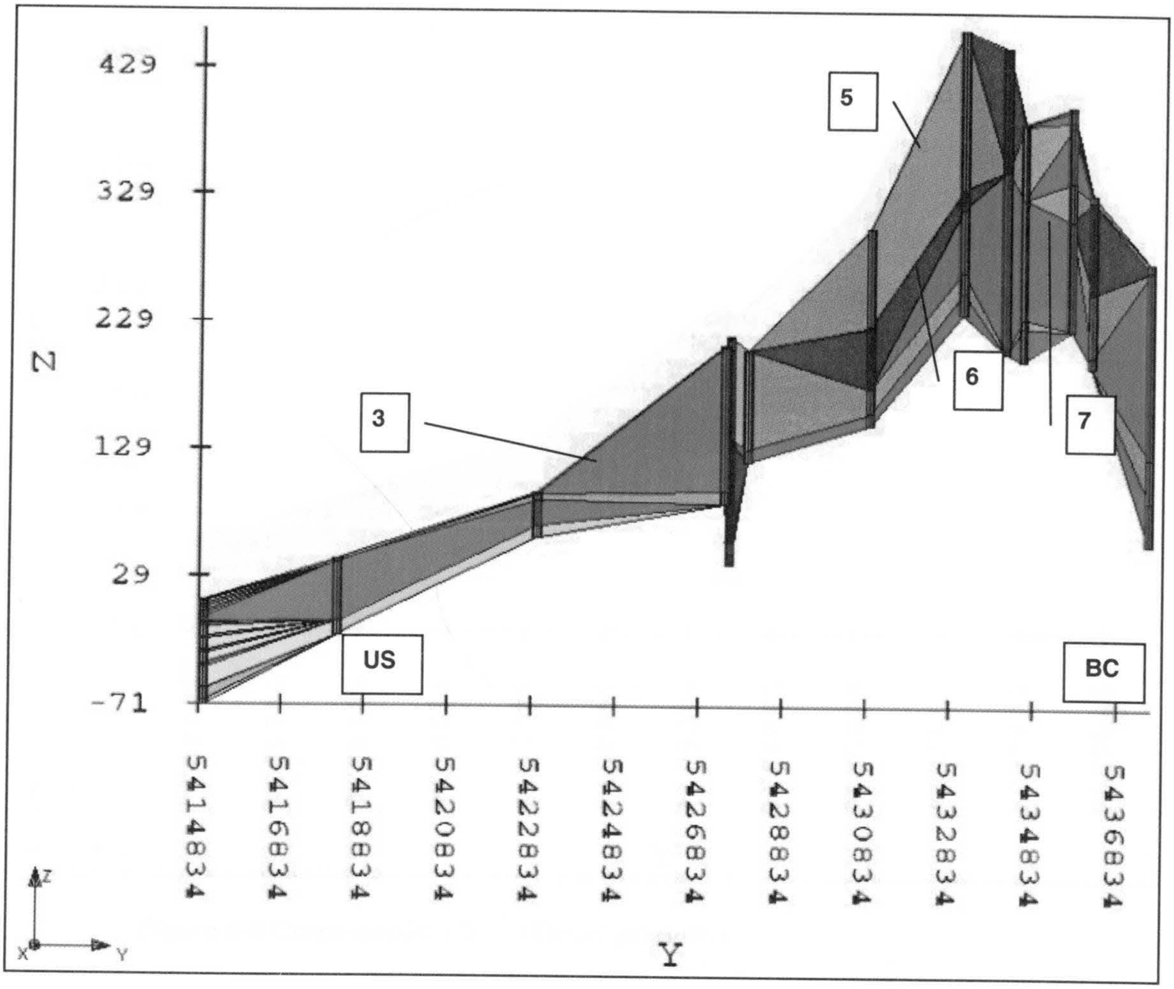
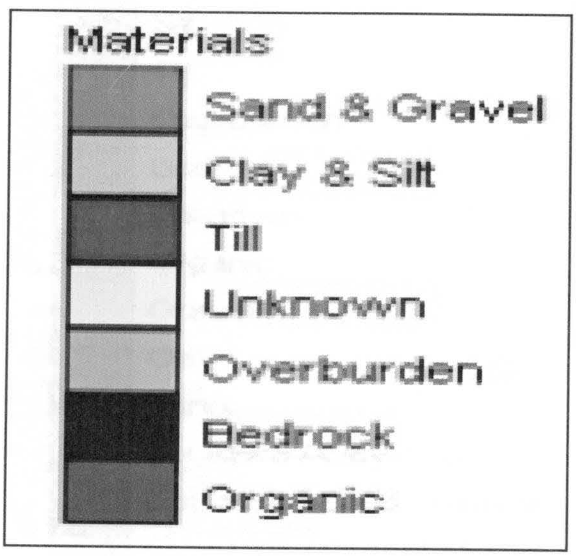


Figure 4-8 Cross-section O-O' (generalized hydrostratigraphic model)



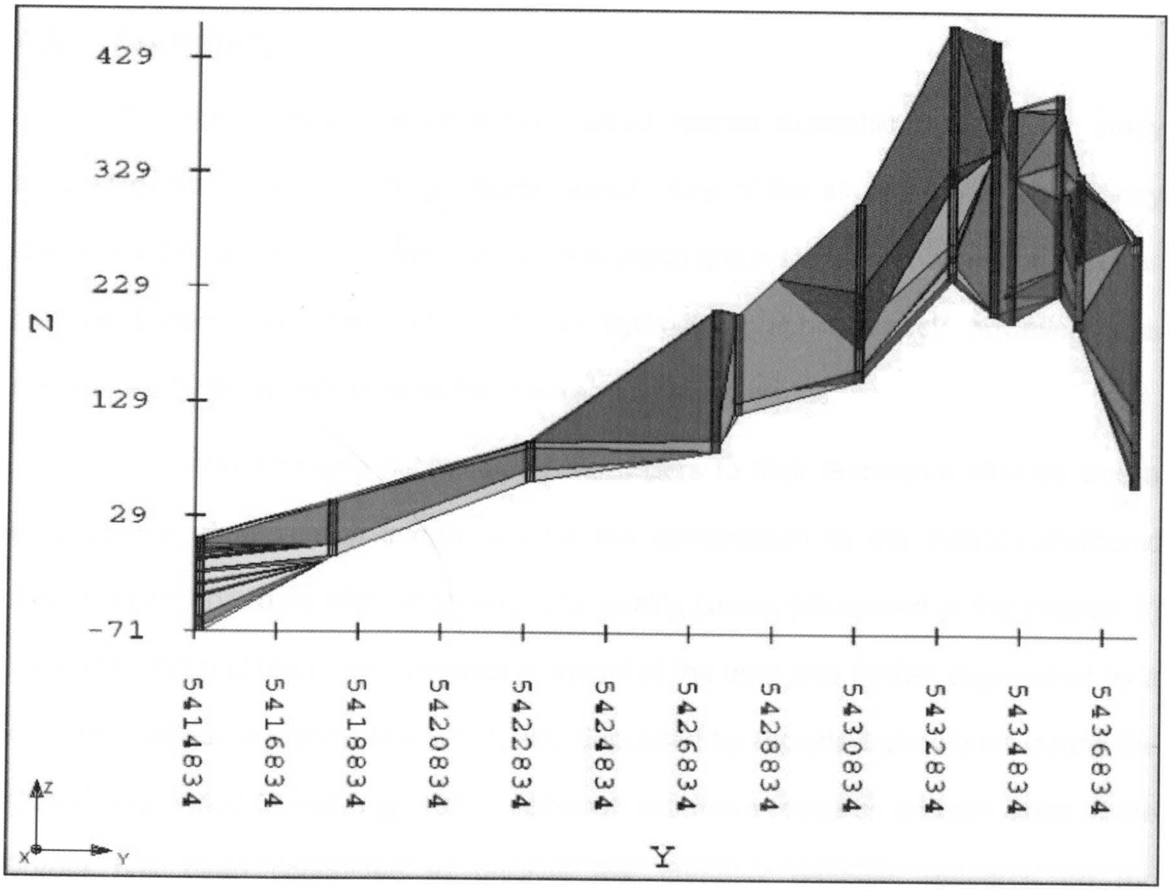
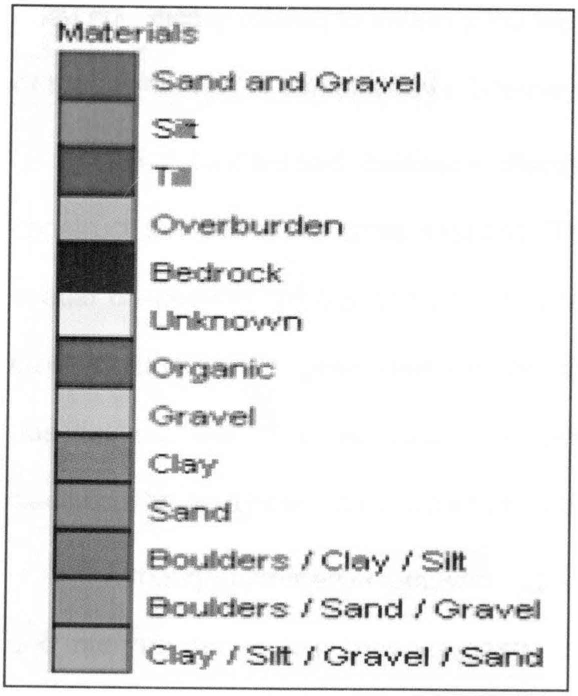


Figure 4-9 Cross-section O-O' (lithologic units)



4.6 Summary

This chapter has provided a generalized hydrostratigraphic model of the study area. Understanding the hydrogeological architecture of the study area is the primary step in building a conceptual framework upon which groundwater management activities can be based. The construction of the hydrostratigraphic model, however, was constrained by the quality of available data and software issues.

The water well data submitted by the drillers to their respective Ministry was a chief source of geological information for the construction of the hydrostratigraphic model. Apart from data integration and data quality issues (discussed in the chapter 3) the construction of the hydrostratigraphic model of the area was further aggravated by a complex geological architecture which was deposited by repeated glaciations during the Quaternary period (Armstrong, 1981). Although extensive research on both sides of the border has been conducted to unravel this complex geology, research on the hydrogeology of the aquifer has concentrated on specific portions of the aquifer, and has been specifically related to determining the hydrostratigraphy, groundwater flow patterns or the distribution of contaminants (Stasney, 2000) at a local level.

The standardized database discussed in the previous chapter was used to construct the geological cross-sections. The wells were selected as a function of their spatial distribution, depths and proximity to cross-section lines. A total of twenty eight cross-sections were generated for the study area. The variation in the geological descriptions, lack of co-relations in adjacent wells prompted the use of a simplified classification (see Appendix A) to develop the hydrostratigraphic model.

Although commercial software packages are available to develop cross-sections and interpret the geology they are based on an assumption that the user is familiar with the conceptual geologic model of the area. In the event of extremely poor quality data, and complicated by a complex geological architecture, the software proved ineffective in developing the conceptual model of the study area. This was further complicated by lack of visualization capabilities afforded by the software. Consequently, a 3-D physical model was manually constructed to visualize the geology of the area and provide a foundation for identifying the hydrostratigraphic units of the study area.

Three main lithology types: Sand & Gravel; Clay & Silt and Till) were identified. These lithologic units alternate with each other and constitute the 9 hydrostratigraphic units (table 4.1). Sand & Gravel lithologic units constitute hydrostratigraphic units HU3 and HU7. These two units are separated by a thick Clay & Silt layer which constitutes hydrostratigraphic unit HU5. Clay & Silt layers occur throughout the study area and constitute hydrostratigraphic units HU2, HU5, and HU8. The Till lithologic units constitute hydrostratigraphic units HU1, HU4, and HU6. Till layers are dispersed throughout the study area and the lateral continuity of till layers was difficult to identify. A continuous till layer (HU6), however, was observed in the western Canadian portion of the study area.

The hydrostratigraphic model was fairly consistent with Halstead's (1986) Hydrogeologic Fence diagrams. The cross-sections are also fairly consistent with Cox and Kahle's (1999) cross-sections although they have used a surficial geological classification for identifying the units. In conclusion, despite the poor quality and inconsistent geological descriptions the wells can be used for constructing the hydrostratigraphic model although provide a generalized hydrostratigraphic model of the area.

The construction of the hydrostratigraphic model for the aquifer was possible due to the development of the integrated database discussed in the previous chapter. Today interoperability remains an unresolved issue and must be considered as a primary step for any environmental management project. Various interoperability issues, discussed in the previous section, can hinder the successful execution of environmental projects. Such issues underscore the importance of interoperability for environmental projects.

5 CONCLUSION AND RECOMMENDATIONS

This research has analyzed the challenges of integrating data for cross-border environmental projects and has evaluated the usability of the resultant lithological databases for constructing a hydrostratigraphic model of the Abbotsford-Sumas aquifer. The problems encountered throughout this research program attest to the reality that interoperability is yet to be achieved (Vckovski, 1998).

Increasing urbanization, anthropogenic activities and the adverse effects of climate change are stressing the environment. Environmental phenomena, however, do not follow political boundaries and their management requires data from multiple sources.[48] Consequently, environmental phenomena need to be monitored to predict future events and reduce damage to humans and property. Interoperability, an unresolved issue for a single jurisdiction, was further exacerbated for cross-border studies where different data cultures exist. Although these research are dependent on the availability of integrated datasets, sparse literature exists on the experiences and challenges of integrating datasets for cross-border studies.[49]

Groundwater, a renewable resource is an important component of the hydrologic cycle and is an example of an environmental issue that does not respect political boundaries. Exploitation of groundwater can produce negative repercussions ranging

---

[48] This is exemplified by the Red River flooding of 1997 which caused extensive damage in parts of Canada and the US.

[49] Two notable exceptions are the OCG and the ISO/TC 211 initiatives.

from direct effects, such as water table lowering to indirect effects, such as subsidence, salt-water intrusion and contamination.

This research examined the Abbotsford–Sumas aquifer; an aquifer that is classified[50] as being highly vulnerable to contamination. This shallow unconfined sand and gravel aquifer supports the activities of approximately 200,000 people who live in this area. Groundwater is dominantly used on both sides of the border for industrial activities, irrigation, poultry farming and municipal and domestic use (Kohut, 1987). Agricultural practices, around such activities as berry cropping, dairy and poultry farming have increased nitrate levels in the groundwater. There are also growing concerns about groundwater issues within the two countries: Canada is concerned with the excessive groundwater withdrawal south of the border (Kohut, 1987), while the US is increasingly concerned about groundwater contamination originating in Canada (Cox and Kahle, 1999). Ensuring a sustainable supply of high quality groundwater, for the various communities reliant on groundwater necessitates groundwater management activities. The ability to do so, however, is constrained by the lack of good quality lithological data with which to constrain the hydrogeologic conceptual model of the aquifer. Currently, geological information contained in water well reports for the study area represents the chief source of depth specific lithological information, although interpretation based on geological mapping have been developed. The use of this information for hydrostratigraphic model development, however, is constrained by the inconsistencies in the geological descriptions. In the absence of geophysical studies, which are useful for delineating subsurface layering, water well reports can be a source of invaluable information. The lack of protocols for integrating cross-border datasets, the lack of

[50] BC aquifer classification has identified the Abbotsford-Sumas aquifer as type 1A (Highly used, highly developed and highly vulnerable to contamination) (Berardinucci and Ronneseth, 2002).

standardized and integrated groundwater datasets, and the lack of geophysical data to complement the development of a hydrostratigraphic model further complicates the ability of researchers to construct a hydrostratigraphic model.

This research has addressed the following three objectives:

1. The challenges of integrating datasets for cross-border projects;
2. The creation of an integrated dataset for groundwater management activities; and
3. The development of a hydrostratigraphic model to evaluate the usability of the water well database.

To achieve the above mentioned objectives spatial and attributed data were obtained from various organizations in Canada and US Data were either available in paper or digital format. Data in paper format necessitated digitizing; an error-prone data input strategy. The data from various sources were also corrected for errors. The errors in the lithological database were resolved by writing Visual Basic and database scripts. The data from Canada and the US were then converted to a common format in order to facilitate data integration. Due to a lack of defined methodologies for integrating datasets, the Canadian and US datasets were integrated on an ad hoc basic. Schematic heterogeneities were resolved by mapping the US datasets onto the Canadian dataset. Semantic heterogeneities such as inconsistencies in the geological descriptions were resolved by standardizing the data. This integrated dataset was then used for constructing a hydrostratigraphic model of the study area.

Numerous interoperability challenges, including the availability of data, metadata, data formats and quality, database structure, semantics, policies and cooperation were identified as inhibitors of data integration for cross-border studies. Such interoperability issues must be considered as a starting point for any environmental project, and can be addressed on the basis of data constraints and institutional constraints.

5.1 Data Constraints

Data are the driving force of the GIS industry, yet the inherent nature of spatial data proves to be the source of most interoperability problems. Data format is one of the foremost criteria that should be considered for data integration. Apart from proprietary software formats, data may exist either in a paper format or may be structured in some other way that limits direct use for analysis. Consequently, converting the data to a usable format may require digitizing or extensive programming. Furthermore, new technologies including internet mapping applications, which are being adopted by organizations to distribute data, may not yield data in format that can be used for analysis. Currently, a user is required to convert data to a digital format, irrespective of whether the data are downloaded from an Internet Mapping application or from the Ministry's[51] website. This results in a multitude of databases, which further exacerbates the interoperability problems.

Data quality, defined by its fitness for use, is another important factor to be considered during integration. The inconsistent geological descriptions and the extremely poor data quality were the reasons cited by water well consultants for the infrequent use of the water well data. Apart from conventional data quality parameters (including accuracy, precision and consistency), database errors must be rectified prior to analysis as they prohibit database querying and affect the data integration process. A multitude of errors, including a lack of locational information, surface elevations, and layer sequence numbers and an inconsistent number of records in the tables were observed in the course of this research, particularly in the less frequently used datasets that serve a small user community. This is exemplified by the good-quality topographic

[51] Ministry of Water, Land and Air Protection (MWLAP)

data distributed by the Ministry of Sustainable Resource Management (MSRM) and the United States Geological Survey (USGS) as opposed to the free groundwater data distributed by the Ministry of Water, Land and Air Protection (MWLAP), USGS and the Department of Ecology Washington State which contain numerous database errors.

Data heterogeneities were also observed to be a source of interoperability problems. Although database errors were resolved by extensive scripting, a lack of defined methodologies made schematic and semantic heterogeneities difficult to resolve. Unlike schematic heterogeneities, which result due to differences in database structure, semantic heterogeneities result due to differences in meaning of data. Schematic heterogeneities, including different field or table names for semantically similar elements, as well as differing numbers of fields employed to store semantically similar information, were observed. In the absence of a well defined methodology, these schematic heterogeneities were overcome on an ad hoc basis during the mapping of the Washington State database onto the BC database.

Several instances of semantic heterogeneities were also observed, and were the most difficult to resolve. Datasets captured and maintained by organizations usually reflect jurisdictional mandates and policies. The British Columbia TRIM[52] datasets, for example, are based on the CCSM[53] classification, while the United States Geological Survey (USGS) 7.5 topographic quadrangles are based on the Spatial Data Transfer Standard (SDTS) conceptual model. The use of different models by individual jurisdictions introduces semantic issues. The single USGS hydrography dataset, for example, consists of rivers, streams, ditches, ponds and lakes whereas the TRIM

---

[52] TRIM: Terrain Resource Information Management

[53] CCSM: Canadian Council on Surveys and Mapping

hydrography datasets are captured in two themes: Lakes and Rivers & Streams. Similarly, the transportation data were also captured differently in the two datasets. Semantic differences in the geological descriptions were resolved by classifying the datasets using the FSSD software. Due to a lack of semantic interoperability solutions, the Canadian and the American datasets were integrated separately.

5.2 Institutional Factors

Cross-border research necessitates an investigation of the institutional and organizational factors that impact interoperability. Institutional and organizational factors have been identified as the most difficult to resolve, and in extreme cases may result in stalled initiatives (Craig, 1995, Pinto and Rushton, 1995; Onsrud and Rushton, 1995). Institutional factors can have a direct or indirect impact on interoperability, and may slow cross-border studies.

The lack of metadata for Canadian datasets was identified as the most important factor. Lack of metadata not only prevents data discovery, but can also result in erroneous results. Although Discovery Portal of the Canadian Geospatial Data Infrastructure (CGDI) provides a gateway for organizations to advertise their datasets or organizational information, institutions have been slow in submitting this information. In all, only 523[54] organizations in Canada have submitted information to Discovery Portal: of these only two provincial ministries in BC and seven municipalities in Canada have submitted information. Due to a lack of indigenously developed metadata standards, organizations distribute metadata in a text format, which may be self explanatory or by word of mouth. By contrast, metadata for US datasets is based on the Federal Geographic Data Committee Content Standard. The distribution of metadata in FGDC

[54] This number includes municipal, provincial, federal and commercial institutes.

compliant format was mandated by President Clinton's Executive Order 12096 (1994), which established the National Spatial Data Infrastructure (NSDI) and enforced all federal organizations and urged spatial data producing organizations to document their datasets and make them electronically available. The success of spatial data infrastructures is demonstrated by the numerous state and local spatial data infrastructure and availability of a variety of spatial datasets at the state and local levels.

In addition to metadata issues copyright issues, data dissemination policies and cost of datasets are important parameters to be considered. Sieber (2003) quoted Birkenstock, a GIS technician, who reported that Canadian data dissemination and cost recovery policies hindered the development of an online internet mapping application for the benefit of the communities following the 1997 Red River Floods. In this research, the US datasets were available at no charge, whereas the Canadian datasets were available for a charge. The TRIM map sheets, for example, are distributed at a cost of $400 Cdn per map sheet. The issue of data costs was discussed by Klinkenberg (2003) where he has noted that some Canadian datasets may be available at a lower cost from US organizations compared to the same datasets distributed by Canadian organizations.

Institutional reluctance and organizational dynamics are also important parameters that need to be considered for cross-border projects. Institutional inertia in the form of reluctance to provide information or the lack of awareness of activities in the same department or levels of governance, which may result in enormous amount of time being wasted in data entry, was evident in the course of this research. Further, a lack of organizational coordination by the USGS and the Department of Ecology in Washington State has resulted in the creation and maintenance of two independent groundwater datasets which lack cross referencing.

Institutional reluctance to adhere to standards can also result in technical interoperability problems. A lack of legislated mandates enables organizations to distribute data in any format they deem appropriate. Unlike the United States, for example, where federal datasets are distributed in SDTS format, Canadian data are distributed in a format based on jurisdictional mandates. TRIM datasets, the equivalent of the USGS 7.5 quadrangles topographic maps, are distributed in SAIF[55], MOEP[56] or ESRI format despite SAIF being a Canadian National Standard.

The observed interoperability challenges for integrating cross-border data requires resolution at technical as well as institutional levels. This is consistent with Evans' and Ferreira's (1995) argument that as we are living in a technologically evolving world, interoperability research should focus on the overlap of these two approaches.

5.3 Hydrostratigraphic Model and Software Issues

The inconsistencies in the geological description (Russel et al., 1998) and the extremely poor quality (discusses in chapter 3) of the original lithological database constrained the development of the hydrostratigraphic model. Approximately 6000 unique geological descriptions were observed for the study area. These descriptions were further standardized to thirteen categories (Appendix A: Table A-6; Figure A-1). Due to the lack of correlation between adjacent wells these categories were further categorized to seven simplified units (Appendix A, Table A-7) which were then used for construction of the cross-section and hydrostratigraphic model.

---

[55] SAIF: Spatial Data Transfer Standard

[56] MOEP: stands for Ministry of Environment and Parks, however, does not have bearing to the current ministry name (Ministry of Sustainable Resource Management) (Poire, 2004).

Today, although software packages are available that facilitate the development of cross-sections and the construction of a hydrostratigraphic model; they are premised upon the assumption that good quality data, geophysical surveys, or drill core records are available to supplement the development of the hydrostratigraphic model. In the absence of good quality data, the software packages hinder the development of a conceptual model. This is further complicated by the lack of visualization capabilities afforded by the software and their ability to handle the issue of lenses.

The conceptual model for this research was generated by manually constructing a 3-D physical model of the study area. The layers identified were used for constraining hydrostratigraphic units in GMS. Two outputs were generated for each cross-section: the generalized hydrostratigraphic model was generated using the simplified classification, whereas the lithologic units were generated using the final classification (see Appendix A).

The hydrostratigraphic models consist of an alternating sequence of sand & gravel, clay & silt and till units. A total of 9 hydrostratigraphic units[57] were identified. These hydrostratigraphic units comprise three lithologic units:

- **Sand & Gravel:** Two discrete sand & Gravel layer were identified in the study area and constitute hydrostratigraphic units (HU3 and HU7). HU7 occurs in the western portion of the study area and is laterally separated from HU3, which occurs in the eastern and southern portion of the study area by a thick clay & silt unit (HU5). HU3 is overlain, in the eastern portion of the study area by Sumas clay (Cameron, 1989).

[57] The hydrostratigraphic units are based on my interpretation of the layers observed in the cross-sections. Given the complexity of the study area and the uncertainty associated with the dataset relative age relationships could not be determined and hydrostratigraphic units of the same age may be observed higher or lower in the hydrostratigraphic model.

- **Clay & Silt:** Clay & silt layers occur throughout the study area and either overlies or underlies the sand & gravel layers. A thick clay & silt deposit (HU5) laterally separates HU3 and HU7.
- **Till:** Till layers appear as discontinuous layers throughout the study area and constitute hydrostratigraphic units (HU1, HU4 and HU6). Although, Halstead has identified continuous till layers in certain areas, it was difficult to identify the lateral continuity of the till layers. HU6 is a continuous till layer that occurs in the western portion of the study area.

The hydrostratigraphic model generated for the study area is fairly consistent with the Hydrogeologic Fence diagrams developed by Halstead (1986) and the cross-sections developed by Cox and Kahle (1999). Despite the poor quality of the lithological database trends were visible that facilitated the development of the hydrostratigraphic model of the area.

5.4 Future Research and Recommendations

Although this research has provided an exploratory analysis of integration issues for cross-border studies, several opportunities exist for further research. One of the most important issues identified was a lack of metadata for Canadian datasets. An inquiry into the factors that prevent organizations from creating metadata and organizational reluctance to submit this information to CGDI needs to be investigated. Data costs, data dissemination policies and copyright issues are factors often cited with respect to availability and access to datasets. The impacts of metadata on availability and access to information would also serve as a future research topic. Further research is also required to thoroughly understand interoperability issues for cross-border projects. Organizational interoperability research has been situated within an American context; further research is required to understand these issues within a Canadian context.

Database errors were identified as one of the causes that hindered the data integration process. Although these errors can be resolved by writing database scripts; considerable amount of time is wasted in improving the data quality. Database errors, if rectified, during the design stage can significantly improve data quality. The database errors in this research were attributed to the design of the CGDS system. Following are recommendations for improving the structure of the BC lithological database:

Lithology table

- Create a standard list of geological terms. Most drillers use geological terminology from Halstead's (1986) and Armstrong's reports (1981; 1984).
- Increase the Geological description field size to 255 or 'memo type'
- Add a new field for storing layer depths.
- Recalculate sequence numbers. Inconsistent sequences numbers were observed.
- Create a new field (memo type) to store important groundwater information that is included by drillers as part of the geological descriptions.

Location Table

- Transfer fields such as Use, Diameter, ConstructionDate, Method and WellDepth from the general table to the location table.
- Create a new table to store information on screens (Welltagnum, Screen id, Screenfrom, ScreenTo, ScreenDiameter, ScreenSlotSize, and ScreenManufacturer).
- Create a new table to store information on Casings (WellTagNum, Casing ID, CasingFrom, CasingTo)
- Create a standard list for drilling methods.
- Create a new field to store locational accuracy.
- Create a new field to store elevation accuracy.

Database Structure

- Create new tables for storing information on pumping tests, recovery and contamination.

- Create integrity rules so that important fields, such as, geological description, locational information, and surface elevation information cannot accept null values.
- Create integrity rules for tables so that well information is entered in all tables. This will eliminate the problem of inconsistent well records in the tables.

There are several opportunities for future research from a hydrogeological perspective. Development of the hydrostratigraphic model is very important for groundwater management activities as it elucidates the spatial continuity of the sub-surface layers. The hydrostratigraphic model thus helps understand groundwater and contaminant flow. Geophysical field surveys for the study area would definitely help build a more robust hydrostratigraphic model. In the Oak Ridges Moraine project, for example, Russell et al. (1998) found that there was considerable use of the term clay in databases, although sedimentological studies showed that clay constituted only 2% of the total sediments. Also, research on development of standards and software that facilitate better visualization would definitely be an added benefit.

APPENDIX A: DATASETS, INTEROPERABILITY AND STANDARDIZATION

A1. Datasets and Interoperability

Dataset	British Columbia (Canada)	Washington State (US)
Lithological data	1. BC Ministry of Water, Land and Air Protection (MWLAP, 2004) 2. Ministry of Transportation 3. Drill Records (Cameron, 1989)	1. USGS (Doremus, 2002) 2. (Department of Ecology (Washington State) , 2004))
Digital Elevation Model (DEM)	Ministry of Sustainable Resource Management (MSRM, 2004)	United States Geological Survey (USGS, 2004)
Topographic Data	MSRM (map sheets 92G008, 92G009, 92G010) (MSRM, 2004)	USGS (7.5 quadrangles Bertrand Creek, Kendall, Sumas, Lynden) (USGS, 2004)
Surficial Geology Maps	Geological Survey of Canada (paper format)	(Department of Ecology (Washington State, 2004))

Table A-1 List of datasets obtained for this research

Interoperability Issues	British Columbia (Canada)	Washington State (US)
Data format	SAIF / ESRI	SDTS(mandatory by federal agencies)
Projection	BC Albers / UTM	UTM / SPCS
Datum	NAD 83	NAD 27
Metadata	FGDC compliant (Voluntary)	FGDC Complaint (Mandatory)
Scale	1:20000	1:24000
Vertical datum	CGDV 28	NGDV

Table A-2 Interoperability issues

A2. Standardization and Classification

	Lithology
1	Bedrock
2	Boulder
3	Boulder and Clay
4	Boulder and Glacial
5	Boulder and Gravel
6	Boulder and Overburden
7	Boulder and Sand
8	Clay
9	Clay and Boulder
10	Clay and Glacial
11	Clay and Gravel
12	Clay and Organic
13	Clay and Other sedimentary rocks
14	Clay and Overburden
15	Clay and Sand
16	Clay and Shale
17	Clay and Silt
18	Glacial
19	Glacial and Boulder
20	Glacial and Clay
21	Glacial and Gravel
22	Glacial and Sand
23	Glacial and Shale
24	Glacial and Silt
25	Granite and Boulder
26	Gravel
27	Gravel and Boulder
28	Gravel and Clay
29	Gravel and Glacial
30	Gravel and Organic
31	Gravel and Overburden
32	Gravel and Sand
33	Gravel and Shale
34	Gravel and Silt
35	Organic
36	Organic and Clay
37	Organic and Gravel
38	Organic and Silt
39	Overburden
40	Overburden and Boulder
41	Overburden and Clay
42	Overburden and Glacial
43	Overburden and Gravel
44	Overburden and Sand
45	Sand
46	Sand and Boulder
47	Sand and Clay
48	Sand and Glacial
49	Sand and Gravel
50	Sand and Organic
51	Sand and Other sedimentary rocks
52	Sand and Overburden
53	Sand and Silt
54	Sandstone and Clay
55	Shale and Clay
56	Shale and Glacial
57	Shale and Gravel
58	Shale and Sand
59	Silt
60	Silt and Boulder
61	Silt and Clay
62	Silt and Glacial
63	Silt and Gravel
64	Silt and Organic
65	Silt and Sand
66	Unknown

Table A-3 List of categories after standardizing lithologic terms using FSSD

A3. Rule Based Classification

Used to further classify lithologic terms obtained from FSSD

Note: material is a general term used for any unconsolidated material

1) IF

GEOLOGY = Overburden + Material OR Material + Overburden

AND

LAYERORDERNUM = 1

THEN

GEOLOGY = Overburden

ELSE

GEOLOGY = Material

2) IF

GEOLOGY = Till and Boulders / Boulders AND Glacial / Glacial and Boulders

THEN

GEOLOGY = Bouldery Till

Original Description	Final Description
Till And Boulders Boulders And Glacial Glacial And Boulders	Bouldery Till
Clay And Gravel Gravelly Clay	Clay And Gravel / Gravelly Clay
Sand And Gravel Gravelly Sand	Sand And Gravel/ Gravelly Sand
Clay And Sand Sandy Clay	Clay And Sand / Sandy Clay
Sand And Clay Clayey Sand	Sand And Clay / Clayey Sand
Sand And Silt Silty Sand	Sand And Silt/ Silty Sand
Silt And Sand Sandy Silt	Silt And Sand / Sandy Silt
Silt And Gravel Gravelly Silt	Silt And Gravel/ Gravelly Silt
Gravel And Silt Silty Gravel	Gravel And Silt/ Silty Gravel
Clay And Silt Silty Clay	Clay And Silt / Silty Clay
Organic And Material / Material + Organic	Material
Bedrock And Material (original records were verified)	Material
Till And Material Material And Till	Till And Material / Material And Till

Table A-4 List of categories after applying the rule based classification

	Geological Material	Canada	Us	Canada_%	Us_%
1	Bedrock	214	142	1.24	2.80
2	Boulders	88	25	0.51	0.49
3	Boulders and Clay	64	17	0.37	0.34
4	Boulders and Gravel	89	30	0.52	0.59
5	Boulders and Sand	19	0	0.11	0.00
6	Boulders and Silt	2	4	0.01	0.08
7	Bouldery Till	95	7	0.55	0.14
8	Clay	3112	579	18.07	11.43
9	Clay and Gravel / Gravelly Clay	449	262	2.61	5.17
10	Clay and Sand / Sandy Clay	748	146	4.34	2.88
11	Clay and Silt / Silty Clay	226	10	1.31	0.20
12	Gravel	1388	295	8.06	5.82
13	Gravel & Clay / Clayey Gravel	137	94	0.80	1.86
14	Gravel and Sand / Sandy Gravel	2846	695	16.52	13.72
15	Gravel and Silt / Silty Gravel	151	14	0.88	0.28
16	Organic	12	9	0.07	0.18
17	Overburden	1327	820	7.70	16.19
18	Sand	2809	883	16.31	17.43
19	Sand and Clay / Clayey Sand	364	144	2.11	2.84
20	Sand and Silt / Silty Sand	227	40	1.32	0.79
21	Silt	385	5	2.23	0.10
22	Silt and Clay / Clayey Silt	0	15	0.00	0.30
23	Silt and Gravel / Gravelly Silt	37	21	0.21	0.41
24	Silt and Sand / Sandy Silt	95	54	0.55	1.07
25	Stoney Clay	518	0	3.01	0.00
26	Till	1513	83	8.78	1.64
27	Till and Clay / Clay and Till	217	6	1.26	0.12
28	Till and Gravel / Gravel and Till	192	12	1.11	0.24
29	Till and Sand / Sand and Till	97	4	0.56	0.08
30	Till and Silt / Silt and Till	23	0	0.13	0.00
31	Unknown	300	650	1.74	12.83
	TOTAL	17226	5066	100	100

Table A-5 Figure showing percentages of geological material for Canada and US

	Final Classification	Canada	US	Canada_%	US_%
1	**Unknown**	300	650	1.74	12.83
2	**Silt**	385	5	2.23	0.10
3	**Sand**	2809	883	16.31	17.43
4	**Overburden**	1327	820	7.70	16.19
5	**Organic**	12	9	0.07	0.18
6	**Gravel**	1388	295	8.06	5.82
7	**Clay**	3112	579	18.07	11.43
8	**Bedrock**	214	142	1.24	2.80
9	**Till**				
	Till	1513	83	8.78	1.64
	Till and Clay / Clay and Till	217	6	1.26	0.12
	Till and Gravel / Gravel and Till	192	12	1.11	0.24
	Till and Sand / Sand and Till	97	4	0.56	0.08
	Till and Silt / Silt and Till	23	0	0.13	0.00
	Bouldery Till	95	7	0.55	0.14
10	**Sand and Gravel**				
	Gravel and Sand / Sandy Gravel	2846	695	16.52	13.72
	Gravel and Silt / Silty Gravel	151	14	0.88	0.28
	Sand and Silt / Silty Sand	227	40	1.32	0.79
	Silt and Gravel / Gravelly Silt	37	21	0.21	0.41
	Silt and Sand / Sandy Silt	95	54	0.55	1.07
11	**Boulders / Sand / Gravel**				
	Boulders and Gravel	19	0	0.11	0.00
	Boulders and Sand	64	17	0.37	0.34
12	**Boulders / Clay / Silt**				
	Boulders and Clay	64	17	0.37	0.34
	Boulders and Silt	2	4	0.01	0.08
	Boulders	88	25	0.51	0.49
13	**Clay / Silt / Gravel / Sand**				
	Clay and Gravel / Gravelly Clay	449	262	2.61	5.17
	Clay and Sand / Sandy Clay	748	146	4.34	2.88
	Clay and Silt / Silty Clay	226	10	1.31	0.20
	Gravel & Clay / Clayey Gravel	137	94	0.80	1.86
	Sand and Clay / Clayey Sand	364	144	2.11	2.84
	Silt and Clay / Clayey Silt	0	15	0.00	0.30
	Stony Clay	518	0	3.01	0.00

Table A-6 Final Classification

	Simplified Classification
1	**Bedrock**
2	**Unknown**
3	**Organic**
4	**Overburden**
5	**Clay / Silt**
	Clay
	Silt
	Clay / Silt / Gravel / Sand
	Boulders / Clay / Silt
6	**Sand / Gravel**
	Sand
	Gravel
	Sand and Gravel
	Boulders / Sand / Gravel
7	**Till**
	Till

Table A-7 Simplified Classification (used for constructing the hydrostratigraphic model)

A4: Classification

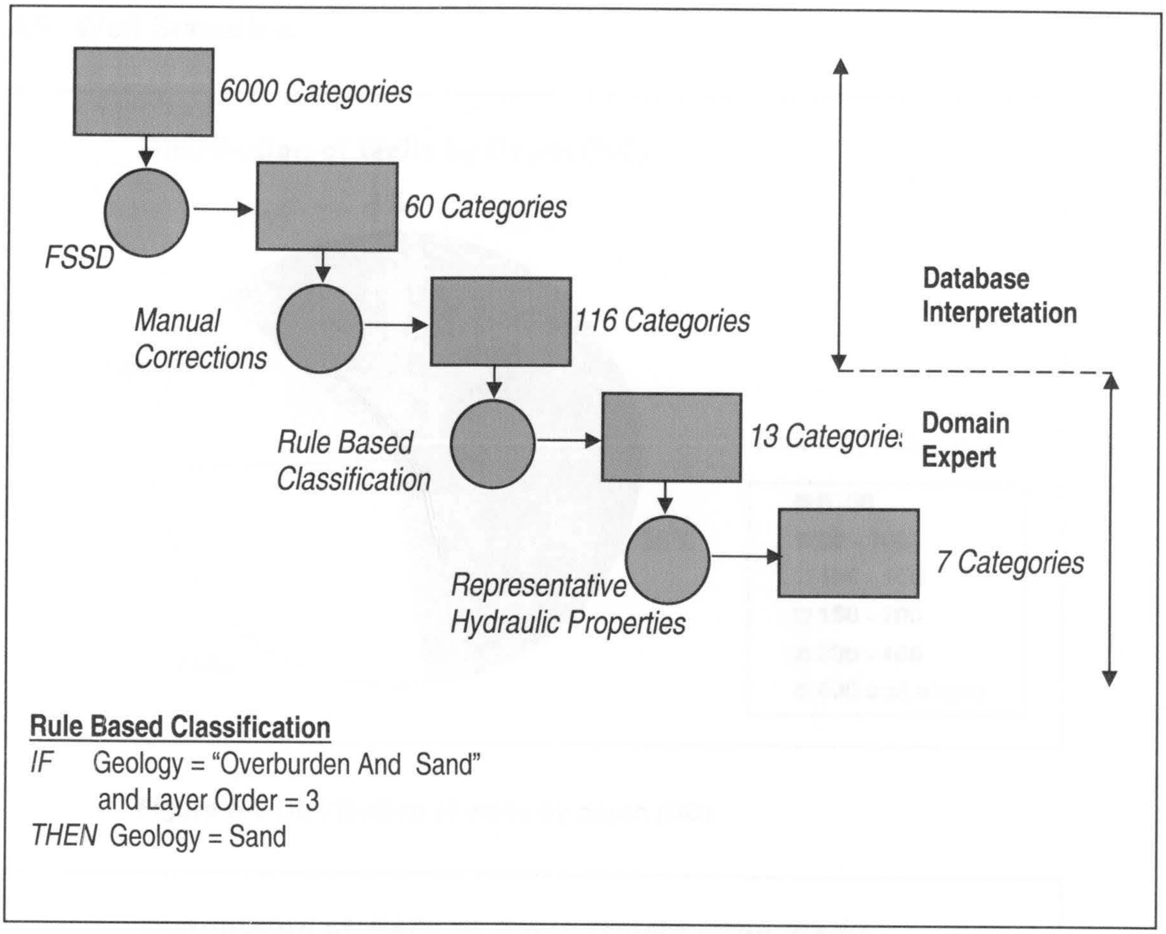


Figure A-1 Classification process

A5. Well Statistics

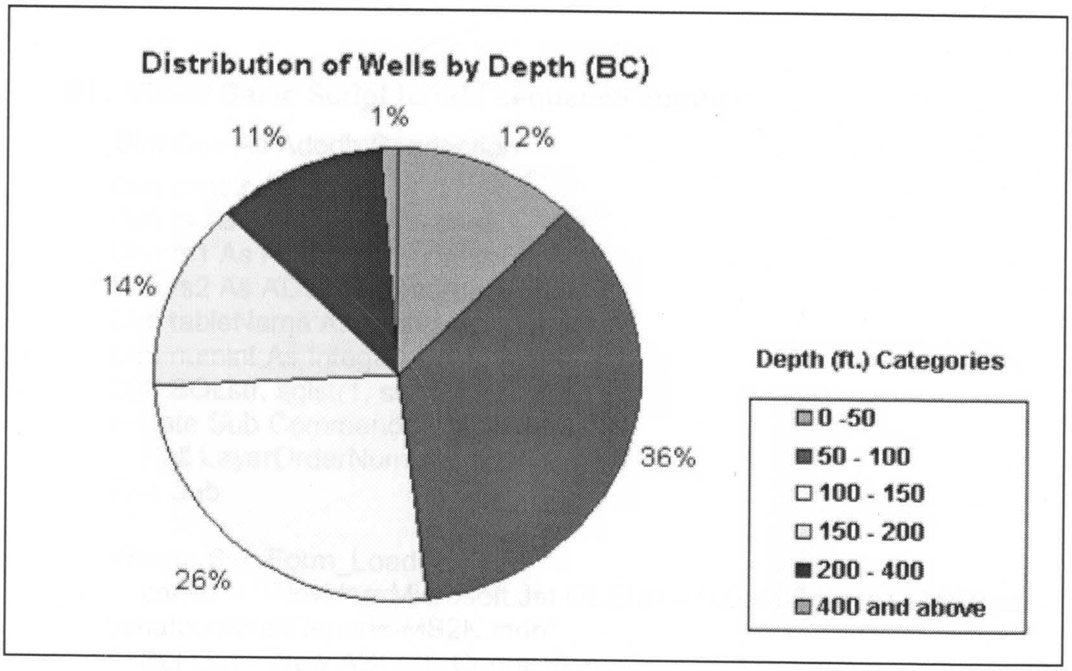


Figure A-2 Distribution of wells by depth (BC)

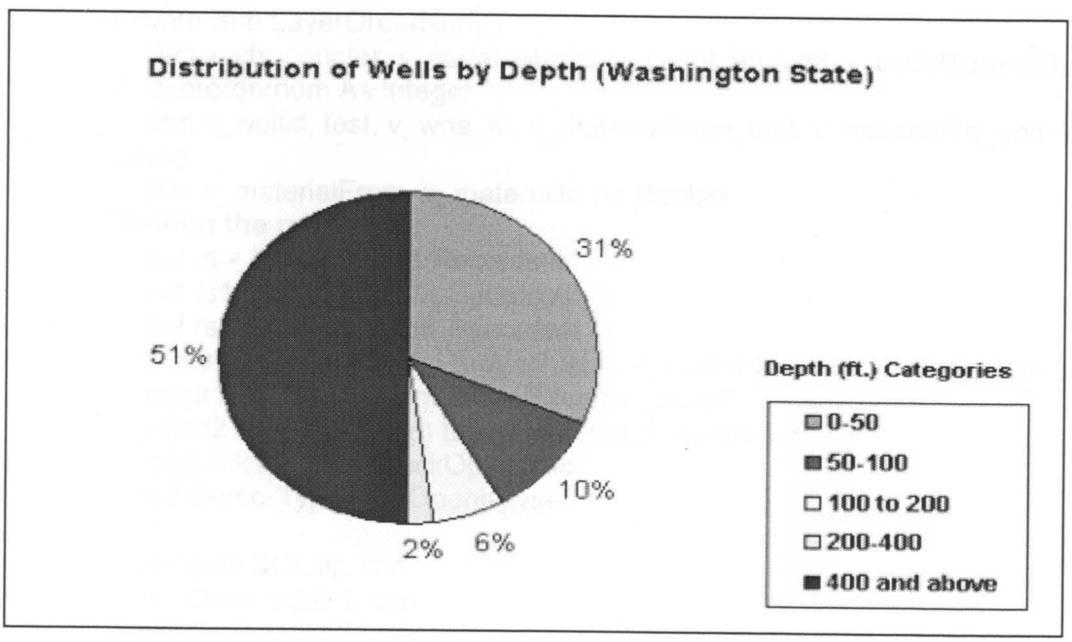


Figure A-3 Distribution of wells by depth (Washington State)

APPENDIX B: VISUAL BASIC SCRIPTS

B1. Visual Basic Script to add sequence number

```
Dim Cnn As Adodb.Connection
Dim cnnstr As String
Dim rs As ADODB.Recordset
Dim rs1 As ADODB.Recordset
Dim rs2 As ADODB.Recordset
Dim tableName As String
Dim numint As Integer
Dim SQLstr, sqlstr1, sqlstr2 As String
Private Sub Command1_Click()
   Call LayerOrderNum
End Sub

Private Sub Form_Load()
   cnnstr = "Provider=Microsoft.Jet.OLEDB.4.0;Data Source=c:\aparna\
\whatcomWellReports-MS2K.mdb"
   Set cnn = New ADODB.Connection
   cnn.Open cnnstr
End Sub

Private Sub LayerOrderNum()
   Dim v_ctr, counter, v_codeMaterial, v_codeMaterial2, v_codeMaterial3,
v_layerordernum As Integer
   Dim v_wellid, test, v_wria_id, v_materialFrom_unit, v_materialTo_unit As
String
   Dim v_materialFrom, v_materialto As Double
'Setting the recordset
   Set rs = New ADODB.Recordset
   Set rs1 = New ADODB.Recordset
   Set rs2 = New ADODB.Recordset
   SQLstr = "Select * from tblWellMaterial_f order by wellid, materialfrom asc"
   sqlstr1 = "Select * from tblWellMaterial_count"
   sqlstr2 = "select * from tblwellmaterial_f_layerorder"
   rs2.LockType = adLockOptimistic
   rs2.CursorType = adOpenKeyset

   rs.Open SQLstr, cnn
   rs1.Open sqlstr1, cnn
   rs2.Open sqlstr2, cnn
   counter = 1
   Do Until rs1.EOF
      v_ctr = rs1.Fields!Count
      test = rs1.Fields!wellid
```

```
v_wellid = rs.Fields!wellid


'counter = 1

If v_wellid = test Then
   For counter = 1 To v_ctr
      v_materialFrom = rs.Fields!Materialfrom
      v_materialto = rs.Fields!materialto
      v_codeMaterial = rs.Fields!codematerial
      v_codeMaterial2 = rs.Fields!codeMaterial2
      v_codeMaterial3 = rs.Fields!codematerial3
      v_wria_id = rs.Fields!wria_id
      v_materialFrom_unit = rs.Fields!materialfrom_unit
      'v_layerordernum = rs.Fields!LayerOrderNum
      'v_materialTo_unit = rs.Fields!materialto_unit
      If rs.Fields!materialto_unit <> "" Then
         v_materialTo_unit = rs.Fields!materialto_unit
Else
         v_materialTo_unit = ""
      End If
      'If rs.Fields!LayerOrderNum <> "" Then
      'v_layerordernum = rs.Fields!LayerOrderNum
      ' Else
      '  v_layerordernum = Null
      ' End If
      If v_materialFrom = 0 Then
         v_layerordernum = 1
      End If
      If v_materialFrom <> 0 Then
         v_layerordernum = counter
      End If
      rs2.Addnew
      rs2.Fields("wellid") = v_wellid
      rs2.Fields("codematerial") = v_codeMaterial
      rs2.Fields("codematerial2") = v_codeMaterial2
      rs2.Fields("codematerial3") = v_codeMaterial3
      rs2.Fields("materialfrom") = v_materialFrom
      rs2.Fields("materialto") = v_materialto
      rs2.Fields("materialfrom_unit") = v_materialFrom_unit
      rs2.Fields("materialto_unit") = v_materialTo_unit
      rs2.Fields("LAyerordernum") = v_layerordernum
      rs2.Update
      rs2.movenext
      rs.movenext
   Next counter
   rs1.movenext

End If
```

```
  Loop
End Sub
```

B2. Visual Basic Script to update elevation

- LITHOLOGY_X1 (stores the records)
- LITHOLOGY_LAYRES _COUNT (store the total number of layers present in each layer)
  ```
  SELECT count(welltagnum) into LITHOLOGY_LAYERS_COUNT
  FROM LITHOLOGY_X1
  GROUP BY Welltagnum
  ```
- LITHOLOGY_X1_FINAL (this table is an empty table and has the same structure as LITHOLOGY_X1. This table will be automatically populated)

```
Option Explicit
Dim cnn As ADODB.Connection
Dim cnnstr As String
Dim rs As ADODB.Recordset
Dim rs1 As ADODB.Recordset
Dim rs2 As ADODB.Recordset
Dim tableName As String
Dim numint As Integer
Dim SQLstr, sqlstr1, sqlstr2 As String

Private Sub Command1_Click()
   Call update_elevation
End Sub

Private Sub Form_Load()
   cnnstr = "Provider=Microsoft.Jet.OLEDB.4.0;Data
Source=d:\conv_water_well(aug4).mdb"
   Set cnn = New ADODB.Connection
   cnn.Open cnnstr
  End Sub
Private Sub update_elevation()
  Dim v_ub, v_lb, v_ld, v_lon As Integer
  Dim v_wtn As String
  Dim v_cnt As Integer
  Dim vn_wtn As String
  Dim vn_ub, vn_lb, elevation, vn_ld, vn_lon, v_liRcdID As Integer
 Dim v_description, v_errcomment, v_manual_correction, v_comment,
 v_description1, v_utm_accuracy As String
  Dim counter As Integer
  Dim update_sqlstr As String

  'setting the recordset
  Set rs = New ADODB.Recordset
  Set rs1 = New ADODB.Recordset
  Set rs2 = New ADODB.Recordset
    SQLstr = "select * From test1_008_temp_final order by
    welltagnum,layerordernum"
  sqlstr1 = "select * from test1_008_temp_count"
```

```
sqlstr2 = "select * from test1_008_temp_Updatedelevation"
rs2.LockType = adLockOptimistic
rs2.CursorType = adOpenKeyset
rs.Open SQLstr, cnn
rs1.Open sqlstr1, cnn
rs2.Open sqlstr2, cnn
Do Until rs1.EOF
   v_cnt = rs1.Fields!Count
   'If rs.Fields!UpperBound Is Null Then
   v_ub = rs.Fields!UpperBound
   v_lon = 0
   For counter = 1 to v_cnt
      v_wtn = rs1.Fields!welltagnum
      v_liRcdID = rs.Fields!liRcdID
      v_lb = rs.Fields!LowerBound
      v_ld = rs.Fields!layerdepth
      v_lon = v_lon + 1
      v_description = rs.Fields!Description
      If rs.Fields!errcomment <> "" Then
         v_errcomment = rs.Fields!errcomment
      Else
        v_errcomment = Null
      End If
      'v_errcomment = rs.Fields!errcomment
      If rs.Fields!manual_correction <> "" Then
        v_manual_correction = rs.Fields!manual_correction
      Else

         v_manual_correction = Null
      End If
      'v_manual_correction = rs.Fields!manual_correction
      If rs.Fields!comment <> "" Then
         v_comment = rs.Fields!comment
      Else
         v_comment = Null
      End If
      If rs.Fields!utm_accuracy <> "" Then
         v_utm_accuracy = rs.Fields!utm_accuracy
      Else
         v_utm_accuracy = ""
      End If
      v_description1 = rs.Fields!description1
      'v_utm_accuracy = rs.Fields!utm_accuracy
      'v_comment = rs.Fields!comment
      vn_lb = v_ub - v_ld
      vn_ub = v_ub
      vn_ld = v_ld
      vn_lon = v_lon
      elevation = vn_lb
         rs2.AddNew
         rs2.Fields("liRcdID") = v_liRcdID
```

```
            rs2.Fields("welltagnum") = v_wtn
            rs2.Fields!UpperBound = v_ub
            rs2.Fields!LowerBound = vn_lb
            rs2.Fields!layerdepth = v_ld
            rs2.Fields!layerordernum = v_lon
            rs2.Fields!Description = v_description
            rs2.Fields!errcomment = v_errcomment
            rs2.Fields!manual_correction = v_manual_correction
            rs2.Fields!comment = v_comment
            rs2.Fields!description1 = v_description1
            rs2.Fields("utm_accuracy") = v_utm_accuracy
         rs2.Update
         rs2.MoveNext
         rs.MoveNext
         v_ub = elevation
      Next counter
      rs1.MoveNext
   Loop
End Sub
```

APPENDIX C: SEDIMENT CLASSIFICATION SYSTEMS AND HYDRAULIC PROPERTIES

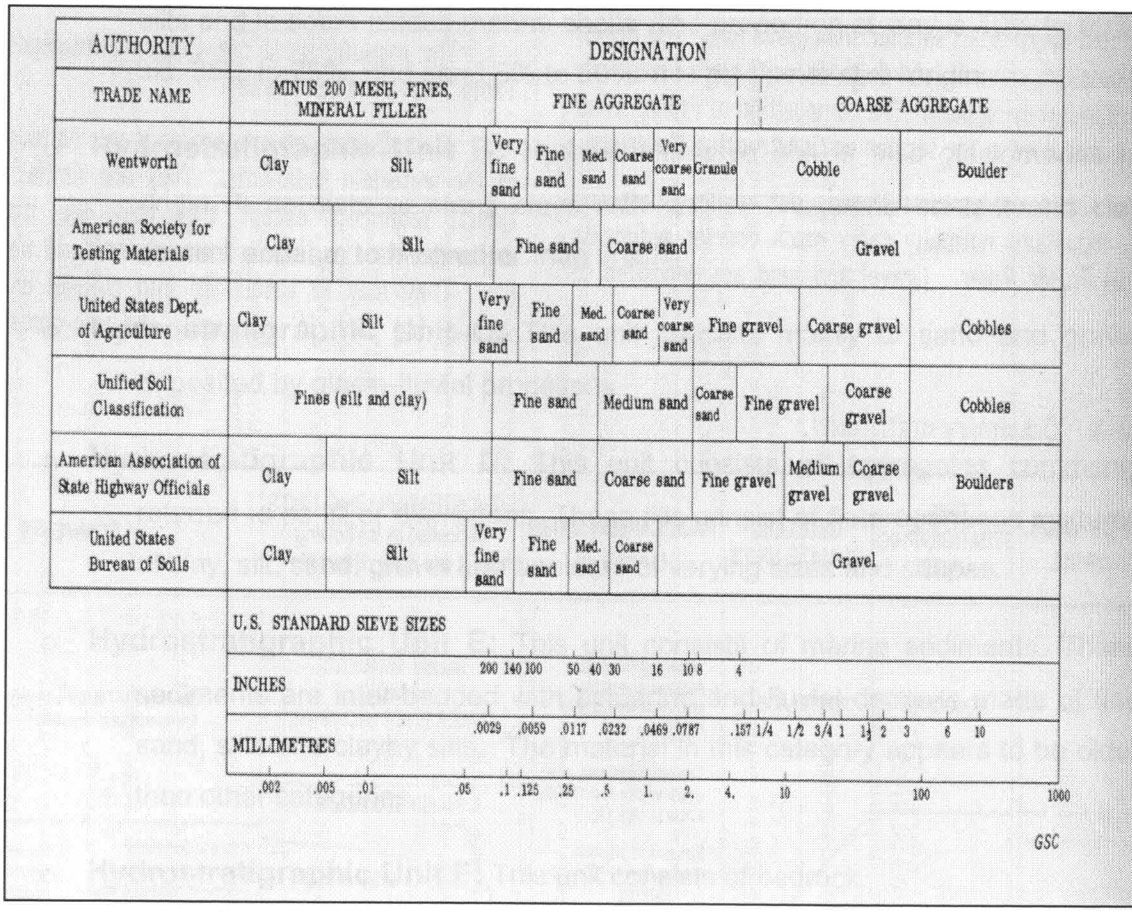


Figure C-1 Sediment Classification Systems[58]

[58]Reproduced with permission of the Minster of Public Works and Government Services Canada, 2004: Courtesy of Natural Resources Canada.

C1. Halstead's Hydrostratigraphic Classification (Source: Halstead, 1986)

(Refer to Halstead (1986) for a detailed classification).

- **Hydrostratigraphic Unit A:** Hydrostratigraphic Unit A consists of clay stony clay and silty clays with varying stone content, as well as silty lenses, sandy silts and in some places marine shells. The proportion of clay is 10% to 50%; silt, 35% to 75%, and sand 5% to 60%. It is glaciomarine in origin.
- **Hydrostratigraphic Unit B:** Hydrostratigraphic unit is also glaciomarine in origin. It consists of stony clays with shells. The stone content and clay content appears to be greater than unit A.
- **Hydrostratigraphic Unit C:** This unit consists mainly of sand and gravel deposited by glacio-fluvial processes.
- **Hydrostratigraphic Unit D:** This unit consists of aggregates commonly referred to as till or diamictons. These tills consist of heterogeneous mixtures of clay, silt, sand, gravel and boulders of varying sizes and shapes.
- **Hydrostratigraphic Unit E:** This unit consists of marine sediments. These sediments are inter-bedded with estuarine and fluvial deposits made of fine sand, silts and clayey silts. The material in this category appears to be older than other categories.
- **Hydrostratigraphic Unit F:** This unit consists of bedrock.

C2. Relative Hydraulic Properties

- **Porosity:** Porosity for earth material is defined as the percentage of the rock or soil that is void of material (Fetter, 1994).

Material	Porosity (%)
well sorted sand or gravel	25 – 50
sand & gravel, mixed	20 - 35
glacial till	10 - 20
silt	30 - 50
clay	33 - 60

Table C-1 Porosity Ranges for sediments (adapted from Fetter, 1994)

- **Specific Yield:** Specific Yield is defined as the ratio of the volume of water that drains from a saturated rock under the influence of gravity to the total volume of rock (Fetter, 1994).

Material	Average Specific Yield (%)
clay	2
sandy clay	7
silt	18
fine sand	21
medium sand	26
coarse sand	27
gravely sand	27
fine gravel	25
medium gravel	23
coarse gravel	22

Table C-2 Specific yield for sediments (adapted from Fetter, 2001)

- **Hydraulic Conductivity:** Hydraulic conductivity or permeability is the ability of a rock or sediment to transmit the flow of a fluid through it. It is also defined as the volume flow rate of water through a unit cross-sectional area of a porous medium under the influence of hydraulic gradient of unity, at a specified temperature. It is usually measured in units of m/s or m/day (Oxford University Press, 1999).

Material	**Hydraulic Conductivity (cm/s)**
Clay	$10^{-9} – 10^{-6}$
Silt, sandy silt, clayey sands and till	$10^{-6} – 10^{-4}$
Silty sands, fine sand	$10^{-5} – 10^{-3}$
Well sorted sands, glacial outwash	$10^{-3} – 10^{-1}$
Well sorted gravel	10^{-2} - 1

Table C-3 Hydraulic conductivity for sediments (adapted from Fetter, 2001)

APPENDIX D: CROSS-SECTIONS

Please refer to the CD-ROM for cross-sections A-A' through BB-B'B', page 135 for the cross-sectional plan and page136 for the legend.

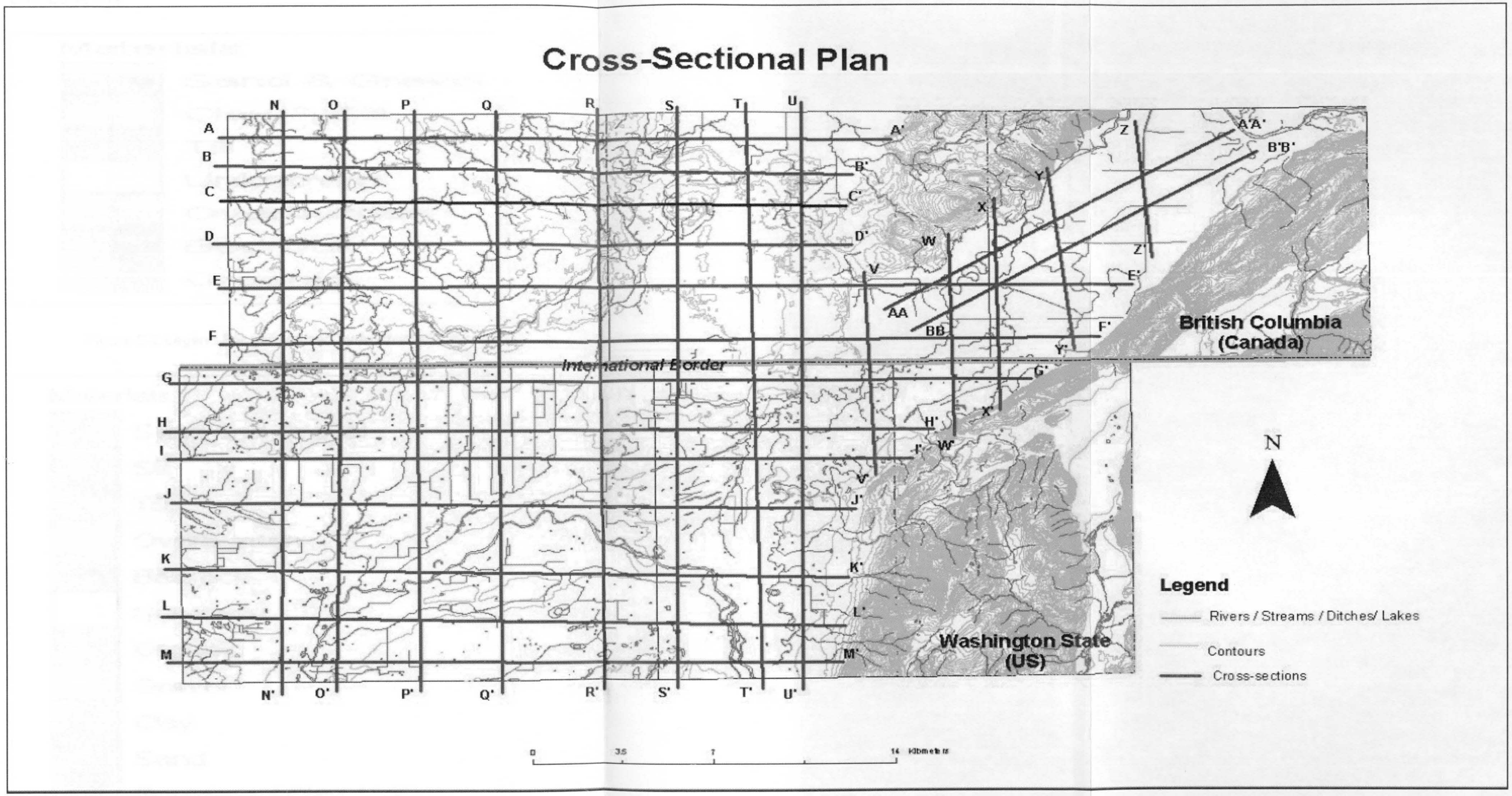


Table D-2 Cross-Sectional plan

D1. Legends

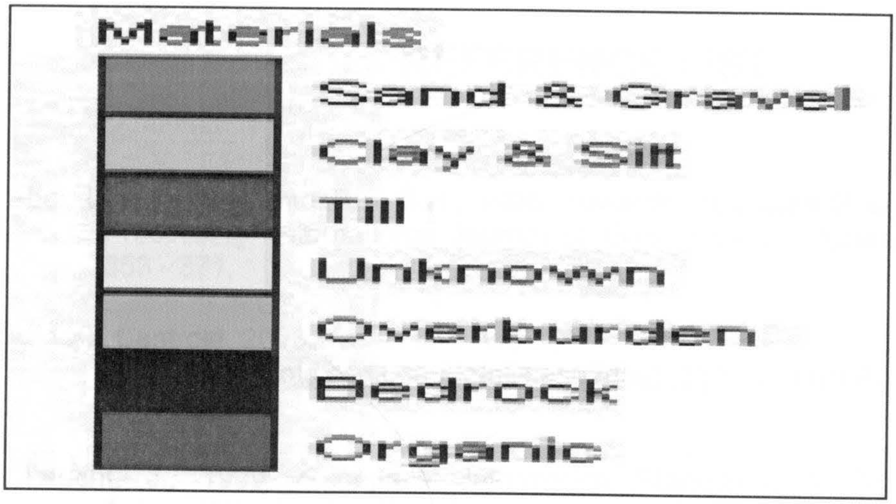


Figure D-2 Legend For generalized hydrostratigraphic model

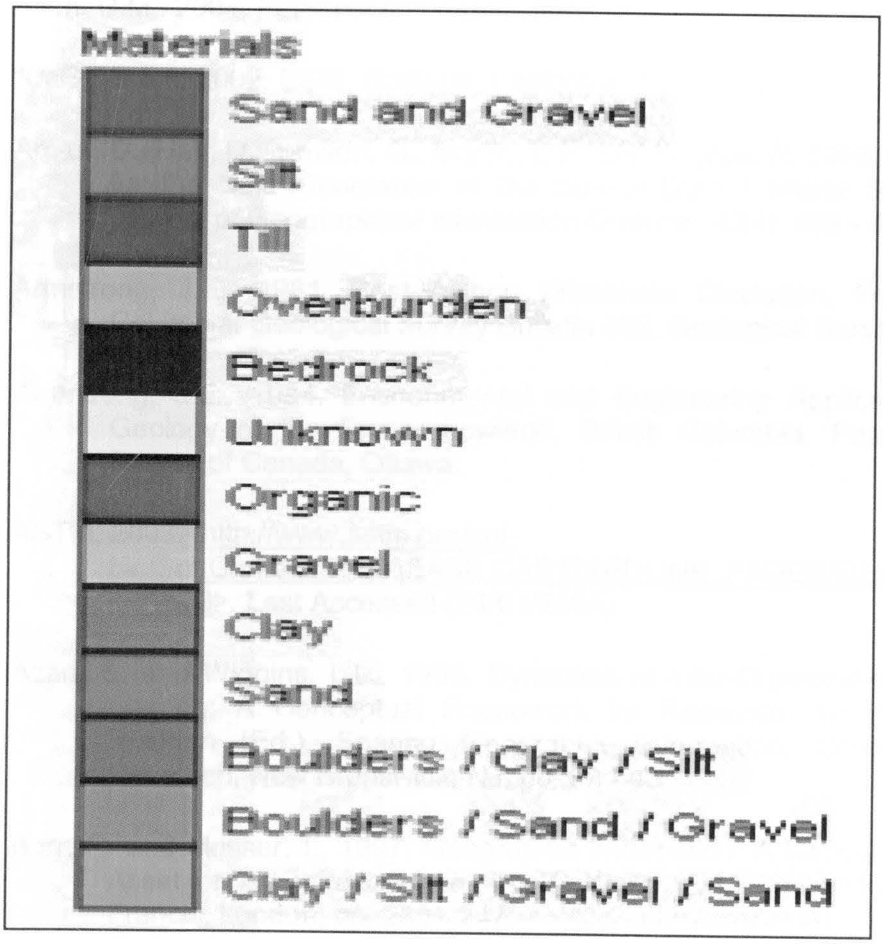


Figure D-3 Legend for lithologic units

REFERENCE LIST

Abel, D., Ooi, B.C. and Tan, S.H., 1998. Towards Integrated Geographical Information Processing. International Journal of Geographical Information Science, 12(4), 353 - 371.

Add XY Centroid, 2003. Trent Hare (USGS). <http://arcscripts.esri.com/details.asp?dbid=11978>. Last Accessed (14/03/2004).

Albrecht, J., 1999. Geospatial Information Standards. A Comparative Study of Approaches in the Standardization of Geospatial Information. Computers & Geoscience, 25, 9 - 24.

Allen, D.M., 2004. Personal communication.

ArcGIS® 8.3, 2002, ESRI, Redland, California.

Arctur, D., Hair, D., Timson, G., Martin, E.P. and Fegeas, R, 1998. Issues and Prospects for the Next Generation of the Spatial Data Transfer Standard. International Journal of Geographical Information Science, 12(4), 403 - 425.

Armstrong, J.E., 1981. Post-Vashon Wisconsin Glaciation, Fraser lowland, British Columbia. Geological Survey Bulletin 322, Geological Survey of Canada, Ottawa.

Armstrong, J.E., 1984. Environmental and Engineering Applications of the Surficial Geology of the Fraser Lowland, British Columbia. Paper 83-23, Geological Survey of Canada, Ottawa.

ASTM, 2003.<http://www.astm.org/cgi-bin/SoftCart.exe/DATABASE.CART/REDLINE_PAGES/D2487.htm?L+mystore+xjbq1693>. Last Accessed (24/01/2004).

Azad, B. and Wiggins, L.L., 1995. Dynamics of Inter-Organizational Geographic Data Sharing: A Conceptual Framework for Research. In: H. J. Onsrud and G. Rushton (Ed.), Sharing Geographic Information. Centre for Urban Policy Research, New Brunswick, NJ, pp. 22 - 43.

Barr, R. and Masser, I., 1997. Geographic Information: A Resource, A Commodity, An Asset or An Infrastructure. In: Z. Kemp (Ed.), Innovations in GIS. Taylor & Francis, London, pp. 234 - 247.

Batty, M., 1999. New Technology and GIS. In: M. Goodchild P. Longley, D. Maguire, and D. Rhind (Ed.), Geographical Information Systems. John Wiley & Sons, Inc., New York, pp. 309 - 316.

Berardinucci, J. and Ronneseth, K., 2002. Guide to Using the BC Aquifer Classification Maps for the Protection and Management of Groundwater, Ministry of Water, Land and Air Protection, Victoria. <http://wlapwww.gov.bc.ca/wat/aquifers/reports/aquifer_maps.pdf>. Last Accessed (14/03/2004).

Bishr, Y., 1998. Overcoming the Semantic and Other Barriers to GIS Interoperability. International Journal of Geographical Information Science, 12(4), 299 - 314.

Bishr, Y., Pundt, H., Kuhn, W. and Radwan, M., 1999. Probing the Concept of Information Communities - A First Step Towards Semantic Interoperability. In: M. Goodchild, M. Egenhofer, R. Fegeas, R. and C. Kottman (Ed.), Interoperating Geographic Systems. Kluwer Academic Publishers, Boston, pp. 39 - 54.

Buehler, K. and McKee, L. (Ed.), 1998. Introduction to Interoperable Geoprocessing and the OpenGIS Specification. Open GS Consortium, Inc., Wayland, Massachusetts.< www.geo.ulg.ac.be/interne/Fichiers_Interet_General/SIG/OpenGISGuide.pdf>. Last Accessed (20/12/2002).

Buehler, K. and Farley, J.A, 1994. Interoperability Of Geographic Data And Processes : The OGIS Approach. StandardView, 2(3).

Cameron, V.J., 1989. The Late Quaternary Geomorphic History of the Sumas Valley, Simon Fraser University, Burnaby.

Cassettari, S., 1993. Introduction to Integrated Geo - Information Management. Chapman and Hall, London.

Chan, T.O., Feeney, M., Rajabifard, A. and Williamson, L., 2001. The Dynamic Nature of Spatial Data Infrastructures: A Method of Descriptive Classification. Geomatica, 55(1), 60 - 72.

Clague, J.J., 1994. Quaternary Stratigraphy and History of South-Coastal British Columbia; In J.W.H. Monger (Ed.), Geology and Geological Hazards of the Vancouver Region, Southwestern British Columbia, Geological Survey of Canada, Bulletin 481, pp. 181 – 192.

Clarke, A.L., 1991, Australian Standards for Spatial Data transfer. In: H. Moellering (Ed.), Spatial Database Transfer Standards: Current International Status. Elsevier Science on behalf of the International Cartographic Association, London, pp. 28 - 36.

Coleman, D.J. and McLaughlin, J.D., 1998. Defining Global Data Infrastructures. Geomatica, 52(2), 129 - 144.

Cox, S.E. and Kahle, S.C., 1999. Hydrogeology, Ground-Water Quality, And Sources of Nitrate in Lowland Glacial Aquifers of Whatcom County, Washington, and British Columbia, Canada: U.S. Geological Survey Water-Resources Investigations Report, 98-4195, 5 plates, pp.251.

Cox, S., Daisy, P., Lake, R., Portele, C. and Whiteside, A, 2003. OpenGIS® Geography Markup Language (GML) Implementation Specification, Open GIS Consortium.

Craig, W.J. (Ed.), 1995. Why We Can't Share Data: Institutional Inertia. In: H.J. Rushton and G. Rushton (Ed.), Sharing Geographic Information. Center for Urban Policy Research, New Brunswick, NJ., pp. 107 - 117.

Department of Ecology, Washington State, 2004. < http://apps.ecy.wa.gov/welllog/>. Last Accessed (11/04/2004).

Devogele, T., Parent, C. and Spaccapietra, S., 1998. On Spatial Database Integration. International Journal of Geographical Information Science, 12(4), 335 - 352.

Dickin, R., 2002. Personal Communication.

Doremus, L., 2002. Personal Communication.

Environment Canada, 2003. Last Updated (08/08/2003). <http://www.ec.gc.ca/water/en/nature/grdwtr/e_intro.htm>. Last Accessed (12/12/2003).

Evangelatos, T.V. and Allam, M.M, 1991. Canadian Efforts to Develop Spatial Data exchange Standards. In: H. Moellering (Ed.), Spatial Database Transfer Standards: Current International Status. Elsevier Science on behalf of the International Cartographic Association, London, pp. 45 - 67.

Evans, J. and Ferreira Jr., J., 1995. Sharing Spatial Information in an Imperfect World: Interactions between Technical and Organizational Issues. In: H. J. Rushton G. Onsrud (Ed.), Sharing Geographic Information. Centre for Urban Policy Research, New Brunswick, NJ, pp. 448 - 460.

Federal Geographic data Committee (FGDC) Home Page. 2003. Last Updated 2/12/2003, FGDC,USGS <http://www.fgdc.gov/publications/a16final.html#appendixc>. Last Accessed (6/12/2003).

Fetter, C.W., 1994, Applied Hydrogeology. Prentice Hall, New Jersey.

Flowerdew, R., 1991. Spatial Data Integration. In: M. Goodchild D. Maguire, D. Rhind (Ed.), Geographic Information Systems: Principles and Applications. Longman Scientific &Technical, Essex, pp. 375 - 387.

FME (Feature Manipulation Engine) Suite 2003X2, 2003, Safe Software, Surrey, BC, Canada.

Fonseca, F.T., Egenhofer, M.J., Davis Jr., C.A. and Borges, K.A.V., 2000. Ontologies and Knowledge Sharing in Urban GIS. Computers, Environment and Urban Systems, 24, 251 - 271.

Frank, A.U., 1998, Metamodels for Data Quality Description. In: R. Jeansoulin and M.F. Goodchild (Ed.), Data Quality in Geographic Information - From Error to Uncertainty. Editions Hermès, Paris, pp. 15 - 29.

Frank, A.U. and Raubal, M., 1999. Formal Specification of Image Schemata - A Step Towards Interoperability in Geographic Information Systems. Spatial Cognition and Computation, 1, 67 - 101.

FSSD, 2002, Simon Fraser University, Dr. N.C. Schuurman (Dept. of Geography) and Dr. D.M. Allen (Dept. of Earth Sciences), Vancouver, BC, Canada.

Fulton, T., 2004, Personal Communication.

GeoConnection, 2003. Last Updated (9/09/2003) <http://cgdi.gc.ca/CGDI.cfm/fuseaction/home.welcome/lang/E/gcs.cfm>. Last Accessed (14/03/2004).

Geographic data Files (GDF) 2003. < http://www.ertico.com/links/gdf/gdf.htm>. Last Accessed 12/15/2003).

Goodchild, M.F., Egenhofer, M. and Fegeas, R., 1997. Interoperating GISs: Report of a Specialist Meeting Held under the Auspices of the Varenius Project, Panel on Computational Implementation of Geographic Concepts., National Centre for Geographic Information and Analysis, Santa Barbara.

Gill, P. (2002) Personal Communication.

GMS (Groundwater Modelling Software), 4.0, 2003, Brigham Yong University, Utah.

Gruber, T., 1995. Towards principles for the Design of Ontologies Used for Knowledge Sharing. Int. J. Human -Computer Studies, 43, 907 - 928.

Guarino, N., 1998. Formal Ontology and Information. Amended version of a paper appeared in N. Guarino (ed.), *Formal Ontology in Information Systems. Proceedings of FOIS'98, Trento, Italy, 6-8 June 1998.* Amsterdam, IOS Press, pp. 3-15. . <www.ladseb.pd.cnr.it/infor/Ontology/Papers/FOIS98.pdf>. Last Accessed 10/12/2003

Guptill, S.C., 1991. Spatial Data Exchange and Standardization. In: M. Goodchild D. Maguire, D. Rhind (Ed.), Geographic Information Systems: Principles and Applications. Longman Scientific &Technical, Essex, pp. 515 - 530.

Guptill, S.C., 1999. Metadata and Data Catalogues. In: M. Goodchild, P. Longley, D. Maguire and D. Rhind (Ed.), Geographical Information Systems: Management Issues and Application. John Wiley & Sons, Inc., New York, pp. 677 - 692.

Halstead, E.C., 1986. GroundWater Supply - Fraser Lowland, British Columbia. NHRI Paper No.26, IWD Scientific Series No. 145, National hydrology Research Institute, Saskatoon.

Harvey, F., Buttenfield, B. and Lambert, S.C., 1999a. Integrating Geodata Infrastructures from Ground Up. Photogrammetric Engineering and Remote Sensing, 11, 1287 - 1291.

Harvey, F., Kuhn, W., Pundt, H. and Bishr,Y., 1999b. Semantic Interoperability: A Central Issue for Sharing Geographic Information. Annals of Regional Science, 33, 213 - 232.

Hogan, R. and Sondhiem, M., 1996. Spatial Data Standards Activities in North America. In: H. Moellering (Ed.), Spatial Database Transfer Standards 2: Characteristics for assessing Standards for the Transfer of Spatial Data and Full Description of the National and International Standards in the World. Elsevier Science on behalf of the International Cartographic Association, Oxford, pp. 31 - 38.

Houlding, S.W., 2001. XML - An opportunity for < meaningful > Data Standards in the Geosciences. Computers and Geoscience, 27, 839 - 849.

ICIS, 2004 < ICIS, 2004) http://www.icisociety.ca >. Last Accessed (26/01/2004).

International Standardization Organization (ISO), 2003. Last Updated (12/12/2003) <http://www.isotc211.org/>. Last Accessed (6/12/2003).

Jones, M. A., 1999. Geologic Framework for the Puget Sound Aquifer System, Washington State and British Columbia. Professional paper 1424-C, US Geological Survey, Reston, Virginia.

Kahle, S.C., 1990. Hydrostratigraphic and Groundwater Flow in the Sumas Area, Whatcom County, Washington, Western Washington University, Bellingham.

Kashyap, V. and Sheth, A, 1996. Semantic and Schematic Similarities Between Database Objects: A Context - Based Approach. The VLDB Journal, 5, 276 - 304.

Kevany, M.J., 1995. A Proposed Structure for Observing Data Sharing. In: H. J. Rushton and G. Onsrud (Ed.), Sharing Geographic Information. Centre for Urban Policy Research, New Brunswick, NJ, pp. 76 - 100.

Kim, T.J., 1999. Metadata for Geo-Spatial Data Sharing. Annals of Regional Science, 33, 171 - 181.

Kim, W. and Seo, J., 1991. Classifying Schematic and Data Heterogeneities in Multidatabase Systems. IEEE, 24(12), 12 - 18.

Kleinrock, L., 1992. Technology issues in the Design of NREN. In: B. Kahin (Ed.), Building Information Infrastructure. McGraw-Hill, New York.

Klinkenberg, B., 2003. The True Cost of Spatial Data in Canada. The Canadian Cartographer, 47(1), 37 - 49.

Kohut, A.P., 1987. Groundwater Supply Capability Abbotsford Upland, Province of British Columbia, Ministry of Environment and Parks, Water Management Branch, Victoria, BC.

Kokla, M. and Kavouras, M., 2001. Fusion of Top-Level and Geographical Domain Ontologies Based on Context Formation and Complementarity. International Journal of Geographical Information Science, 15(7), 679 - 687.

Kottam, C.A., 1999. The Open GIS Consortium and Progress Towards Interoperability in GIS. In: M. Egenhofer M. Goodchild, R. Fegeas, and C. Kottam (Ed.), Interoperating Geographic Information Systems. Kluwer Academic Publishers, Boston, pp. 39 - 54.

Kuhn, W., 1994. Defining Semantics for Spatial data Transfers. In: T.C. Waugh and R.G. Healey (Ed.), Advances in GIS Research: Proceedings of the Sixth International Symposium on Spatial Data Handling. Dept. of Geoinformation, Tech University Vienna, Vienna.

Lake, R., 2001. GML 2.0 Enabling the Geospatial Web, Geospatial Solutions, pp. 38 - 41.

Laurini, R., 1998. Spatial Multi - Database Topological Continuity and Indexing: A Step Towards Seamless GIS Data Interoperability. International Journal of Geographical Information Science, 12(4).

Leake, S.A., Last Updated (6/01/2004). Land Subsidence from Groundwater Pumping. <http://geochange.er.usgs.gov/sw/changes/anthropogenic/subside/>. Last Accessed 16/01/2004).

Leclercq, E., Benslimane, D. and Yétongnon, K., 1999. Semantic Mediation For Cooperative Spatial Information Systems: The AMUN Data Model. Proceedings of the IEEE Forum on Research and Technology Advances in Digital Libraries.

Livingstone, E., 2002. Personal communication.

Maitra, n.d. The National Spatial Data Infrastructure in the United States: Standards, Metadata, Clearinghouse and Data Access, www.GISDevelopment.Net <http://www.gisdevelopment.net/policy/gii/gii0002pf.htm>. Last Accessed (14/03/2004).

Maguire, D.J., Goodchild. M.F. and Rhind, D.W, 1991, An Overview and Definitions of GIS. In: M. Goodchild D. Maguire, D. Rhind (Ed.), Geographic Information Systems: Principles and Applications. Longman Scientific &Technical, Essex, pp. 9 - 20.

Masser, I., 1998. "Governments and Geographic Information" Taylor and Francis, London.

Masser, I., 1999. All Shape and Sizes: the First Generation of National Spatial Data Infrastructure. International Journal of Geographical Information Science, 13(1), 67 - 84.

McKee, L., 1996. OGC: User - Mediated Technology Drives Vendor Opportunity. StandardView, 4(4), 192 - 197.

Microsoft® Access 2002, 2002, Microsoft Corporation, Redmond, Washington State.

Moellering, H., 1991, Approaches to Spatial Database Transfer Standards: Introduction:. In: H. Moellering (Ed.), Spatial Database Transfer Standards: Current International Status. Elsevier Science on behalf of the International Cartographic Association, London, pp. 1 - 27.

Moellering, H., 1996, Spatial Database Transfer Standards 2: Characteristics for assessing Standards for the Transfer of Spatial Data and Full Description of the National and International Standards in the World. Elsevier Science on behalf of the International Cartographic Association, Oxford.

Mohammed, D.A.M., 1997. International Standards and National Mapping Organizations. In: D. Rhind (Ed.), Framework for the World. Geoinformation International, Cambridge, pp. 206 - 210.

Mooney, D.J. and Grant, D.M., 1997. The Australian National Spatial Data Infrastructure. In: D. Rhind (Ed.), Framework for the World. Geoinformation International, Cambridge, pp. 187 201.

MSRM, 2004. Ministry of Sustainable Resource Management
< http://www.landdata.gov.bc.ca/>. Last Accessed (11/04/2004).

MWLAP, 2001. Water Well Registration Form, Last Updated (13/05/1997).
<http://wlapwww.gov.bc.ca/wat/waterbot/input/altform.html>. Last Accessed (24/01/2004).

MWLAP, 2002, Abbotsford-Sumas International Taskforce.
<http://wlapwww.gov.bc.ca/wat/aquifers/absumas.html>. Last Accessed (24/01/2004).

MWLAP, 2004, Ministry of Water, Land and Air Protection
< http://wlapwww.gov.bc.ca/wat/waterbot/gwell-out.html>. Last Accessed (11/04/2004).

National Spatial Data Infrastructure (NSDI), 2003. Last Updated 27/03/2003, FGDC, USGS < http://www.fgdc.gov/nsdi/nsdi.html>. Last Accessed (6/12/2003).

Nedovic-Budic, Z., Pinto, J.K, 2000. Information Sharing in an Inter-organizational GIS Environment. Environment and Planning B: Planning and Design, 27(3), 455 - 474.

Nedovic-Budic, Z., Pinto, J.K., 1999. Interorganization GIS: Issues and Prospects. The Annals of Regional Science, 33, 183 - 195.

Nedovic-Budic, Z., Pinto, J.K., 2001. Organizational (soft) GIS interoperability: lessons from the US JAG, 3(3).

O'Donnell, J.H. and Penton, C.R., 1997. Canadian Perspective on the Future of National Mapping Organization. In: D. Rhind (Ed.), Framework for the World. Geoinformation International, Cambridge, pp. 214 - 225.

Office of Management and Budgets (OMB), 2004, <http://www.whitehouse.gov/omb/circulars/a016/a016_rev.html#appc>. Last Accessed (14/03/2004).

Onsrud, H. J. and Rushton, G. (Ed.). 1995. Sharing Geographic Information. Center for Urban Policy Research, New Brunswick, NJ.

Open GIS consortium (OGC) Home Page, 2003a. Open GIS consortium, Web Feature Service < http://www.opengis.org/specs/?page=specs >. Last Accessed (2/06/2003).

Open GIS consortium (OGC) Home Page, 2003b. Open GIS consortium <http://www.opengis.org/about/?page>. Last Accessed (12/06/2003).

Open GIS consortium (OGC) Home Page, 2003c, Open GIS consortium <http://www.opengis.org/resources/?page=testing>. Last Accessed (12/06/2003).

Open GIS consortium (OGC) Home Page, 2003d. Open GIS consortium, Topic 11 Metadata < http://www.opengis.org/specs/?page=abstract >. Last Accessed (12/06/2003).

Oxford University Press, 1999. "permeability" A Dictionary of Earth Sciences. Ed. Ailsa Allaby and Michael Allaby. Oxford University Press, 1999. Oxford Reference Online. Oxford University Press. Simon Fraser University. 15 March 2004 http://www.oxfordreference.com/views/ENTRY.html?subview=Main&entry=t13.e6283. Last Accessed (15/03/2004).

Pinto, J. and Rushton, G., 1995. Sharing Geographic Information across Organizational boundaries: A Research framework. In: H. J. Rushton and G. Onsrud, (Ed.), Sharing Geographic Information. Center for Urban Policy Research, New Brunswick, NJ, pp. 44 - 64.

Piteau Associates and Turner Groundwater Consultants, 1993. Groundwater Mapping and Assessment in British Columbia Volume 1: Review and Recommendations, Environment Canada and Resource Inventory Committee (Earth Science Task Force), North Vancouver, British Columbia.

Pitt, N. (Ed.), 2001. XML Black Book. The Coriolis Group, LLC, Scottsdale, Az.

Poire, M., 2004. Personal Communication.

Quakenbash, P., 2002, Personal Communication.

Reed, C., 2003, personal communication.

Regier, R., 2002, Personal Communication.

Rhind, D.W., 1999. National and International Geospatial data Policies. In: M. Goodchild P. Longley, D. Maguire and D. Rhind (Ed.), Geographical Information Systems: Management Issues and Application. John Wiley & Sons, Inc., New York, pp. 767 - 787.

Ricketts, B.D., 1997. The Fraser Lowland Hydrogeology Project: an Overview. Open File D3828, Geological Survey of Canada, Vancouver, British Columbia.

Rivera, A., Crowe, A., Kohut, A.L., Rudolph, D., Baker, C., Pupek, D., Shaheen, N., Lewis, M. and Parks, K., 2003. Canadian Framework for Collaboration on Groundwater Collaboration, Natural Resources Canada, Quebec. <http://www.cgq-qgc.ca/cgsi/index.cfm?flag=0&CFID=65781&CFTOKEN=50816061>. Last Accessed 24/01/2004).

Robertson, W.A. and Gartner, C., 1997. The Reform of national Mapping Organizations: The New Zealand Experience. In: D. Rhind (Ed.), Framework for the World. Geoinformation International, Cambridge, pp. 247 - 264.

Rosssmeissl, H.J. and Rugg, R.D., 1991. The United States Spatial Data Transfer Standard. In: H. Moellering (Ed.), Spatial Database Transfer Standards: Current International Status. Elsevier Science on behalf of the International Cartographic Association, London, pp. 204 - 216.

Russell, H.A.J., Brennand, T.A., Logan, C. and Sharpe, D.R., 1998. Standardization and Assessment of Geological Descriptions from Water well records: Greater Toronto and Oak Ridges Moraine Areas, Southern Ontario. Current Research 1998-E, Geological Survey of Canada, Ottawa.

Salgé, F., 1997, International Standards and National Mapping Organizations. In: D. Rhind (Ed.), Framework for the World. Geoinformation International, Cambridge, pp. 160 -172.

Salgé, F., 1999, National and International Data Standards. In: M. Goodchild, P. Longley, D. Maguire and D. Rhind (Ed.), Geographical Information Systems: Management Issues and Application. John Wiley & Sons, Inc., New York, pp. 693 - 706.

Schuurman, N., 2002. Flexible Standardization: Making Interoperability Accessible to Agencies with Limited Resources. Cartography and Geographic Information Science, 29(4), 343 - 353.

SDTS Home Page 2002. Last Updated (17/02/2004). <http://mcmcweb.er.usgs.gov/sdts/>. Last Accessed (14/03/2004).

SDTS Home Page 2003. Last Updateds14 /11/2003, Us Department of Interior / US Geological Survey <http://mcmcweb.er.usgs.gov/sdts/>. Last Accessed (10/20/2003).

Sears, G., 2001. Geospatial Data Policy, KPMG Consulting: Prepared for GeoConnections Policy Advisory Node, Ottawa.

Sharpe, D., n.d. Oak Ridge Moraine, Project overview, Geological Survey of Canada, Ottawa. < http: //sts.gsc.nrcan.gc.ca/page1/envir/orm/orm.htm.> Last Accessed 24/01/2004).

Shepherd, I.D.H., 1991. Information Integration and GIS. In: M. Goodchild D. Maguire, D. Rhind (Ed.), Geographic Information Systems: Principles and Applications. Longman Scientific &Technical, Essex, pp. 337 - 360.

Sheth, A. and Larson, J.A., 1990. Federated Database Systems for Managing Distributed, Heterogeneous, and Autonomous Databases. ACM Computing Surveys (CSUR), 22(3), 183 - 236.

Sheth, A.P., 1999. Changing the Focus on Interoperability in Information Systems: from System, Syntax, Structure to Semantics. In: M. Egenhofer M. Goodchild, R. Fegeas, and C. Kottman (Ed.), Interoperating Geographic Information Systems. Kluwer Academic Publishers, Boston, pp. 5 - 29.

Sieber, R.E., 2003. Public Participation Geographic Information Systems across Borders. The Canadian Cartographer, 47(1), 50 - 61.

Smith, I., 1991. Digital Geographic Information working Group: Exchange Standards. In: H. Moellering (Ed.), Spatial Database Transfer Standards: Current International Status. Elsevier Science on behalf of the International Cartographic Association, London, pp. 223 - 235.

Smith, B. and Mark, D.M., 2001. Geographical Categories: an Ontological Investigation. International Journal of Geographical Information Science, 15(7), 591- 612.

Sondheim, M., Gardels, K. and Buehler, K., 1999. GIS Interoperability. In: M. Goodchild P. Longley, D. Maguire, and D. Rhind (Ed.), Geographical Information Systems. John Wiley & Sons, Inc., New York, pp. 359 -369.

Sowton, M., 1991. The National Transfer Format. In: H. Moellering (Ed.), Spatial Database Transfer Standards: Current International Status. Elsevier Science on behalf of the International Cartographic Association, London, pp. 191 - 203.

Sp3dPntzval, 1998. ESRI, Redlands, California <http://arcscripts.esri.com/details.asp?dbid=10957>. Last Accessed (14/03/2004).

Stasney, D.E., 2000. Hydrostratigraphy, Groundwater Flow and Nitrate Transport with the Abbotsford-Sumas Aquifer, Whatcom County, Washington, Western Washington University, Bellingham.

Stock, K and Pullar, D, 1999. Identifying Semantically Similar Elements in Heterogeneous Spatial Databases Using Predicate Logic. In: K. Brassel and H. Schek A. Vckovski (Ed.), Interoperating Geographic Information Systems: Second International Conference, INTEROP'99 Zurich, Switzerland, 1999. Springer - Verlag, Berlin, pp. 231 - 252.

Strand, E.J., Mehta, R.P. and Jairam, R., 1994. Applications Thrive on Open Systems Standards. StandardView, 2(3), 148 - 154.

Strong, D.M., Lee, Y and Wang, R., 1997. Data Quality in Context. Communications of the ACM, 40(5), 103 - 110.

Symington, R., 2002. Personal Communication.

Taylor, D.R.F., 1996. Presidential foreword. In: H. Moellering (Ed.), Spatial Database Transfer Standards 2: Characteristics for assessing Standards for the Transfer of Spatial Data and Full Description of the National and International Standards in the World. Elsevier Science on behalf of the International Cartographic Association, Oxford.

Tosta, N., 1994. Standards to Support the National Spatial Data Infrastructure. StandardView, 2(3), 143 - 147.

Tosta, N., 1997. National Spatial Data Infrastructures and the Roles of National Mapping Organizations. In: D. Rhind (Ed.), Framework of the World. Geoinformation International, pp. 173 - 186.

Tsou, M-H and Buttenfield, B.P., 2002. A Dynamic Architecture for Distributing Geographic Information Services. Transactions in GIS, 6(4), 355 - 381.

Uschold, M. and Gruniger, M., 1996. Ontologies: Principles, Methods and Applications. Knowledge Engineering Review, 11(2), 93 - 155.

USGS, 2004 United States Geological Survey <http://edc.usgs.gov/>. Last Accessed (11/04/2004).

Vckovski, A., 1998. Special Issue: Interoperability in GIS (Guest Editorial). International Journal of Geographical Information Science, 12(4), 297 - 298.

Vckovski, A., 1999. Interoperability and Spatial Information Theory. In: M. Egenhofer M. Goodchild, R. Fegeas, and C. Kottman (Ed.), Interoperating Geographic Information Systems. Kluwer Academic Publisher, Boston, pp. 31 - 37.

Visser, U., Stuckenschmidt, H. and Schlieder, C, 2002a. Interoperability in GIS – Enabling Technologies, 5th AGILE Conference on Geographic Information Science, Palma (Balearic Islands, Spain).

Visser, U., Stuckenschmidt, H., Schuster,G. and Vogele, T., 2002b. Ontologies for Geographic Information Processing. Computers & Geoscience, 28, 103 - 117.

Voisard, A. and Jurgens, M., 1999. Geospatial Information Extraction: Querying or Quarrying. In: M. Egenhofer M. Goodchild, R. Fegeas, and C. Kottman (Ed.), Interoperating Geographic Information Systems. Kluwer Academic Publisher, Boston, pp. 149 - 164.

Webopedia, 2004. Last Updated (17/03/2003) <http://www.webopedia.com/TERM/I/interface.html>. Last Accessed (14/03/2004).

Whitten, D.G.A. and Brooks, J.R.V., 1985, Dictionary of Geology. Penguin Books Ltd., Harmondsworth, Middlesex, England.

Widom, J., 1995. Integrating Heterogeneous Databases: Lazy or Eager. ACM Computing Surveys, 28(4).

Winter, S., 2001. Ontology: Buzzword or Paradigm Shift in GI Science. International Journal of Geographical Information Science, 15(7), 587 - 590.

World Wide Web Consortium (W3C), 2003. Last Updated (11/12/2003), <http://www.w3.org/XML/>. Last Accessed (6/12/2003).

Wortman, K., 1992. The Spatial Data Transfer Standard (FIPS 173): A Management Perspective. Cartography and Geographic Information Science, 19(5).

WRIA, 2002. Last Updated (2/12/2003). <http://apps.ecy.wa.gov/welllog/textsearch.asp>. Last Accessed (24/01/2004).