DEVELOPMENT OF AN INTEROPERABLE GEOGRAPHIC INFORMATION SYSTEM PLATFORM FOR TRANSPORTATION APPLICATIONS

by

Kai Han

A Thesis
Submitted to the Faculty of Graduate Studies
in Partial Fulfillment of the Requirements
for the Degree of

MASTER OF SCIENCE

Department of Civil and Geological Engineering
University of Manitoba
Winnipeg, Manitoba
Canada

0-612-62747-0

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES

*****

COPYRIGHT PERMISSION

DEVELOPMENT OF AN INTEROPERABLE GEOGRAPHIC INFORMATION SYSTEM PLATFORM FOR TRANSPORTATION APPLICATIONS

BY

KAI HAN

A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University of Manitoba in partial fulfillment of the requirement of the degree

of

MASTER OF SCIENCE

ABSTRACT

Interoperability and data sharing are important issues in the Geographic Information Systems (GIS) arena. The thesis examines the interoperability of GIS, establishes a theoretical framework together with associated techniques to facilitate GIS data sharing, and creates an interoperable GIS platform for transportation applications in Manitoba by applying the framework and techniques.

The thesis has constructed the theoretical framework based on the need to share GIS data to facilitate transportation analysis and planning. Techniques to fulfill data sharing tasks are developed to improve interoperability at the spatial data level. The established interoperable platform supports transportation applications through more accurate, current, and interoperable GIS systems. The research enhances GIS interoperability between different data sources.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor, Professor Alan Clayton, for his wisdom, insight, patience, assistance, and unfailing encouragement to complete this thesis. I also wish to thank Dr. Attahiru Alfa, Dr. Ahmed Shalaby, and Mr. Grant Forsman for their interest and suggestions.

Thanks are also due to my student colleagues and the staff of the University of Manitoba Transport Information Group--Jeannette Montufar, Errol Tan, Kean Lew, David Tang, Scott Minty, and Ed Lemke for their encouragement and friendship which made my work in the group a pleasure.

I want to extend my appreciation to the Manitoba Transportation and Government Services and the City of Winnipeg Public Works Department for their assistance in this research project.

A special thanks is extended to Mr. Ian Clayton.

Finally, I would like to thank my wife, Shouchen Sun, and my parents for their love, support, and understanding throughout the course of this research.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1.0 INTRODUCTION

1.1 Purpose

The thesis examines the interoperability of Geographic Information Systems (GIS), establishes a theoretical framework with associated techniques to facilitate GIS data sharing, and creates an interoperable GIS platform for transportation applications in Manitoba. With improved interoperability at the spatial data level, the platform supports transportation applications through more accurate, current, and interoperable GIS systems.

1.2 Background

With their ability to conduct spatial analysis and visually represent information, GIS have evolved as a powerful tool for comprehensive analysis and decision-making. They have been widely adopted over the past decade in planning, forestry, agriculture, transportation, and many other fields (Goodchild et al., 1997).

GIS combines electronic maps with traditional databases, and makes visual representations of information. Visualization helps people approach problems in the dimensions of space and time with intuitive maps instead of dimensionally-restricted data tables and graphs.

Transportation involves the movement of people and goods. The spatial and temporal dimensions of transportation systems make GIS a useful tool to analyze transportation problems. Transportation analyses involve multimodal, multisectoral, multiproblem, and multidisciplinary studies focusing on the interaction between the transportation and activity systems of a region (Manheim, 1979). Realizing the advantages of GIS spatial analyzing, network analyzing, data integrating, and information visualizing abilities, transportation is one of the earliest application fields for GIS (George, 1991). Many GIS for Transportation (GIS-T) applications have been developed to collect, store, manage, analyze, and distribute information.

Benefits of GIS-T Data Sharing

Transportation systems continuously interact with the world they serve (Manheim, 1979). Information integrated by GIS from multiple data sources can provide better understanding and deeper insight into transportation issues. However, because it can be tedious, time-consuming, and expensive to develop GIS spatial data and associate attribute data, GIS-T data sharing is desirable.

Good examples of GIS-T data sharing are the creation of an integral GIS-T platform from individual provincial GIS for the Winter Weight Premiums and Spring Weight Restrictions study for the Prairie Provinces (Montufar et al., 2000) and the development of the Manitoba Capital Region transportation analysis GIS network through the combination of the city street network with the provincial highway network (Lew, 2000). However, the full potential of GIS-T data sharing has yet to be realized, mainly because of the lack of interoperability between existing GIS systems.

GIS Interoperability

GIS interoperability refers to components of one GIS that are able to function in different environments (OGC, 2001). For example, interoperability occurs when GIS are able to share data and other resources. Interoperability is an increasingly important topic in GIS research. There have been numerous efforts to promote interoperability among GIS to overcome problems concerning GIS software, computer platforms, and data format issues (NCGIA, 1997), (OGC, 2001). Among these issues, GIS data related problems are fundamental to GIS interoperability.

Transportation Applications and Interoperability

Transportation applications of GIS have unique characteristics, which makes GIS-T an interesting and demanding field for implementing interoperability. GIS-T applications such as networking, routing, and linear referencing make diverse demands on GIS spatial data. To accommodate these applications, spatial data structures of GIS-T systems vary significantly. Therefore, it is necessary to handle interoperability issues right from the spatial data level to accomplish data sharing among GIS-T systems.

Major interoperability issues concerning GIS-T spatial data are: accuracy, currency, level of generalization, spatial data structure, networking/routing capabilities, and linear referencing/dynamic segmentation. The complexity of these issues has hindered the development and implementation of more advanced transportation applications of GIS.

Current GIS-T systems are limited in their ability to share data from different data sources. For example, a regional analysis covering different jurisdictions can be difficult to conduct due to the lack of a common, uniform platform to integrate information from individual GIS.

Efforts have been made to overcome these problems (Noronha, 1997), (Beskpalko et al., 1997), (Xiong et al., 1999), (Dueker and Butler, 2000), (Han et al., 2000). Often, individual methods and techniques have been developed to solve problems encountered in the course of specific projects. To achieve a high level of interoperability, this research focuses on interoperability issues and uses systematic approaches to solve principal GIS-T data sharing problems.

1.3 Objectives and Scope

Objectives

The thesis has two objectives:

To create a framework and develop techniques to facilitate data sharing and interoperability in GIS-T applications.

The framework addresses transportation-related data sharing issues with a focus on interoperability at the spatial data level. The techniques developed are based on real-world transportation projects.

To apply the framework and techniques developed to create a GIS-T platform for Manitoba with improved interoperability, accuracy, and currency at the spatial data level.

Spatial data from various public and private organizations are integrated to make a GIS-T platform for Manitoba. Using this platform, intra-agency, inter-agency and inter-jurisdictional planning and analysis can be conducted. Existing GIS-T can also be enhanced by adopting higher quality data without jeopardizing compatibility between spatial data and current systems.

Scope

The thesis is limited to interoperability issues regarding GIS platforms for transportation applications, with a geographical coverage of the province of Manitoba and the City of Winnipeg. Connections to adjacent jurisdictions are also included. While the framework and techniques are developed in a specific transportation context, their potential is not limited to either this geographical realm, or the transportation realm itself. A wider implementation of the framework and techniques will benefit the entire GIS community.

1.4 Approach and Organization

Approach

The thesis addresses interoperability issues from a GIS-T user perspective. GIS-T applications are discussed, and the requirements they impose on GIS platforms are covered in detail. A conceptual framework is developed to provide guidance for the GIS-T data sharing process. Techniques automating data sharing process are developed to support the interoperability framework.

Engineering judgments and practices are utilized throughout the research. Practical experience with GIS-T applications and their platforms are included to provide valuable insights on problems in the real world and to facilitate solutions. The following transportation research projects lay the foundation for the thesis:

- Development and Enhancement of a New Manitoba Transportation and Government Services (MT&GS) GIS-T Base Map
- Manitoba Highway Traffic Information System (UMTIG, 2000)
- Development of a New Transportation Planning Model for the Manitoba Capital Region (Lew, 2000)
- Harmonization of Spring Weight Restrictions & Winter Premiums for Roads in the Prairie Region (Montufar et al., 2000)

- Application of GIS to the analysis of truck accidents in Manitoba & Saskatchewan (Han et al., 2000), (Montufar and Sanderson, 1999)
- Development of a Road Network Management System for Manitoba Rural Municipalities (Lew et al., 1999)
- Development of the City of Winnipeg Traffic Information System (Alam et al., 1998)

To establish an interoperable GIS-T platform, spatial data from various sources are employed in the research. The data sources used for the research are as follows:

- Three versions of the Manitoba provincial highway network map
- The City of Winnipeg street centerline (SCL) map, traffic zone, map and land use map
- Canadian Pacific and Canadian National Railway map in Winnipeg region developed by University of Manitoba Transport Information Group (UMTIG)
- Rural Municipalities (R.M.) of Cartier, Grey, North Norfolk, Portage la Prairie, South Norfolk, and Victoria municipal road network maps developed by UMTIG
- Digital map fabric and aerial photos from Natural Resources Canada (NRC)
- Engineering drawings of various road construction projects provided by the City of Winnipeg and MT&GS
- Aerial photos of Winnipeg and vicinity provided by City of Winnipeg and R.M. of Cartier
- Satellite Imagery of R.M. of South Norfolk provided by Prairie Farm Rehabilitation Administration (PFRA)
- GPS surveys in Winnipeg and southern Manitoba conducted by University of Manitoba students and UMTIG staff
- GIS-T base maps from various sources including the provinces of Alberta and Saskatchewan, the cities of Edmonton, Saskatoon, and Regina, and the states of Minnesota, North Dakota, and Montana regarding jurisdictional winter weigh premiums (WWP) and spring weight restrictions (SWR)

Organization

The thesis consists of six chapters. They are:

- Chapter 1 defines the thesis and gives the background of the research.
- Chapter 2 provides the technical background of the research regarding GIS and spatial data.
- Chapter 3 discusses the importance of GIS-T, and issues related to data sharing.
- Chapter 4 describes the development of a framework to facilitate GIS-T data sharing.
- Chapter 5 gives technical details of implementing the framework in the process of building an interoperable GIS-T platform for Manitoba.
- Chapter 6 provides the conclusion and recommendations.

1.5 Thesis Concepts and Terminology

This section provides a basic understanding of the thesis concepts. A more detailed discussion of GIS and GIS-T is given in Chapters 2, 3, and 5.

GIS – Geographic Information Systems. GIS are computerized information systems integrating tabular attribute data with spatial-referenced geographical data to facilitate information management, mapping, spatial analysis, and decision-making.

GIS-T – Geographic Information Systems for Transportation. It emphasizes infrastructure management, network planning, traffic flow analysis, safety analysis, and information distribution. The GIS platforms for transportation applications are most commonly vector-based, which is more suitable for network concepts and analysis. Specialized algorithms such as linear referencing and connectivity checking are often required for GIS-T.

Spatial data – One of the two key components of GIS data. Spatial data facilitates the visual representation of the real world. Spatial data are points, lines, and polygons, or pixels, with location information attached to them, representing real-life features. For example, a bus stop can be represented as a point with longitude and latitude in GIS, city streets can be represented as lines following the street centerline, and traffic zones can be represented as separate polygons with shapes corresponding to their boundaries.

Attribute data – The second of the two key components of GIS data. It reflects the properties of the features and may contain any type of information. Attribute information is usually stored in databases and associated with spatial data features through unique identifiers provided by GIS.

Digitizing – One of the processes of creating GIS spatial data, i.e., electronic maps. Using a digitizer and/or specialized computer software, location information can be obtained manually or automatically from various sources, such as existing paper maps, ortho-corrected aerial photos and satellite imagery.

Map projection – Methodology used to represent the curved surface of the earth on a flat paper or computer screen. There are many projections established for various purposes and each of them maintains a certain amount of distortion. Albers, Lambert, and Transverse Mercator are commonly used projections (Anderson and Mikhail, 1998).

Coordinate system – Reference system to help locate features on the Earth's surface. It can take the form of longitude and latitude, or projected coordinates (e.g., Universal Transverse Mercator (UTM) and U.S. State Plane), which is particularly suitable for a smaller region. Both of them provide location information relative to other geo-spatial features (Heywood et al., 1998).

Linear referencing/dynamic segmentation – Techniques commonly used in transportation GIS applications to dis-aggregate sections of linear features (e.g., a section of highway) through one-dimensional offset referencing. Since the difference between the two techniques is small, they are often used interchangeably. These techniques are performed dynamically by GIS-T to save computer storage and gain flexibility.

GPS – Global Positioning System. GPS uses satellite signals to help position an object on the Earth's surface. It was developed and operated by the U.S. Military, and comprises 24-plus satellites and ground control. Users need receivers to pick up satellite signals and calculate their positions through triangulation. The accuracy of GPS can be improved through differential GPS (DGPS) and other techniques and technologies (Wyman, 2000).

2.0 Basics Concepts of Geographic Information Systems

This chapter discusses the basic concepts, general functions, and components of Geographical Information Systems (GIS) to provide a better understanding of the thesis topic and the research process. Commonly used GIS software is also discussed.

2.1 Defining GIS

GIS are defined in a variety of ways due to their wide range of applications (ESRI, 2001), (Heywood et al., 1998), (Intergraph, 1998b). The common characteristics of GIS can be summarized as follows:

- GIS is a computer-based system, including computer hardware, software, data, and users
- GIS deals with location-related information
- GIS can be used to collect, store, and analyze data, and disseminate information

In short, GIS is a computer information system using spatial-referenced geographical data to combine electronic maps with conventional databases and produce visual representations of various types of information.

2.1.1 Defining GIS as an Information System

Like other computer information systems, GIS comes in a combination of computer hardware and software. GIS are built on basic computer hardware, which includes computer server, workstation computers, digitizer, and plotter. On top of the hardware, GIS is supported by a computer operating system which takes care of routine processes such as file saving and network communication. Through the operating system, GIS manages databases, storing spatial data and attribute data with the support of a spatial sub-system and database management sub-system. The result of GIS analyses is presented to users through a Graphical User Interface (GUI). Figure 2.1 shows the basic physical structure of a typical GIS.

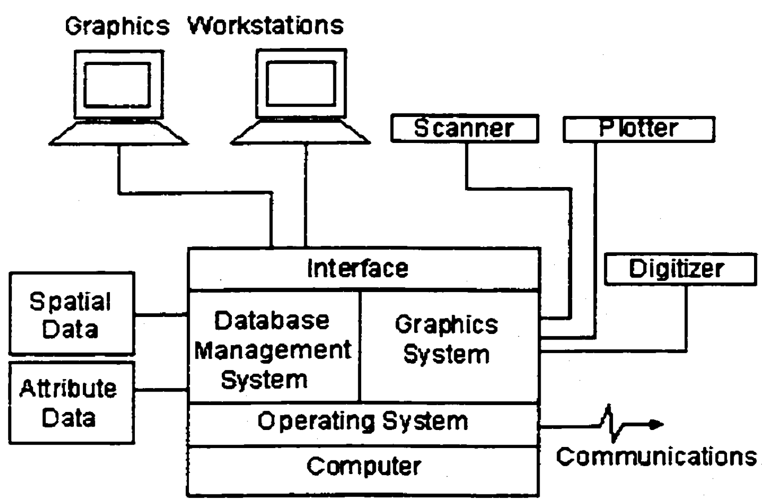
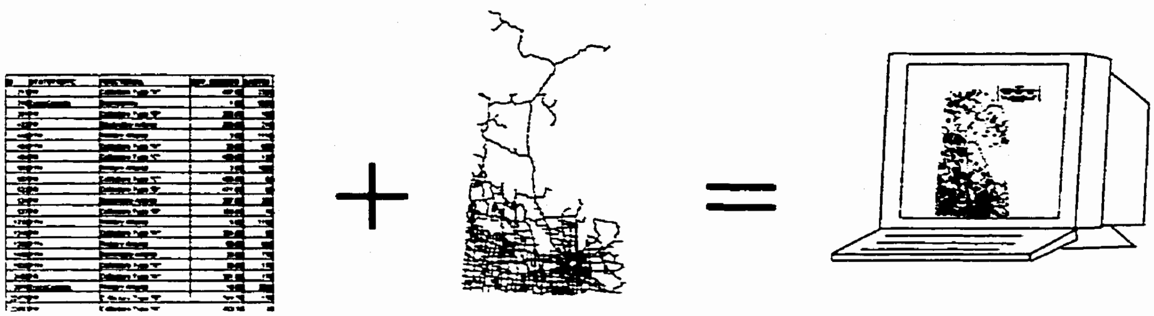


Figure 2.1 Basic Physical Structure of a GIS (AGI, 1999)

2.1.2 Defining GIS as a Geographic System

GIS support the acquisition, management, analysis, and visualization of spatial-referenced data. To render data spatially referenced, GIS integrate location information with conventional tabular data to create "mappable" data sets organized by spatial coordinates. This is illustrated in Figure 2.2.



Tabular Data Location Information GIS

Figure 2.2 The Integration of Tabular Data and Location Information

2.2 GIS Data Components

In order to perform analytical and management functions, GIS require two sets of data--spatial data and attribute data. Spatial data is data about location. Attribute data is data about property. In simple words, spatial data is "x, y, z", and attribute data is "A, B, C".

2.2.1 Spatial Data

Spatial data is a set of geographic features that are used to represent the real world. Spatial data is also referred to as "base map" and "map database" in GIS terminology. Contained in the map databases, there are spatial data features coupled with coordinate information to correctly represent the real-world entities. With coordinate information, spatial data can be visualized on a computer screen as a map.

In order to enable and facilitate computerized data handling, spatial data are structured with the following characteristics:

<u>Spatial Data Layers</u>

To facilitate managing spatial data features, GIS adopts layering technology (Heywood et al., 1998), which is commonly used in map making and the CAD field. Spatial features are grouped by feature type, and by the real-world entities they represent. Each layer is handled by GIS separately, and has its own set of spatial and attribute data. However, these separate layers can be and often are stacked together when the same coordinate/projection systems are used. This layering technology also allows GIS users to perform cross-layer analysis, which is a powerful tool for comprehensive analysis and difficult to conduct without GIS. Figure 2.3 shows an example of cross-layer analysis.

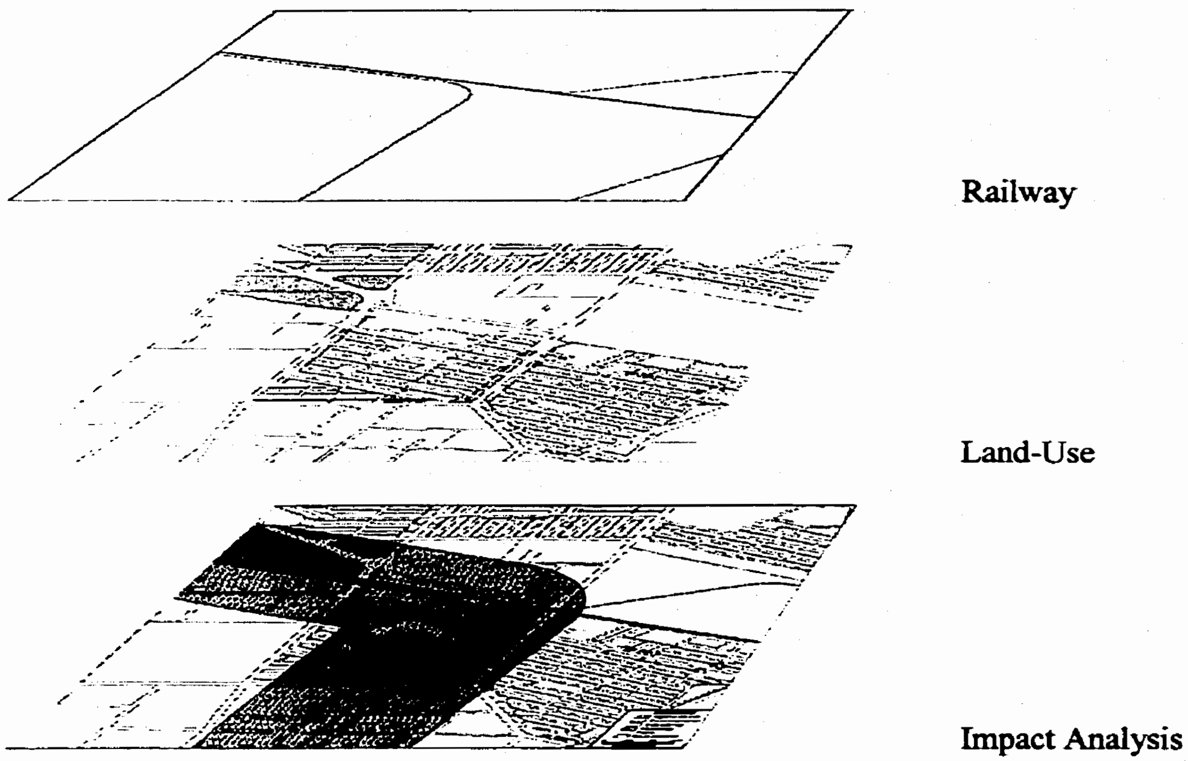


Figure 2.3 Cross-layer Analysis on Railway Impact and Land Usage

Raster and Vector Spatial Data

Due to different schemes of rendition, spatial data can be classified into two groups--raster and vector.

Raster Data

Raster spatial data are a set of pixels containing location and basic attribute information. Each pixel of raster spatial data is of equal size, but a number of pixels can be grouped together based on similar attribution to represent real-world entities. This type of spatial data is suitable for analyses involving continuously varying attributes (i.e., forestry coverage, distribution of precipitation, and satellite imagery). An example of raster data is shown in Figure 2.4.

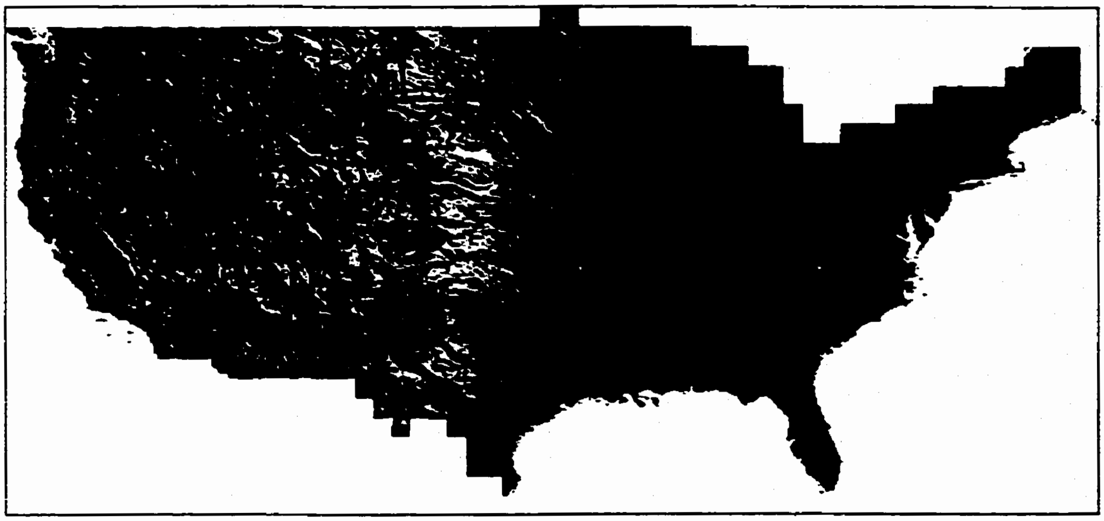


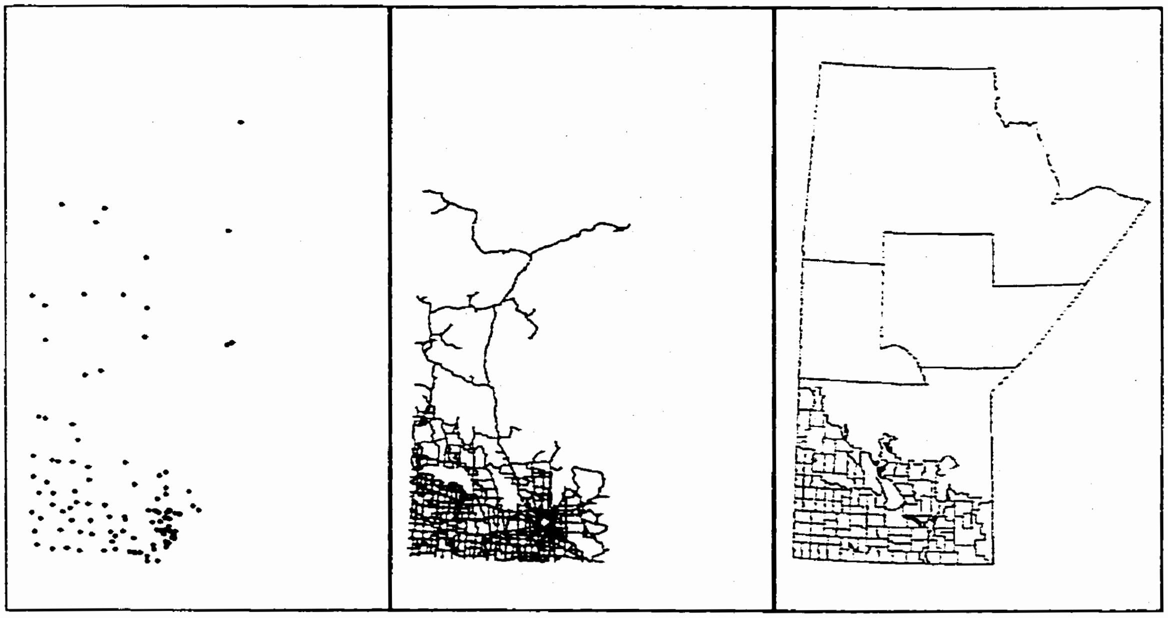
Figure 2.4 Raster Data Example--Digital Elevation Model (DEM)
(Earth Recourse Mapping, 2001)

Vector Data

Vector spatial data use abstract features to represent the real world. Three types of features are used in vector spatial data:

- Point
- Line (including polyline, poly-polyline)
- Polygon (including poly-polygon)

Polyline is a series of connected line segments and is treated as a whole. Using the same concept, poly-polyline and poly-polygon are created. Figure 2.5 gives some examples of vector spatial data features.



Points (Towns) Lines (Highways) Polygons (Municipalities)

Figure 2.5 Examples of Vector Data

Vector spatial data are generally considered to be more efficient in storage and map regeneration. Their data structure also facilitates network analysis and linear referencing, which are used extensively in transportation fields.

2.2.2 Attribute Data

Attribute data are another important component of GIS data. Attribute data are various types of information collected to reflect real-world features. For example, attribute information may include city populations, employment numbers, highway route numbers, bus stop names, and pavement condition at a specific location. Attribute data often take the form of tables, and are stored in databases.

2.3 GIS Functions

2.3.1 Spatial Data Capturing and Editing

GIS provide functions supporting the acquisition of spatial data from various sources. Common spatial data sources include digitizing paper maps, direct digitizing from aerial photos, and satellite imagery.

As real-world entities change, spatial data representation of the real world needs to be edited accordingly. Spatial data editing is often supported in manual and automatic/programmable fashions.

2.3.2 Attribute Data Management

GIS use databases to store and manage attribute data from various sources. Developing attribute data may involve field surveys and labor-intensive data input and linking. The database management system helps to take care of these valuable data and make the best use of them through automated processes.

GIS access attribute data stored in databases either through built-in database engines locally or through an enterprise-wide Relational Database Management System (RDBMS). The enterprise-wide approach brings better security and data integrity to GIS. The associated central database control makes data-warehousing techniques easy to deploy. However, local access to databases is more convenient for users and more common.

2.3.3 Analysis and Modeling

GIS is an intuitive analysis tool. In addition to numbers or charts, users get a real-world perspective that helps provide a clear idea of the situation, what resources are available, and the results of human inputs. The benefit of using GIS is to help users think geographically as opposed to analyzing data without consideration of specific location properties. This is critically important in many transportation engineering and planning activities and in other fields.

Through its spatial and tabular query and modeling ability, GIS helps handle large amounts of spatial-related data and conducts complicated analysis in and of the decision-making process.

Spatial and Tabular Query

GIS provides functions to make tabular or spatial queries of its data sets. GIS can perform structured queries of tabular data, and topological queries of spatial data. In addition, GIS can integrate query results from two types of queries to conduct sophisticated cross-referencing analyses that a conventional information system cannot perform. The query results are also represented in two forms--data view tables or selections of spatial features.

Modeling

GIS users can analyze complicated problems through modeling. Sophisticated models can be built on top of a GIS platform, and supported by various analysis tools provided by GIS. A good example is the recent development of a Transportation Planning Model for the Manitoba Capital Region (Lew, 2000).

2.3.4 Information Integration

Traditionally, information systems have been designed and operated as individual systems. Today, users want to increase the value of individual systems by integrating them on or through a common platform (Montufar et al., 1999 and 2000).

GIS differs from other information systems in its ability to integrate information from different data sources through common location information. An individual information system has its data structured to serve specific purposes. Differences in data structures between one information system and another are often large enough to prevent efficient data sharing and integration. For example, a city's traffic database has a different data structure than a grocery chain's distribution database. However, once location information is attached to these two databases, integration can be achieved using GIS through the location linking process. Queries, such as determining traffic conditions around certain stores and shortest-path routing from warehouse to stores, can be conducted based on integrated information from the two databases.

"*The whole is greater than the sum of its parts.*" By using common location information contained in individual databases, GIS can act as an information integrator by processing information/data on a wide range of subject matters (e.g., traffic, road condition inventory, weather, and accidents), which are related to a specific geographical area. Through data integration, GIS users make the most use of spatial-related information.

GIS also interfaces with different data sources, helps data sharing, and lays the foundation for an integrated comprehensive multi-modal transportation information system.

2.3.5 Information Visualization and Visual Presentation

"*A picture is worth a thousand words.*" The visualization capability of GIS makes it possible for engineers, analysts, decision-makers, and the public to perceive the "picture" of the problem they are facing, and envision the possible circumstances resulted from sophisticated interaction between many factors. In addition, GIS helps deliver technical information to decision-makers who may not have a technical background by using maps to clearly present the results of analyses. GIS allows users to control individual map components, such as scope, theme, scale, and labeling, to make maps efficiently. Also, GIS can present data through tables and charts.

In the transportation information dissemination field, GIS can serve as an efficient data carrier for information exchange between major transportation system components. By establishing an interoperability framework for ITS spatial data, the information exchange process can be facilitated (Goodwin, 1996). For example, transport information such as accidents, and congestion can be linked to GIS and information can be exchanged between police, emergency response teams, and the road authority (Lehtonen and Lähesmaa, 1999).

2.4 GIS software

GIS software allows GIS to interact with different hardware and software components, coordinate their work, interface with users, query databases conventionally or spatially, and automate the workflow to aid users. GIS software also manages the creation, maintenance, dissemination, and storage of spatial and attribute data.

There are many GIS software packages available. It is essential for users to know what GIS software can do for them, as well as the limit of its ability. A major consideration of this research is the functions that various GIS software provide to users. Five categories of GIS software are identified based on functions provided and level of users.

Group I is full-feature GIS software. Principal examples are ARC/INFO, MGE (Modular GIS Environment), and GeoMedia Professional. They are targeted at higher-level users engaged in sophisticated analysis and the development of spatial data for lower-level GIS users. This group of GIS software often provides sophisticated analyzing tools and spatial functions for the creation, editing, management, analysis, display, and mapping of geographic information. However, they are relatively expensive, and require substantial computing power and GIS specialists to operate.

Group II is general-purpose GIS software and is the most commonly used. Examples are ArcView, MapInfo, GeoMedia, and Maptitude. They contain the most commonly required functions of GIS, and support many user-friendly features. Due to its reasonable system requirements and ease in learning and using, this group has the broadest user base.

Group III is specialized GIS software for individual application fields. Examples include TransCAD for transportation applications, FRAMME for utility industries, and ER Mapper for image processing applications. This group of GIS software concentrates on specific fields, providing special functions to assist their users.

Group IV originates from CAD software. Examples are "AutoCAD Map" and "MicroStation Geographics". They incorporate common GIS functions into basic CAD software, and take advantage of the capabilities and flexibilities of CAD functions and the reusability of existing CAD drawing data.

Group V is new and accompanies the rapidly growing field of Internet applications. This group of software can be further sub-grouped into two categories—server software and client software. Examples of Internet GIS servers are ArcIMS, GeoMedia Web Map, MapXtreme, and MapGuide. Examples of client GIS software are ArcPad and OnSite running on hand-held computers (Barnes, 2000).

The responsibility of performing GIS functions is shared by the software on the server side and the client side.

The information is summarized in the following table.

Table 2.1 Comparison of GIS Software Groups

	User Base	**Functionality**	**Pros and Cons**	**Examples**
Group I	GIS Professionals, Advanced Analysts	Complete functions	Powerful tools, Hard to learn, expensive	ARC/INFO, MGE, GeoMedia Pro
Group II	General GIS Users	Commonly used functions	User-friendly, easy to learn, lack of advanced tools	ArcView, MapInfo, Maptitude
Group III	Specialists	Specialized functions	Capable of special analysis, requires support from others	TransCAD, FRAMME, ER Mapper
Group IV	CAD Specialists	Common GIS functions + full CAD functions	Easy access to CAD drawings, lack of analysis tools	AutoCAD Map, MicroStation GeoGraphics
Group V	Field Staff, General Public	Shared functions between Servers and Clients	Highly scalable and flexible, can be used everywhere, limited functions	ArcIMS, MapXtreme, ArcPad

Detailed discussion about each group of GIS software is provided in Appendix A.

3.0 GIS-T Applications and Data Sharing

This chapter outlines the transportation applications of GIS and discusses principal issues related to GIS-T data sharing.

3.1 Transportation Applications of GIS

Transportation is one of the basic human needs. Various transportation systems have been developed to satisfy the increasing demand for mobility. Transportation engineering studies various aspects of these systems in order to plan, construct, and manage them efficiently and safely.

Transportation systems are complicated and have perpetually changing characteristics. The ever-changing nature of transportation and the world it serves requires "continual updating of our understanding of the transportation system, how it is used, what it contributes, and what it affects" (BTS, 1997).

To support this understanding, transportation data and corresponding data-handling information systems are required. For example, highway inventory data, traffic data, accident data, pavement condition data, regulatory data, and environmental data are continuously collected by highway agencies. These data come in different formats, in very large quantities, and increasingly on a close to real-time basis. Information is extracted from raw data. Consequently, comprehension of the information leads to understanding, and understanding lays the foundation for informed decision-making.

As a useful tool for transportation decision-making processes, GIS-T helps transportation engineers, planners, managers, and operators gain better understanding of the transportation system and make well-informed decisions based on geo-referenced information.

3.1.1 Characteristics of Transportation

Transportation has unique characteristics which make GIS a very suitable tool for transportation applications.

Transportation Movements

Transportation involves movements of goods and people, and consequently vehicle movements. GIS provide an efficient way to handle data with spatial and temporal attributes. This makes GIS an excellent platform for transportation data integration, presentation, spatial analysis, and decision-making.

Transportation Networks

The spatial structure of a transportation system is reflected in its network characteristics (Manheim, 1979). Given the power of GIS, it is natural to construct an analysis network on top of GIS to perform transportation analysis.

Historically, transportation agencies have depended on network models to conduct analyses. Network models simplify the world into a network by taking advantage of the fact that land-based vehicle movement is confined to roads, railroads, canals, etc. These simplified models also help position data on the network (Goodwin, 1996).

GIS represents transportation networks as links and nodes, and applies network analysis on a computerized platform. This helps GIS users solve sophisticated network and routing problems by searching through thousands of possible options within a very short time. In fact, GIS road networks play an important role in motor vehicle navigation systems, traffic advisories, scheduling, and emergency vehicle routing (Nystuen et al., 1997).

Systematic Interactions

Transportation is a complicated system consisting of many components (e.g., people, goods, infrastructure, vehicles, rules, and regulations). The relationships between these components are sophisticated and inter-related (e.g., social-economical, environmental, and inter-modal relationships). GIS relate information from different sources using common location identifiers. By providing an integrated information system platform, GIS help analyze the interactions between each component of the transportation system.

3.1.2 GIS-T Application Fields

There are many GIS-T applications. A detailed list is shown in Appendix B. Among these application fields, three require particular attention:

Transportation Planning

Transportation planning involves correlating factors such as land use, environmental concerns, economic development, and components of transportation systems. With its data integration ability, GIS has become a useful tool for the planning process. To further support transportation planning, several GIS are customized to provide sophisticated tools for planning/analysis/engineering, such as network modeling based on a GIS platform.

Transportation Asset Management

Transportation infrastructure includes a wide spectrum of facilities (highways, bridges, airports, harbors, canals, etc.). GIS help to store, manage, and visualize information about these transportation assets, which are spread over space and which interact over time.

Sophisticated management, such as geographical distribution analysis, and network balancing can be supported through GIS functions.

Intelligent Transportation Systems (ITS)

ITS are the integrated application of advanced sensor, computer electronics, and communications technologies and management strategies to provide traveler information and increase the safety and efficiency of the surface transportation system (USDOT, 1999).

A major aspect of ITS is about information--the collection, sharing, processing, and redistribution of information--to move people and goods more efficiently. Information lets travelers make better decisions, and helps improve the efficiency and safety of the various elements of our surface transportation infrastructure: for example, transit systems, freeways, toll facilities, rail intersections, truck regulatory facilities/services, and rural roadways (USDOT, 2001).

With the development of ITS, a large number of sensors and monitors are being installed. Inbound data flow is enormous. To efficiently retrieve information from these data, a suitable GIS platform is needed. Through efficient visual representation, GIS can be a bridge between data and end-users. The success of many ITS applications hinges on the ability to communicate a location message unambiguously across dissimilar map bases.

Transportation applications will benefit greatly by accessing and communicating with multiple GIS data sources. Through GIS data sharing, various parties can share transportation-related information easily and efficiently.

3.2 GIS Data Sharing

In most cases, the development of GIS spatial data is an expensive, time-consuming and labor-intensive process. After a GIS is established, attribute data are continuously gathered to serve the designed purpose of the system and to enhance the information system. In current GIS, these attribute information are attached to specific spatial data sets. The value of this information can be multiplied through data sharing among agencies and departments.

However, current GIS are limited in their capability for data sharing and interoperability. They operate on their own spatial data sets. Different spatial data sets have different scales, levels of details, and data structures. These differences among GIS platforms are large enough to prevent efficient data sharing.

3.2.1 Current Status of GIS-T Data Sharing

GIS-T Base Map Development

To accommodate all types of transportation applications, GIS professionals have developed a multitude of GIS platforms. However, these GIS are developed as individual systems using combinations of different software, hardware, spatial data sets, and representation schemes.

For example, over the past decade, a number of transportation GIS platforms have been created for Manitoba. Most were designed and constructed to cover a specific portion or component of the transportation system in the province. These platforms vary in terms of structure, scale, projection, accuracy, and completeness. A detailed review of these GIS-T base maps is given in Appendix C.

Data Sharing Research and Development

Data sharing and interoperability between GIS are issues of increasing concern in GIS research. Dedicated conferences were held in 1997 and 1999 (NCGIA, 1997), (OGC, 2001); interested parties formed Open GIS Consortium, and recent efforts include linking information at various levels of government. However, efforts are concentrated in GIS software, data format, and linear referencing standard fields. Many efforts are limited to conceptual research. The unique characteristics of GIS-T, and its special spatial data requirements make it necessary to implement GIS interoperability at the spatial data level. Some of the major GIS-T related data sharing and interoperability endeavors are:

Open GIS Consortium (OGC)

One initiative to achieve GIS interoperability is the Open GIS Consortium. This is an association of government agencies, research organizations, software developers, and systems integrators looking to define a set of requirements, standards, and specifications that will support GIS interoperability. The objective of the Open GIS Consortium is to develop technology that will enable an application developer to use any geodata and any geoprocessing function or process available on "the net" within a single environment and a single work flow (OGC, 2001).

Spatial Data Standard

A common need in spatial operations and analysis is the ability to transfer data between digital databases. The requirements for creating such capacity have received a great deal of attention--for example, through the efforts to develop Spatial Data Transfer Standards (SDTS) (Nystuen et al., 1997).

NCHRP proposal

The GIS-T community is establishing a common data model and linear referencing system to enable spatial data sharing (National Cooperative Highway Research Program (NCHRP) 20-27) (Goodwin, 1996).

Geographical Data File (GDF) Standard

Geographical Data File (GDF) is a standard currently under development by the Central European Normalisation (CEN) and International Organisation for Standardisation (ISO) (Lehtonen and Lähesmaa, 1999).

Interoperability Support of GIS Software

When it comes to supporting interoperability, GIS software are weak in many respects. Even though some software provide programs/functions supporting low-level interoperability-related manipulation of spatial data, they do not have a systematic approach to help users fulfill required tasks. For example, ARC/INFO has a line-to-polygon conversion function, but moving the attribute information from line layer to polygon layer (or vice versa) requires manual intervention.

Given the limitations of existing GIS software for supporting interoperability of transportation applications, it is useful to take a systematic approach--developing a framework and associated techniques--to help users overcome the limitations.

3.2.2 Obstacles to GIS Data Sharing

Reasons for difficulties in GIS data sharing are both institutional and technical. Historically, GIS were designed as isolated systems operating locally. Communication or interaction with other GIS was not a major concern because GIS were rare and expensive at that time. The increasing demand for GIS data sharing has exposed the current dedicated GIS differ considerably in terms of data structure, datum, projection, etc. As GIS is being seen and used as a major tool for integrating transportation databases over larger domains, changes to the existing GIS are needed.

However, many billions of dollars have been spent on spatial databases by thousands of agencies and private companies. These GIS and the organizations that support them, with employees and systems skilled in particular ways of viewing and using spatial data, make change expensive (Goodwin, 1996).

Technically, GIS are very complicated and have evolved over a long period of time. Different GIS software use different file formats. During the process of data sharing, complete translation/accommodation is difficult to achieve. Spatial data sets serving as GIS base maps are constructed differently in terms of map projection, datum, and coordinate system. The meta data (data about data) information is often neglected or unavailable, making spatial data linking/matching that is based on a common map projection and coordinate system difficult. Furthermore, spatial linking/matching process can be hard to perform and is often affected by the issues related to spatial data sets, such as differences in data structure, representation scheme, accuracy, currency, and completeness.

Although the relationship between the two aspects--technical and institutional--is complicated and inter-related, the institutional issues often have their roots in technical fields. Therefore, the thesis focuses its attention on technical aspects of GIS data sharing and interoperability issues.

3.2.3 Common Interoperability Issues Concerning Spatial Data

Many issues arise when attempting to share data between GIS. Since GIS spatial data are often at the center of the issues, the thesis pays special attention to the following recurring issues related to spatial data.

<u>Accuracy</u>

Two types of accuracy are involved in the data sharing process. They are absolute accuracy and relative accuracy, as illustrated in Figure 3.1. Interoperability is more affected by relative accuracy. Poor relative accuracy causes large discrepancies between spatial data sets, which can lead to inaccurate and inefficient linking/matching results. Since relative accuracy is often inter-related with absolute accuracy, better accuracy is desired in general for conducting GIS data sharing.

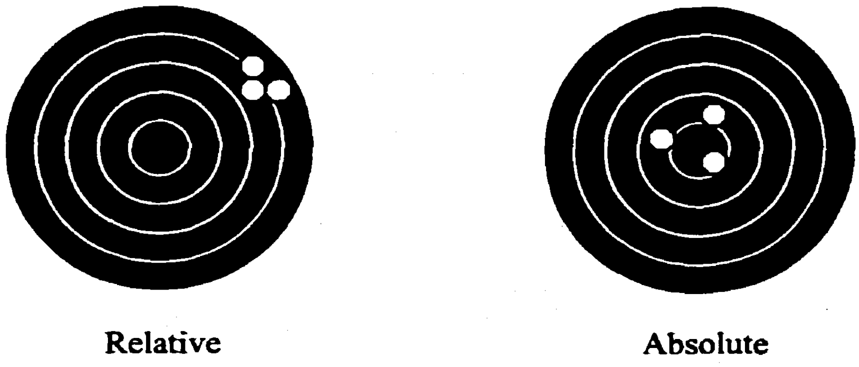


Figure 3.1 Relative Accuracy and Absolute Accuracy (Minty et al., 2001)

The accuracy of the spatial data set of a specific GIS is often limited to the resources available at a certain time and within a certain budget. Computational ability is another factor affecting spatial data accuracy. New technologies provide increasingly accurate spatial data. More and more spatial data with improved quality are available to the general public. And new computers are more capable of handling complicated spatial information.

Accuracy of the spatial information not only helps provide better representation of the data, but also makes possible the analysis pertaining to data from different sources. Good relative accuracy is important to ensure correct spatial analysis.

Level of Generalization

Generalization refers to the mapping process in which complicated real-world features are simplified for clarity. All maps and spatial data sets are simplified models of reality. In conjunction with thematic symbols and labels, the generalization process helps make a map clear. Depending on the purposes of the GIS and the scale of the base maps, the levels of generalization vary greatly from one GIS to another. For example, a city can be represented as a point in one GIS, an area in another, and thousands of links and nodes in another.

When the same real-world features are represented differently in various GIS, data sharing among these GIS becomes complicated. The issue is further complicated by the fact that the level of generalization is often tied together with scale, accuracy and level of detail. Highly generalized spatial data often has a smaller scale, lower accuracy and less detailed spatial data features. On the other hand, less generalized spatial data is closer to reality, containing more detail, and often associated with better accuracy which is resulted from large-scale mapping.

Highly detailed spatial data are expensive and time-consuming to create and handle. And they may not be suitable for some applications. For example, it may not be appropriate to incorporate highway interchange details into a national highway network designed for long distance routing purposes. But the turning possibilities might be very important in and around major urban centers.

Differences in level of generalization and consequently in the spatial data structures make matching/linking of these data a difficult process. Matching spatial data using the buffering technique and cross-layer linking is required to identify the relationships between spatial data features from different data sets.

Currency and Completeness

Transportation networks are continually changing: new facilities are built; existing roads may be rerouted; a railway may be realigned, etc. GIS spatial data should reflect these changes in a timely fashion. Otherwise, the information provided by that specific GIS and the analysis conducted based on the information will soon be rendered incomplete and incorrect.

Relying on a single source to build and update spatial data, most current GIS lack the flexibility of updating their spatial data easily and promptly. Updating and improving spatial data becomes difficult and often relies on new GIS development. A long updating period for spatial data hinders data sharing. That is because mis-linking and mis-matching often occur during the data sharing process when two data sets were collected at different time.

<u>Difference in Spatial Data Structure</u>

The structure of the information storage in a specific GIS application often impedes the adoption of newer, more accurate spatial information. GIS spatial data were often structured to represent the transportation system which they were designed to deal with, and often re-structured during use to accommodate a certain application. Newly available spatial data may not be compatible with the original data structure; making direct transformation impossible. Examples of this issue are different spatial data structures for highway networks. In a provincial-level network, highways are often segmented from one highway intersection to another, leaving local accesses not represented. On the other hand, a local street network typically segments roads by intersection, no matter how big or small (e.g., back lane). In that case, a provincial highway going through an urban area is represented as many small sections in a local street network.

Meta Data for Spatial Data

Meta data is data about data (Microsoft, 1998). It gives the GIS user information about how to interpret and handle the spatial data under consideration. A complete set of meta data should include:

- Identification information
- Data quality information
- Spatial data organization information
- Spatial reference information
- Entity and attribute information
- Distribution information
- Reference information
- Citation information
- Time period information
- Contact information

The U.S. Federal Geographic Data Committee (FGDC) has created a Standard for Digital Geospatial Meta Data. It provides a common set of terminology and definitions for the documentation of digital geospatial data (FGDC, 2001). However, this standard has not been widely adopted, and very few spatial data sets are provided with standard meta data. Also, GIS software has very limited support for meta data.

3.3 Sharing Information on Broader Bases

As geospatial information technologies are more commonly adopted in various transportation fields, GIS data sharing is not limited to existing GIS data sources. New spatial data sources are becoming more promising as new technologies become available (e.g., satellite imagery). At the same time, historical data remain valuable despite the dramatic changes in technology.

Accommodating historical data

GIS spatial data often concentrate on representing the current situation of the transportation system. However, in many cases, transportation analyses need support from historical data. For example, highway accident analyses may be conducted by relating historical roadway alignment with accidents to be able to determine the contributing factors.

Comparisons of a network over different time periods may also be of interest (e.g., the rationalization of the prairie region rail network) (Dick and Clayton, 2001). Overlaying a digital database of a network from one time with the database from another time allows the differences to be highlighted. The changes that have occurred often provide useful insights. In general, being able to merge or match different databases characterizing the same spatial network is useful in many transportation contexts.

Spatial data of current GIS often have different data structure than the historical spatial data. To accommodate historic spatial data, interoperability is required to transform data between current and historical data sets. And a translation capability to convert legacy databases into GIS compatible formats is also required.

<u>Alternative Spatial Data Sources</u>

GIS can take advantage of existing spatial data from alternative sources through the data import/conversion process. For example, engineering designs (field surveys and computer aided design (CAD) drawings) are efficient and accurate sources of spatial data. Global Positioning System (GPS) and its derivative Deferential GPS (DGPS) also provide opportunities to generate spatial data more quickly and less expensively than conventional digitization. Although there are limitations associated with them, these alternative spatial data sources are readily available and at lower cost than conventional sources.

GPS

> GPS provides reasonably accurate spatial data at fairly low cost. With the termination of Selective Availability (SA) from the U.S. Military, and the establishment of a network of Differential GPS (DGPS) beacons covering a large portion of North America, GPS has become an attractive data source providing easy updates to existing spatial data sets.

CAD

CAD drawings are easy to convert to GIS spatial data. There are many existing engineering drawings in CAD format that are very accurate and can easily meet GIS accuracy requirements. However, engineering design and actual construction often have discrepancies; as-built checking is desirable in converting CAD drawings to GIS data sets. (Dias, 2001), (Lopez, 2000)

Aerial Photography and Satellite Imagery

Ortho-corrected aerial photography and satellite imagery are not only useful backdrops in GIS analysis, they are also basic sources of spatial data in today's GIS mapping practice. This imagery contains ground truths and culture information, which can provide better understanding of the real world when used in conjunction with other GIS spatial data.

Integrating spatial data from multiple sources can also enhance existing GIS by providing flexibility to update spatial data sets more easily and more frequently, and consequently provide better understanding of the transportation system.

4.0 Framework for GIS-T Interoperability

This chapter discusses key characteristics of spatial data in relation to GIS-T; and describes the design, structure, and construction of the framework developed in the research to facilitate GIS-T interoperability. Basic ideas and methodologies incorporated in the framework are discussed.

4.1 Objectives of the Framework

The framework is designed to solve interoperability problems encountered in the realm of GIS-T. Thus, the framework is transportation-oriented. This is reflected in the fact that the framework deals with vector spatial data and has an emphasis on networks. Nevertheless, the basic idea and techniques can be applied to much broader areas.

To accommodate GIS-T data from various data sources which may differ in coverage, accuracy, data structure, and generalization, the framework is designed to provide a "party" place. Consequently, integration ability is developed to establish linkages between GIS-T spatial data. Through the linkage established between spatial data, attribute data is also rendered interoperable.

4.2 Framework Structure

The linking and matching of GIS-T spatial data is a sophisticated process which may involve several phases and employ several techniques. The steps to follow vary from case to case. To incorporate these various components, and provide guidelines to follow when new spatial data are to be processed, a framework is established to serve as the fundamental structure and facilitate the data sharing process. Consolidated basic methods and improved techniques can fit into the framework and support interoperability.

The conceptual framework is a skeletal structure for enclosing GIS-T spatial data sets from various data sources. It also provides a place for the techniques to conduct the actual linking processes on spatial data sets and consequently on attached attribute data.

To simplify the data sharing process, the framework handles spatial data and attribute data separately. Since the spatial data--by definition--have common location information that can naturally be used as a linking agent, the framework is constructed around, and has its emphasis on, spatial data. When spatial data sets are linked, attribute data can use interoperable spatial data as bridges to allow the free flow of attribute information.

4.2.1 Form

The framework forms a pyramid shape in accordance with the generally-accepted view of the scaling factor in mapping and spatial data generalization. The shape of the framework is illustrated in Figure 4.1.

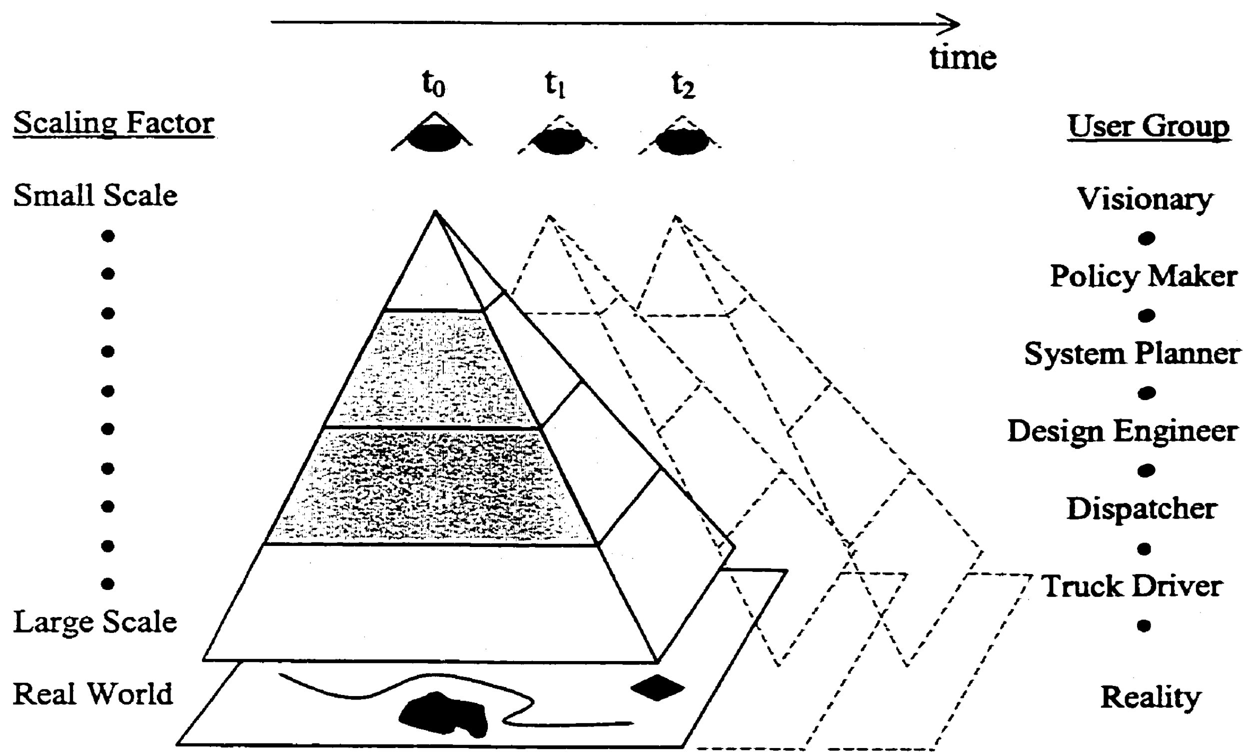


Figure 4.1 The Conceptual Framework

Layers in Levels of Abstraction

The framework consists of layers representing different levels of abstraction of the real world. These layers are ordered by their levels of abstraction. Naturally, the level of abstraction is correlated with mapping generalization and map scale.

The lower the level in the framework, the larger the scale, and the closer spatial data get to the real world. Consequently, to render the detail of the real world, spatial data become more complicated and often have better accuracy. Conversely, less spatial data is necessary as the abstraction level goes up in the framework. The result is that no two spatial data sets are exactly alike. However, the real world entities they represent are the same.

Tiers in Levels of Abstraction

Depending on the scale of the spatial data set, spatial data are structured following certain generalization rules. For example, a city can be represented as a myriad of points and lines depicting a dense urban street network in the city transit planning data set, and merely a point in the world city data set. To identify these abrupt changes in the generalization process, spatial data layers are further grouped into tiers. Within one tier, the spatial data representing the same real-world entities do not change type from one spatial data layer to another. This hierarchical structure of the framework facilitates the selection of appropriate methods and techniques to be applied to render different spatial data sets interoperable.

The number of layers (or tiers) in the framework is not fixed. It varies depending on how many different spatial data sources are used in the data sharing project.

Spatial Coverage

To focus the data sharing effort, the conceptual framework's spatial coverage has a limited scope covering the study region. The study region can vary from case to case. Only spatial data sets (or portions of them) that are inside the study region are handled by the framework.

4.2.2 Dimensions

There is more than one dimension in the framework. Besides the scaling/level-of-abstraction point-of-view, another view is provided from the user group point-of-view.

User Level Dimension

GIS-T users range from high-level users to low-level users, from sand truck drivers who determine when/where/how much sand to spread, to policy makers who plan long-term multi-modal transportation systems. These different user groups require different things from the spatial data for their specific tasks. In the GIS realm, spatial data are application-driven. "Higher" level user groups often require a larger scope of spatial coverage and are concerned with systematic interaction more than specific details. On the other hand, "lower" level user groups coping with local issues need accuracy and a greater level of detail to support their applications. Therefore, there is a correlation between scaling/level-of-abstraction factor and user group level. The framework helps to bring these two perspectives together.

Vertical Views

Besides the horizontal layer/tier-type views, the framework also provides GIS-T users with vertical views. One vertical view is top-down. From this viewpoint, GIS-T users approach data sharing from a higher level. For example, a province-wide transportation planner can fit lower level spatial data features from detailed municipal road GIS into the provincial GIS to share information residing on lower level GIS. Conversely, the bottom-up view of the framework assists data sharing efforts by "lower level" users taking advantage of higher level layer information.

Time Dimensions

The framework has two time dimensions. The first one is the common time. Since transportation systems are dynamic, systems structure and components change as time goes by. The second dimension related to time is the perception time. Often, it takes time for GIS-T to reflect real-world changes. The time from the actual change taking place to the change being perceived by GIS-T forms another dimension of the framework.

In order to ensure the correctness of analyses, appropriate time measurement should be applied to both time dimensions. Where data sharing is concerned, faster may not always be better. Synchronization between GIS-T data sources is essential.

4.2.3 Data Sharing Techniques

To conduct information sharing processes in practice, the framework must be supported by a group of techniques. These techniques can be applied to spatial data layers embraced in the framework to help link or match spatial data sets, manipulate spatial data structures, integrate spatial data sets, and populate the attribute data table. Although these techniques are developed to solve individual problems, they can be standardized to handle more general tasks in future processes. Some of the techniques employ existing functions found in GIS software off-the-shelf. Typically, however, majority of the techniques involve in-house computer programs developed to solve particular problems. The accumulation of these techniques forms a growing toolbox for data sharing. Subsequently, the data sharing process can take advantage of previous endeavors, save time, and keep the quality of the results.

4.2.4 Framework Guided Data Sharing Process

In the research, the data sharing process is to be conducted in the following manner:

- GIS-T data sets from different information sources are separated into spatial data sets and attribute data sets.

- The spatial data sets are inserted into the framework according to their levels of abstraction.
- GIS-T users decide the flow directions (upward, downward, or both) for the shared information.
- According to the direction of the information flow and relative position of the spatial data layers (or tiers), suitable methods are chosen, and relevant techniques are applied or developed/applied to the spatial data sets.
- The resulting linking/matching tables are utilized to help move attached attribute information across spatial data layers.
- To take the result of the framework a step further, the GIS-T user can derive a spatial data set based on the best quality data sources available with a suitable level of abstraction to enhance the quality of their particular GIS-T application--thus facilitating future data sharing projects.

4.3 Theoretical Foundations

Location Linker

Although spatial data sets from different sources differ in accuracy, data structure, and generalization, they represent the same real-world entities. The location information that is inherent to spatial data provides the foundation for the framework to identify spatial relationships between different spatial data sets. Based on these spatial relationships, matching and linking of the spatial data can be conducted effectively.

Spatial Relationship

The spatial relationship defines the way in which two spatial data features (points, lines, and polygons) interact with each other. Common examples include "containing", "intersecting", and "within certain distance". They can be further broken down to different groups by spatial data feature types. It is also important to note that the actual "relationship identifying" process is conducted in a cross-layer manner, which differs from ordinary within-one-layer topological analyses.

Explicit and Implicit Data Sharing Approaches

The data sharing approaches of the framework utilize the spatial relationship in different ways. These approaches can be grouped into two categories--explicit and implicit.

- *Explicit approaches* directly investigate spatial relationships and derive a matching/linking relationship based on spatial similarity. Generally, spatial data from different sources are in the vicinity of absolute location of the real-world entity they represent. The spatial similarity can be explored using buffering, which creates a buffer zone around one spatial feature--a point, line, or polygon--and finds spatial data features inside this buffer zone. This spatial similarity renders the matching, linking relationships (Han et al., 2000).

- *Implicit approaches* employ advanced techniques such as linear referencing or centerline-derivation to extend the scope of spatial data sharing. In these approaches, spatial relationships are utilized internally under the cover of mechanical processes. For example, the centerline-deriving technique finds points having equal distances from both sides of the roadway along a specific route, and the linear referencing technique converts one-dimensional tabular data into 2D (sometimes 3D) spatial data along the line-set with certain distance bearings, e.g., mile post. Through these approaches, spatial data can be restructured and rendered in different fashions to fit the needs of different applications. New usage opportunities are also created for information residing on legacy databases which can be converted into a GIS format compatible with newly developed GIS.

4.4 Data Sources

The framework can handle a broad spectrum of GIS data. Through the application of appropriate methods and techniques, these geo-referenced GIS data can be blended into the framework and have spatial data features linked or matched.

The information sources are not limited to existing GIS data sets. Any location-referenced information can also be incorporated into the framework. These additional information sources will enhance the framework's spatial data and attribute data aspects. These information sources can include:

- Basic land survey systems
- Geodetic control points
- Engineering drawings, CAD drawings
- Aerial photography, satellite imagery
- GPS field surveys
- Mathematically calculated grids

4.5 Outcomes

4.5.1 Spatial and Attribute Data Sharing

The framework helps to achieve better interoperability which benefits spatial data sharing and, consequently, attribute data sharing between different GIS. The interoperable information matching/linking different GIS data sets are generally delivered in the form of a translation table. With the table, further data sharing is made possible on the basis of this foundational relationship.

4.5.2 Improved Spatial Data Quality

Improved spatial data quality benefits GIS analysis and further data sharing efforts. An important output of the framework is that it provides an avenue to inexpensive, better-quality, and more compatible spatial data sets for every data source involved in the framework. With interoperability at the spatial data level, spatial data sets can swap spatial data features interchangeably between different data sources, taking advantage of the best (most accurate, current, complete, and appropriately generalized) spatial data available. The process can be accomplished through the application of techniques associated with the framework. This creates new opportunities for existing GIS-T to enhance and update their spatial data to better reflect the real-world and facilitate future application development.

The output spatial data sets can be more accurate, more complete, constantly updated, and most importantly completely compatible with the original GIS-T application. Therefore, GIS-T applications do not need to make any changes to accommodate new spatial data, and the transition process will be minimized.

4.5.3 Interoperable GIS Platform

As more and more GIS data sets are included in the framework, a repository of data containing original line pieces from different sources (with their sources identified and having the best accuracy and currency) can be created. Based on this repository, an interoperable GIS-T platform can be derived and maintained at a high standard. This approach will greatly facilitate future GIS-T interoperability.

The framework also supports new GIS-T developments by providing spatial data platforms which have the highest quality data available and built-in interoperability with current GIS-T. Ideally, a GIS-T designed for multiple purposes should consist of two (or more) sets of spatial data specifically structured to meet individual needs, such as conceptual analysis, detailed asset management, and visual presentation. The requirements of different purposes can be met for the most part without compromise. Interoperability among these spatial data sets binds them together into an integral platform for transportation applications.

5.0 IMPLEMENTATION OF THE FRAMEWORK

This chapter describes the implementation of the framework discussed in Chapter 4. Design criteria and programming considerations for the framework implementation are discussed in detail, and examples of technique development are provided.

5.1 Techniques Supporting the Framework

Interoperability of spatial data is explained schematically in Figure 5.1. Assume that A and B are spatial data features representing the same entity in the real world. In GIS 1, A is defined by coordinates (x1, y1, z1). In GIS 2, B is defined by coordinates (x2, y2, z2), where x1 $\neq$ x2, y1 $\neq$ y2, z1 $\neq$ z2. Therefore, without interoperability, A $\neq$ B (even though in reality, it does). With the interoperability concept, x1 can be linked to x2 and so on ... in that way GIS users can figure out A is in fact the same as B. Through the framework and techniques, x1, y1, z1 are rendered the way GIS user knows them to be, which bring out the end result of A=B. This process can be time consuming and labor-intensive when a large amount of data is to be handled. Therefore, techniques supporting the framework are developed to automate the process.

in reality	A = B		
but in GIS 1	A is (x1, y1, z1)		
and in GIS 2	B is (x2, y2, z2)		
thus, without interoperability	x1 $\neq$ x2, y1 $\neq$ y2, z1 $\neq$ z2	$\Rightarrow$	A $\neq$ B
but, with interoperability concept	x1 $\leftrightarrow$ x2, y1 $\leftrightarrow$ y2, z1 $\leftrightarrow$ z2	$\Rightarrow$	A $\leftrightarrow$ B
and, through the framework and techniques	x1 = x2, y1 = y2, z1 = z2	$\Rightarrow$	A = B

Figure 5.1: GIS Spatial Data Interoperability

A number of techniques are developed to solve a series of interoperability problems encountered during the development and implementation of the framework. As discussed above, they are focused on spatial data processing. The techniques are grouped into two categories, based on their purposes to provide accessible and reusable tools for future data sharing projects:

- Spatial data restructuring and enhancement--manipulating the spatial data structure to meet specific needs
- Spatial data linking and matching--using location proximity to identify relationships between spatial data sets

To help assign attribute data to their matching spatial data, a special group of techniques is also developed for attribute data integration.

5.2 Design Considerations

Automation and User Intervention

Since spatial data sets often have complicated structures that necessitate extensive numerical calculation when data-sharing techniques are implemented, it is desirable to implement the techniques in an automated fashion. However, user intervention is required to handle extraordinary situations. For example, in the process of linear referencing, when tabular data have mileposts lying outside the highway control section range, user intervention is needed to determine the locations based on prior knowledge and assisting information about the site. Techniques developed in this research involve combinations of automated processes and user intervention.

Flexibility

To accommodate different GIS spatial data sets, the research has developed programs with the flexibility to handle different data structures. This flexibility is achieved through altering the parameters of the program. This practice also provides reusability of the codes to fit into different types of problem solving.

Using existing GIS functions

Whenever possible, the techniques developed in the research employ functions in commonly available GIS software to eliminate unnecessary programming. GIS software provides general-purpose functions and often comes with an internal programming language, which can be used to combine functions and automate the work process. However, these functions are inadequate to perform data sharing in most cases. When GIS software cannot handle the situation, specialized in-house programs are developed to implement the framework.

Data File Formats

To provide the highest degree of accessibility to the spatial data sets, this research has adopted text format as the common spatial file format internal to the framework. Similarly, dBase format is adopted as the common file format for attribute data. Utilizing these common formats facilitates the data sharing process involving existing GIS functions and in-house programs.

Commonly used proprietary GIS data formats are supported during the data input and output process to facilitate end usage. In the proprietary file formats, spatial data and attribute data are often joined together by GIS software. Input data sets need to be disassembled and converted to text (ASCII) and dBase formats respectively when they are processed by computer programs. Output data sets need to go through a reverse process to convert back to individual proprietary format.

In-house Programming

In the research, in-house programs are developed to solve data sharing problems in an efficient manner. A limited number of factors are considered during the programming. "Performance" is a major concern, especially when large amounts of spatial data are to be processed. Non-crucial programming features such as Graphical User Interfaces (GUI) are eliminated from the in-house programs to simplify the program and reduce running time.

"Accuracy" is another factor. To ensure accuracy of the results, the in-house programs often require time-consuming floating-point calculations. To address the large calculation requirements of floating-point calculations, the C++ programming language is chosen for implementing the techniques to provide better performance (i.e., faster calculation) without sacrificing accuracy of the calculations.

Quality Control Process

The results of the matching/linking processes are contained in multiple tables. Before further data sharing, it is necessary to remove ambiguity within the table. This cleaning process uses database programs to search records, spot anomalies and identify ambiguities. Then a combination of automatic and interactive processes is applied to resolve the problems. Using the resulting match/link table, data sharing can be conducted without ambiguity throughout the framework.

5.3 Spatial data Characteristics

In order to address issues related to GIS-T and data sharing, it is essential to have a clear understanding of the characteristics of spatial data.

Spatial data form the basis of GIS and play an important role in the GIS data sharing process. Vector data are highly suitable and widely used in transportation GIS applications. Therefore, a more detailed discussion is given to this type of spatial data.

Vector Data Components

In vector spatial data sets, real-world entities are represented as features. These features consist of pairs of coordinates. A point feature needs one set of coordinates. A line feature (other than a straight line) needs more than two sets of coordinates to determine the shape and location of the line. In other words, lines can be seen as consisting of a combination of points, including a start point and an end point. Typical example of start and end point is at-grade intersection. The rest of the points--vertices, or shape points--are used to depict the shape of the line. The more complicated shape a line has, the more shape points required. Polygon features (sometimes called area features) are formed by joining lines to become the boundary of an area.

Coordinates

Coordinates help to position spatial data features on the Earth surface. Depending on the individual spatial data set, coordinate information contained in spatial data can be 2-dimensional (2-D) or 3-dimensional (3-D). For many transportation purposes, 2-D representation is adequate. For others, the third dimension is necessary. Adding elevation information will significantly increase the complexity of GIS spatial calculation (e.g., distance calculation), and therefore make GIS functions less efficient. However, the added third dimension is valuable for certain applications (e.g., terrain modeling, grade calculation, drainage design, train performance modeling, and determining truck power-gearing requirements).

Topology

Spatial relationships are often referred to as topology. Topology is a mathematical method showing how point, line and area features are related to one another spatially.

There are four general ways that spatial features can be topologically related. They are illustrated by the following figures.

- Next to each other--adjacent

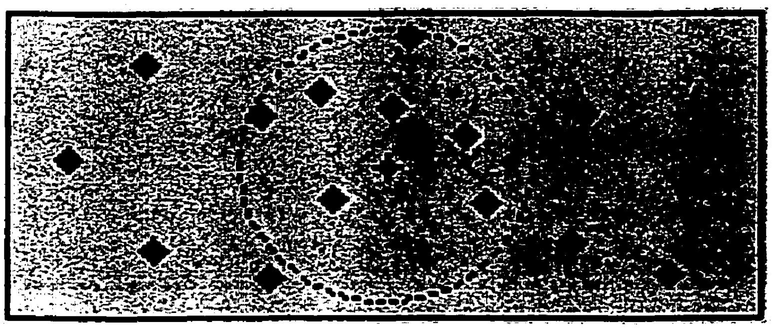


Figure 5.2: Example of Adjacent Spatial Features (Intergraph, 1998b)

- Linked to each other--connected

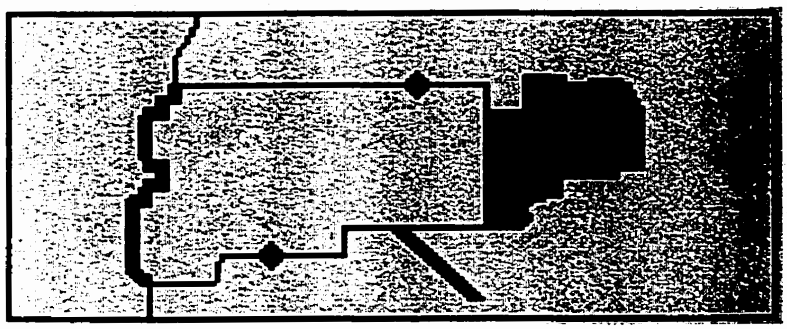


Figure 5.3: Example of Connected Spatial Features (Intergraph, 1998b)

- One feature included within another--contained

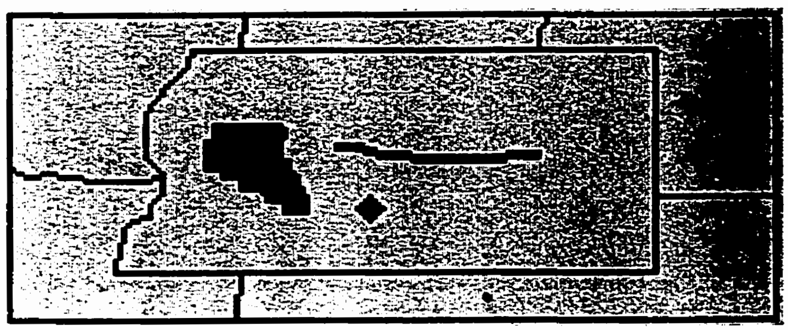


Figure 5.4: Example of Contained Spatial Features (Intergraph, 1998b)

- Two features representing the same location--coincident

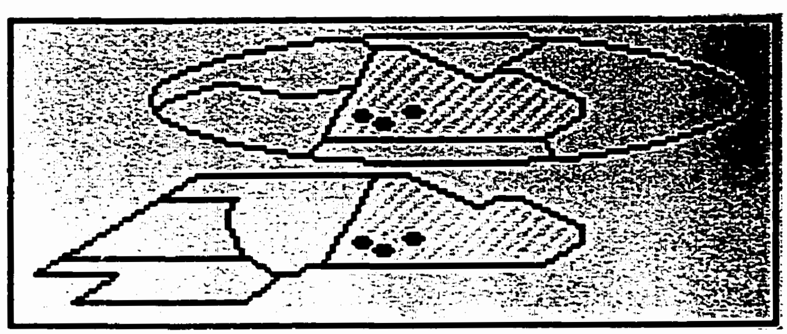


Figure 5.5: Example of Coincident Spatial Features (Intergraph, 1998b)

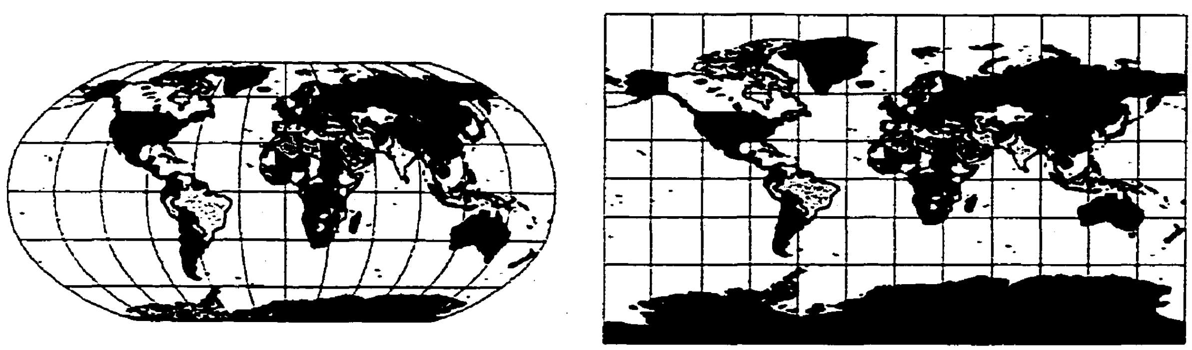
Spatial analysis can be done based on the relationships between spatial features within one spatial layer or crossing layers.

Mapping Characteristics

GIS visually presents information by displaying the spatial data in the form of electronic maps. Spatial data sets used in electronic maps resemble ordinary paper maps in many ways.

Projection

All maps, whether on paper or in the form of a structured database within a computer, are simplified models of reality (Goodwin, 1996). When spatial data are displayed on flat paper or a computer screen, the curved Earth surface needs to be projected. There are many projections developed to represent the real world with as little distortion as possible. Depending on the situation, a certain projection is preferred over other projections. Commonly used projections include Lambert Azimuthal Equal Area, Transverse Mercator, Robinson, and Miller Cylindrical Projection. Examples of the latter two projections are illustrated in Figure 5.6.



Robinson Projection　　　Miller Cylindrical Projection

Figure 5.6: World Maps in Different Projections

Datum

In spatial data and mapping, the earth is referred to as an irregular ellipsoid, which is often called a spheroid. Through the years, more and more accurate spheroids have been developed to better represent the earth. A particular datum is defined in terms of a combination of the reference ellipsoid (spheroid) and a specific geodetic control network. Datum is always used as reference by coordinates contained in spatial data. Commonly used datum in North America are NAD27 (North American Datum 1927), NAD83 (North American Datum 1983), and WGS84 (World Geodetic Spheroid 1984).

Coordinate System

Coordinate systems are used to locate features on the Earth surface. There are two types of coordinate systems--the geographic coordinate system and the projected coordinate system. Based on a specific datum, the geographic coordinate system uses longitude and latitude to locate geographical features. Often referred to as grid systems, projected coordinate systems are created by applying a certain projection based on a specific datum. Coordinate systems can be designed for a region (e.g., U.S. State Plane Coordinate Systems), a nation (e.g., U.K. Ordinance Survey Grid), or worldwide (e.g., Universal Transverse Mercator (UTM)).

Map Scale

As with a paper map, spatial data has a scale. Despite the ability to zoom using GIS, the scale associated with the spatial data remains unchanged. The scale of the spatial data is determined during the digitizing process, and has implications for accuracy and the level of detail. Generally, larger scale spatial data have better accuracy and provide higher levels of detail. As the scale gets smaller, accuracy decreases and detailed spatial data features must be generalized.

5.4 Transportation Characteristics of Spatial Data

The diversity of transportation applications places high demands on GIS software and especially on GIS spatial data. The successful implementation of GIS-T applications requires accurate, timely, and well-attributed spatial data sets. To meet these demands, GIS spatial data are constructed differently from one GIS to another. These differences between dedicated spatial data sets can make GIS data sharing difficult.

5.4.1 Transportation Network Characteristics

Under network terminology, spatial features become network components: point features become nodes; line features become links; polygon features become areas. These network components are used to represent transportation systems. For example, nodes can be used to represent traffic generators or street intersections; links can represent highways or canals; areas can represent traffic zones or political boundaries. In transportation networks, route systems are also introduced to represent a routing path, which consists of a sequence of links connecting two points. Examples are bus routes, highway control sections, and emergency response systems.

Transportation networks have some important characteristics regarding connectivity, directionality, and network partition.

Connectivity

Connectivity defines the accessibility from one link to another when they meet. It affects the accessibility of network components and routing options from node to node. Connectivity should reflect the real-world situation. For example, there should be a connecting node at every intersection in the street network except where one street overpasses another. However, to facilitate higher-level analysis, connectivity can be limited by disconnecting small intersections.

Directionality

Direction determines origins and destinations of flows on links. In transportation, each link implies two-way traffic. There is an inherent direction associated with network links, which is defined by the sequence of shape nodes making up the link. Unless restricted to one-way traffic, links in a transportation network have two directions, one following the inherent direction, and one against it. When direction-related data are attached to networks, GIS-T analyses can be conducted at a more detailed and accurate level. For example, accident sites can be identified on one specific side of the roadway.

Network Partition

Partitioning of the network is a common practice when certain restrictions are applied to specific types of movement. For example, heavy truck movements are restricted to truck routes in many cities. Given good attribute information, these partial networks can be differentiated in a GIS-T network to support various analyses (e.g., safety-related oversize/overweight vehicle permitting (Price et al., 1999) and real-time routing subject to Winter Weight Premiums and Spring Weight Restrictions (Montufar et al., 2000)).

5.4.2 Linear Referencing and Dynamic Segmentation

Linear referencing is a technique by which the position of a network-related object is referenced to a known object and expressed as an offset distance along a link. By using the linear referencing technique, GIS-T can transform conventional tabular data into GIS spatial data. For example, pavement condition ratings may be positioned by highway control section and milepost. Figure 5.7 illustrates linear referencing schematically.

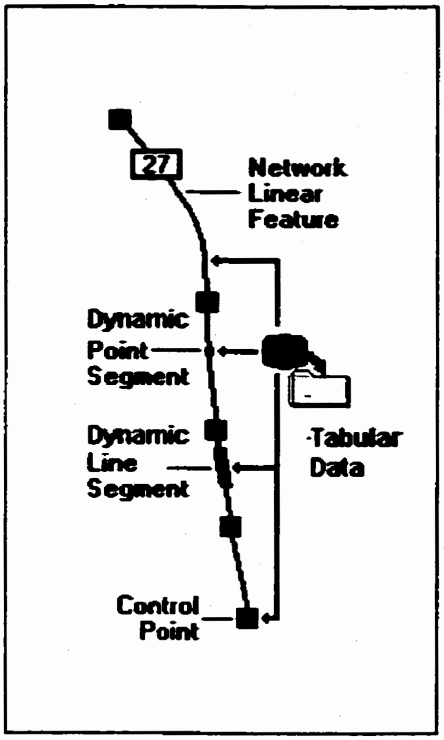


Figure 5.7: Linear Referencing Example (Intergraph, 1998a)

Dynamic segmentation employs the linear referencing technique, deriving sub-segments along existing line features based on reference section and offset information. The process is accomplished automatically by the algorithms implemented through computers, and changes made to referencing information will be reflected in derived results.

Many transportation-related assets and events can be positioned or re-structured using the linear referencing/dynamic segmentation technique (Kennedy et al., 2000), (Dick and Clayton, 2001). GIS-T networks are often constructed to support linear referencing/dynamic segmentation. The advantages of using linear referencing/dynamic segmentation are savings in the storage of the information and flexibility in making changes to the data.

5.4.3 Rendering spatial data GIS-T usable

Although many GIS have components that represent transportation activities, a GIS platform needs to be structured properly to support transportation applications. Newly available spatial data which are not designed and constructed for GIS-T applications should be re-structured according to transportation requirements to make them GIS-T usable. Consequently, GIS-T applications can benefit from the improved accuracy and currency associated with the new spatial data.

5.5 Technique Development

The process to develop individual techniques incorporates the design criteria discussed above. Currently, techniques developed in the research are:

5.5.1 Category I--Spatial Data Restructuring and Enhancement

Two techniques were developed in the research to restructure spatial data to fit special needs.

Standardizing Topological Directions in a Complex Network

The City of Winnipeg transportation department maintains a street centerline (SCL) GIS system to represent the city's street network. Various transportation-related information is attached to this SCL GIS system. Examples include functional classification, accident data, and traffic volume estimates. The most recent published development is a new transportation planning model for the Winnipeg capital region (Lew, 2000).

A topology issue arises recurrently during these endeavors. Due to the lack of uniformity in the direction of line entities in the SCL spatial data set, directional traffic counting data cannot be assigned to the SCL through an automated process. The amount of directional data makes manual assignment prohibitive.

In order to overcome this problem and facilitate future GIS data sharing, the research developed a technique and associated in-house computer program to restructure the spatial data, giving standardized direction to the line features while maintaining the integrity of the original data set. The design criterion for restructuring was to standardize all topology to run mostly west to east, and mostly south to north. These are the directions compatible with the Manitoba provincial highway Control Section system. Figure 5.8 shows the difference between original directions and modified directions of the streets in a Winnipeg urban area. Detailed discussion of the restructuring technique is provided in Appendix D.

Original (haphazard)

Modified (standardized)

Figure 5.8: Difference between Original Directions and Modified Topological Directions

Automated Theoretical Centerline Deriving

This technique was developed to derive a theoretical single centerline from double-centerlines which are used to represent divided highways in certain GIS-T base maps.

Current situation

Two schemes are commonly used in the current GIS-T field to represent divided highways:

- Single Centerline Representation – using one line feature following the centerline of the total roadway to depict a divided highway. When this scheme is carried through the entire highway network, all highway sections are represented by single centerlines, whether they are divided or undivided. The internal compatibility and simplicity of this scheme facilitates high-level transportation planning, network analysis, and small-scale presentation of the highway system.

- Double Centerline Representation – divided highways are portrayed by two centerlines following the centerlines of the two roadways of each side. This representation, often associated with ramp details at interchanges, provides a more detailed picture of the divided highway infrastructure than single centerline. This scheme facilitates asset management processes, representation, accident analysis, and large-scale presentation (e.g., traffic data or IRI measurement). However, network-base applications are not supported very well by the double centerline scheme.

There are advantages and disadvantages associated with both representation schemes. They are summarized as follows:

Table 5.1: Comparison of Single Centerline vs. Double Centerline Representation Schemes

	Advantageous Field	**Disadvantageous Field**
Single Centerline Scheme	network, routine, traffic flow analysis, macro-scope presentation	lack of local detail
Double Centerline Scheme	asset management, micro scope analysis, certain types of visual presentation	network analysis (e.g., routing)

Aware of the advantages and disadvantages of the two representation schemes, transportation agencies often have difficulty choosing scheme. Decisions made depend on many factors, e.g., data source, major applications of the system, and cost. For example, Manitoba Transportation and Government Services has developed three GIS-T base maps--two of them use the single centerline scheme (the first and second generation base map), one uses the double centerline scheme (the third generation base map).

Technical Approach

To take advantage of both representation schemes, interoperability between the two types of base maps is desirable. Although line matching can be established between a single centerline base map and a double centerline base map using buffering techniques, the efficiency of the process and the accuracy of the result is often hard to achieve when the two base maps are of different scales.

Since double centerline base maps are generally newer and have better spatial accuracy and a higher level of detail than single centerline base maps, it is preferable to derive new single centerlines from existing double centerline base maps to ensure better spatial data accuracy.

The technique developed in the research employs several in-house programs, and calculates the theoretical centerline from existing double centerlines in the base map. Figure 5.9 shows examples of the derivation. The double centerlines of the divided roadways of Trans Canada Highway 1 control section 1001290 are shown in the first map. Theoretical centerline based on calculations is added in the second map.

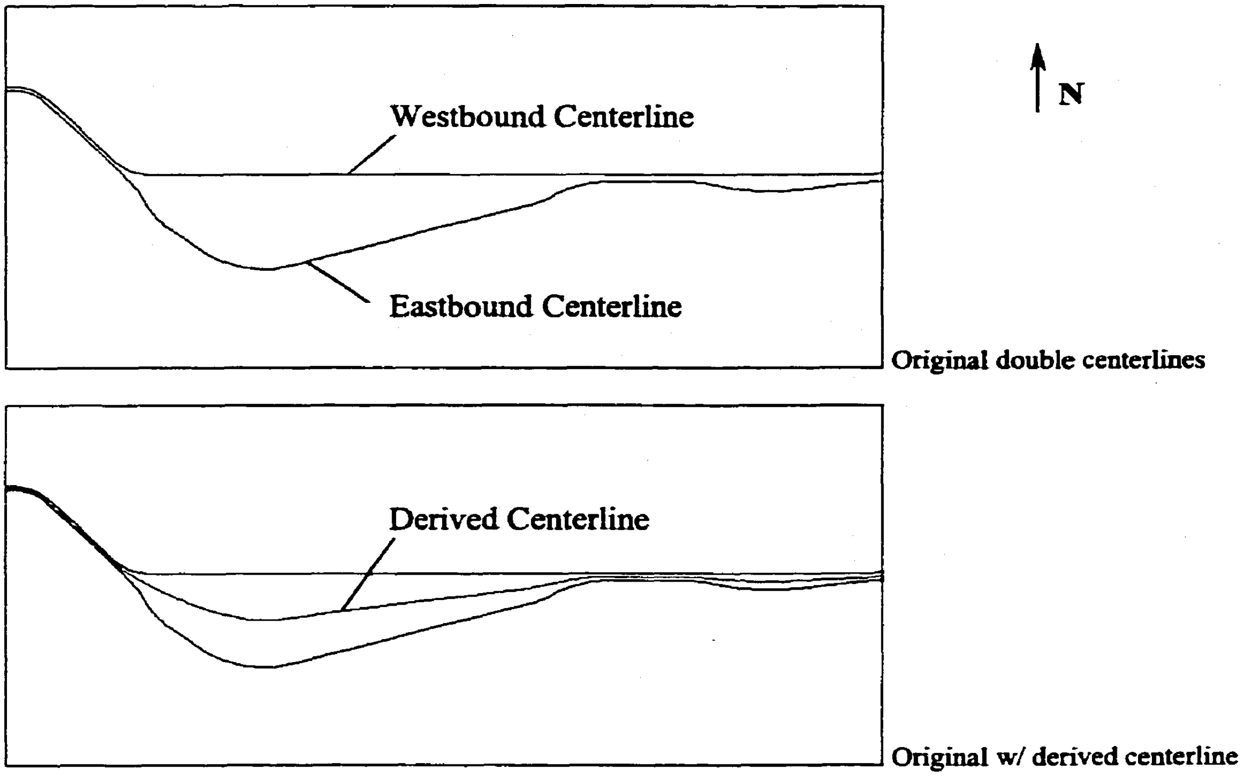


Figure 5.9: Theoretical Centerline Derivation

A discussion of the centerline derivation technique and in-house program is given in Appendix E.

5.5.2 Category II--Spatial Data Linking and Matching

The techniques developed are based on the geographical characteristics of spatial data, as well as buffering methodology and linear referencing. Individual techniques are developed to handle point features, line features, and polygon features respectively.

Point Features

Coordinate information of point features is used to create buffer zones and find the point features from the other data sets which have coordinates within the buffer zone. Figure 5.10 illustrates the application of this technique in the City of Saskatoon to automatically match intersection nodes with intersection unique ID labels residing in a different spatial data layer.

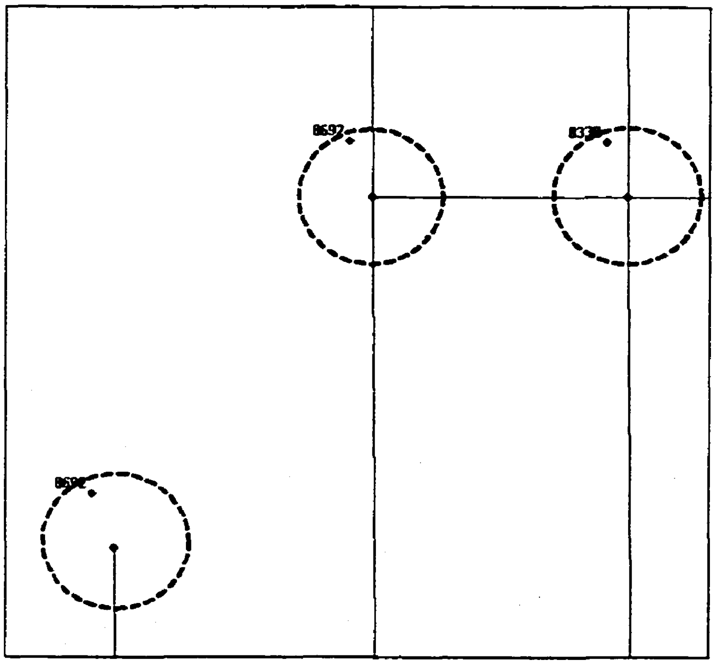


Figure 5.10: Matching Points Using Buffer Zones (Han et al., 2000)

Line Features

Buffer zones can be created around every line feature in one data set (illustrated in Figure 5.11), and matching lines from other data sets should fall within those buffer zones.

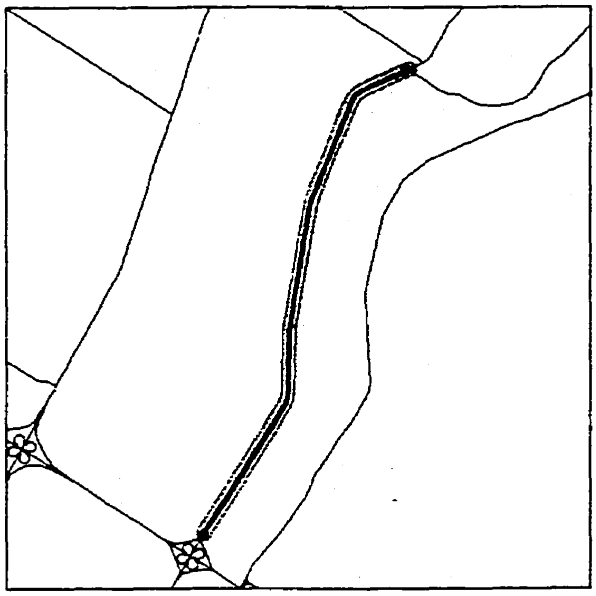


Figure 5.11: Buffer Zone Created around Lines (Han et al., 2000)

Polygon Features

Although the buffer zoning method can also be applied to polygon matching, it is much easier to match centroids. GIS software provides functions to convert polygons into centroids. Via the technique, the problem of matching polygons is converted into a point-matching problem.

5.5.3 Category III--Associating Attribute Data Using Linked Spatial Data

This category of techniques uses linked/enhanced spatial data to transfer attribute data between different GIS data sources. Some examples are cross-layer attribution, or dragging information across layers using linear referencing.

The following is an example of applying the linear referencing technique to the City of Saskatoon GIS map to tie unique IDs from one data set to the street sections data set. Figure 5.12 shows the situation of the two original data sets. In the process, point features containing unique ID information were found along a linear referenced route within a certain distance. After points were "dragged" to street sections based on the milepost given by linear referencing, unique ID's could be transferred to street centerline features. Consequently, through unique ID, other attribute data can be transferred as well to solve this problem which is essentially an engineering issue.

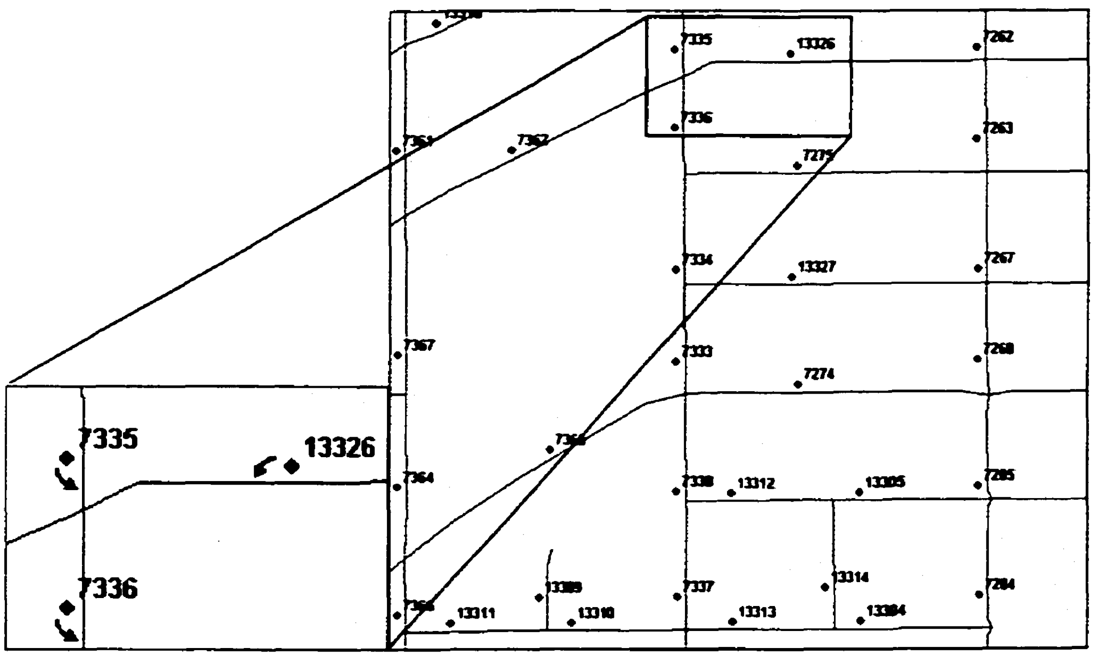


Figure 5.12: Matching City of Saskatoon Street Unique ID (Han et al., 2000)

5.6 Creating an Interoperable GIS-T Platform for Manitoba

By implementing these data sharing techniques, the research developed an interoperable GIS-T platform for Manitoba based on various data sources. The platform facilitates comprehensive transportation analysis and applications through improved interoperability and data sharing capability.

5.6.1 Objectives of the Interoperable GIS-T Platform

The platform covers road and railway facilities within the Province of Manitoba. For geographical coverage, the platform has a provincial highway component, a City of Winnipeg street component, and a rural municipality road component (Lew et al., 1999). Special attention is paid to the Manitoba Capital Region network.

The platform integrates spatial data from various sources: three versions of provincial highway base maps; the City of Winnipeg SCL base map; Natural Resources Canada's map fabric; ortho-aerial photos; rural municipality maps; engineering drawings; and GPS tracking.

The platform swaps and derives its spatial data from these data sources to ensure the best accuracy, level of detail, and currency available at the current time. Attribute information is also integrated into the platform after the interoperability techniques are applied.

The resulting platform accommodates multiple transportation applications (e.g., regional transportation planning (Lew, 2000), the provincial traffic monitoring system) by providing customized spatial data sets specifically directed toward individual applications. The spatial and attribute data sets are constructed to provide compatibility with current applications to eliminate extra modification on current GIS-T.

The platform is not static and has the ability to improve its spatial data quality whenever new (higher quality) spatial data become available.

5.6.2 Structure of the Interoperable GIS-T Platform

The interoperable GIS-T platform supports the Manitoba Transportation Analysis System (MTAS).

The Manitoba Transportation Analysis System (MTAS) is a collection of vector-based GIS networks and associated attribute databases being developed by UMTIG (Lew et al., 1999). The MTAS is designed as a system capable of integrating various transportation databases onto a common platform for analysis. Figure 5.13 illustrates the components of the MTAS. These components can be linked to existing spatial and attribute data developed by other organizations.

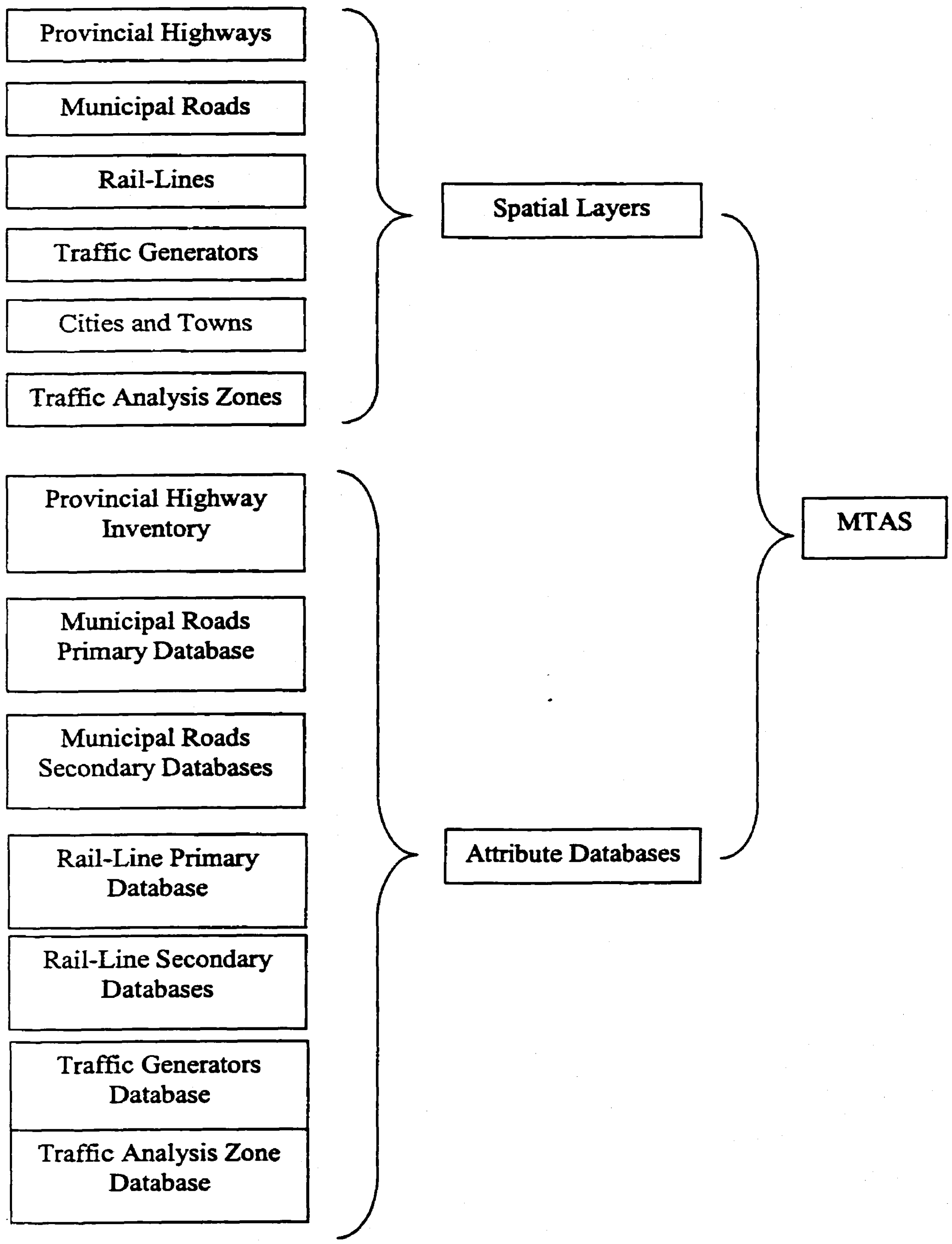


Figure 5.13: Components of MTAS (Lew et al., 1999)

As illustrated in Figure 5.13, the MTAS consists of spatial layers and attribute databases. The spatial layers are: (1) the Provincial Highways, (2) the municipal roads, (3) the railways, (4) the airports, (5) the ports, (6) the traffic generators, (7) the population centers, and (8) the traffic analysis zones for the Winnipeg capital region. The attribute databases are: (1) the provincial highway inventory database, (2) the municipal road primary database; (3) the municipal road secondary databases, which include the road construction and maintenance database and the municipal road traffic database; (4) the rail-line primary database; (5) the rail-line secondary databases, which include rail traffic and accidents; (6) the traffic generator attribute databases; and (7) the traffic analysis zone database.

To facilitate regional transportation analysis, MTAS provides connections to adjacent jurisdictions. For example, as illustrated in Figure 5.14, Manitoba provincial highways are connected with highways of the Prairie Provinces and States to construct a regional network demonstrating Spring Weight Restrictions across the Prairie region.

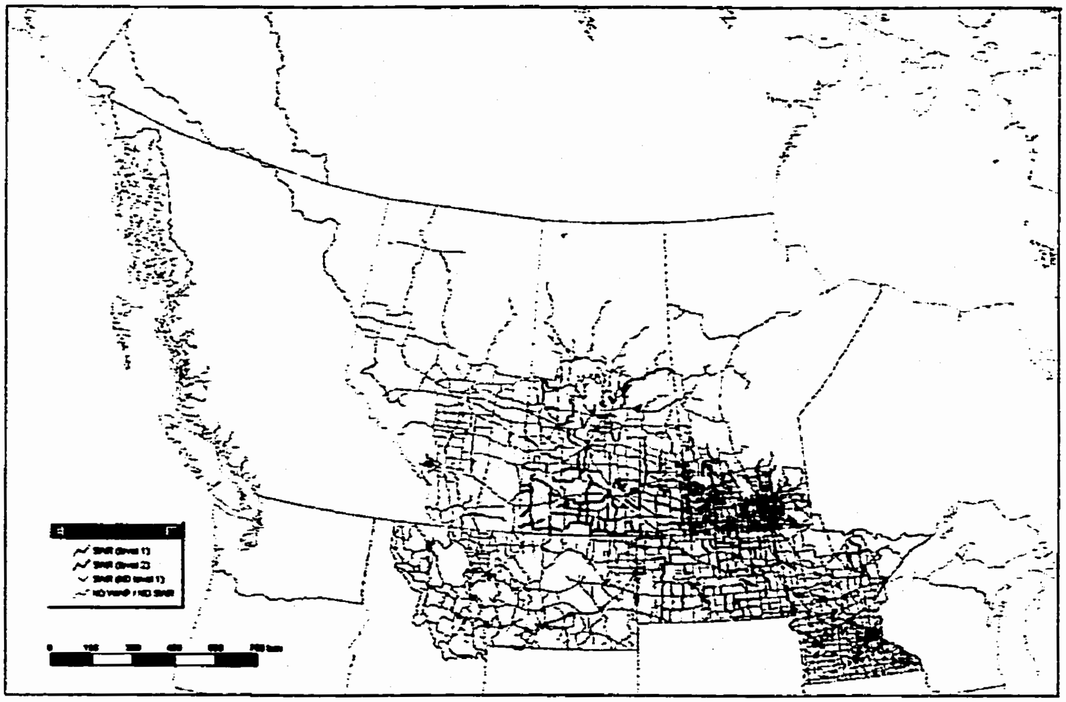


Figure 5.14: Regional Spring Weight Restrictions (Apr. 8 1999) (Montufar et al., 2000)

While the focus of the MTAS is on transportation engineering, other components that are relevant to municipal governments such as drains, culverts, gas lines and other structures can be developed and integrated to be compatible with the MTAS spatial layers. Non-transport information such as land use, grain production, and soil classification are often considered as supporting data, which can consequently be attached to transportation data.

5.6.3 Construction Processes

The platform consists of spatial data and attribute data. Both are constructed individually and assembled together specifically to meet the requirements of different GIS-T applications. For example, the resulting base map for the transportation planning model for Winnipeg (a centerline map) is different from the base map for the provincial highway asset management system in terms of its networking ability and level of generalization.

Spatial Data Construction

To create an interoperable platform, the research has merged spatial data from different sources. This ability to restructure spatial data makes it easy to use the best available pieces of data to create an integrated platform. Figure 5.15 shows an example of spatial data integration. In this example, street centerline data of the City of Winnipeg was integrated with the enhanced "new-new" provincial highway map at the city boundary. Additional ramp data was retrieved from Natural Resources Canada map fabric.

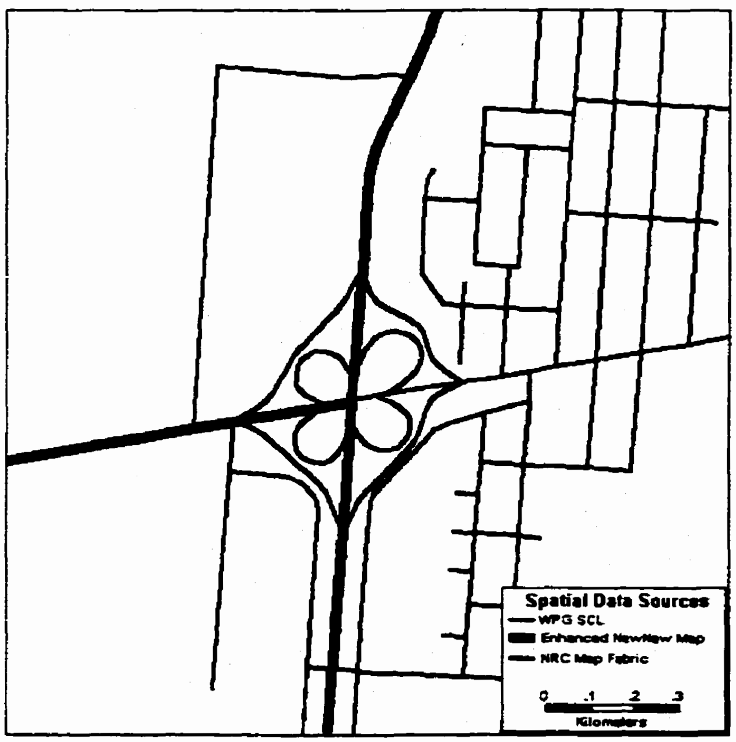


Figure 5.15: Example of Spatial Data Integration

In the merging process, spatial data sets are uncoupled from attribute data and treated as raw material. Spatial data from different sources are merged. The results for this process are networks ready to accept attribute information.

Attribute Data Preparation

Attribute data from different sources often contain property information regarding different aspects of the same spatial entities. To ensure providing a complete and accurate picture of the real world, attribute data need to be rearranged or restructured prior to the integration with interoperable spatial data sets.

Integration of Spatial Data and Attribute Data

Spatial and attribute data sets are merged together to create the GIS-T platform. The integrated platform differs in spatial data structure and attribute data coverage according to individual requirements of different GIS-T applications. However, interoperability between these versions of the platform is ensured by the linkages established internally.

5.6.4 Results

Two of the results of the integration are demonstrated below:

a) *GIS-T Platform for Manitoba Capital Region*. This platform has been used as the GIS platform for the new Manitoba Capital Region Transportation Planning Model (Lew, 2000). Figure 5.16 shows the road portion of the platform.

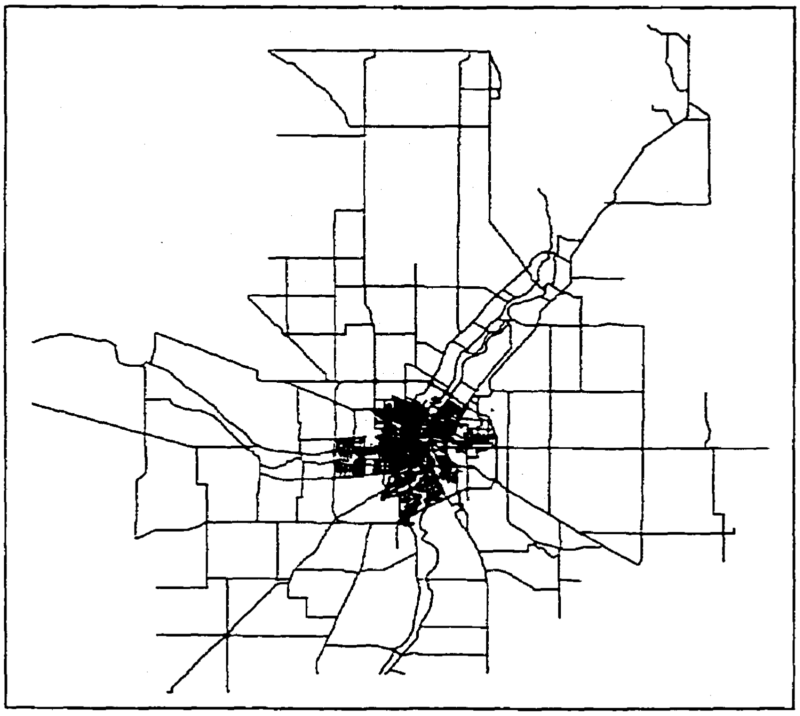


Figure 5.16: Manitoba Capital Region Transportation Network (Road Portion)

b) *GIS-T Platform for Manitoba Transportation and Government Services (MT&GS)*. This research involved the development of a platform including components of provincial highways, city streets, RM roads, and railways. The base maps are designed to support four types of applications--transportation planning, asset management, data exchange, and presentation. Figure 5.17 gives some examples of spatial data components of the platform developed during the research.

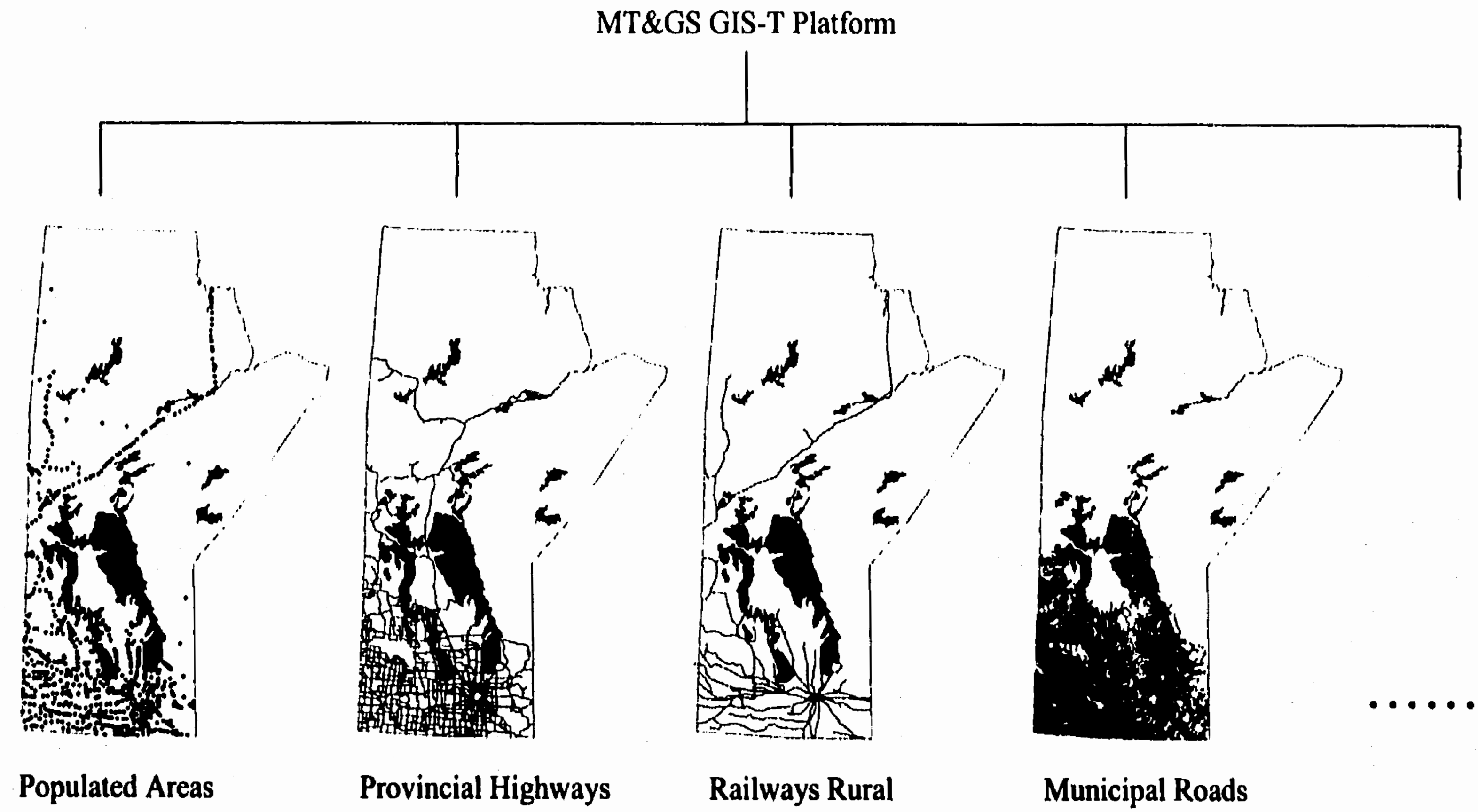


Figure 5.17: Components of MT&GS GIS-T Platform

6.0 CONCLUSION AND RECOMMENDATIONS

This chapter provides the conclusion of the thesis and discusses further research opportunities.

6.1 Conclusion

The research established a framework and developed associated techniques with a focus on spatial data, thereby facilitated total GIS-T interoperability. By applying the framework and techniques, an interoperable GIS-T platform is developed for Manitoba transportation applications. The research also provides better understanding of interoperability issues and enhances data sharing capability in GIS-T field.

6.2 Further Enhancements

6.2.1 Expansion of the Framework

While the theoretical framework developed in the thesis focuses on transportation applications, the framework has potential usage in a variety of fields (i.e., environmental research, agricultural planning). The structure of the framework can be adapted to fit the desired application.

The importance of the framework is that it provides interoperability, which lets users to focus on their prime tasks (e.g., design, analysis, evaluation) instead of spending large amounts of time in specific data sharing processes.

6.2.2 Development of Techniques

The framework can be further enhanced through incremental development of the data sharing techniques. Techniques involved in the research were developed specifically for the GIS-T platform for Manitoba. New techniques can be added to enable and enhance other data sharing processes. The techniques developed in the research can also be modified to accommodate different applications and provide better user-friendly interfaces.

Further development of the techniques will open new opportunities for broader data sharing. For example, techniques can be developed to link spatial features of different types (e.g., linking point features to line features) to encourage data sharing in a wider range.

6.2.3 Dynamic GIS-T Platform

The interoperable platform has the potential to support more diverse applications in Winnipeg, Manitoba, Canada, U.S. and many other places by dynamically deriving customized spatial data sets to meet specific user requirements.

The diversity of GIS-T applications prevents the use of a common set of spatial data for different applications. One solution is to provide derived spatial data sets to multiple GIS-T applications while maintaining interoperability. This allows more applications to share spatial data through the interoperable platform. The advantages of this solution include:

- facilitating access to and use of more information sources
- providing more transportation application friendly spatial data structure
- better updating of spatial data to give a more accurate picture of the world

6.3 Recommendations

6.3.1 Broader Implementation of the Framework

Implementing the framework for applications outside the transportation field will facilitate spatial data sharing by avoiding duplicate efforts and encouraging new product development.

6.3.2 Integration into GIS Software

The benefit of the framework and interoperable platform will become more obvious when more applications adopt the data sharing framework in various fields. With this much broader user base, it will be feasible and desirable to integrate the framework methodology and its associated techniques into GIS software suites to assist users in the whole GIS community.

6.3.3 Data Sharing Standard

Standardized spatial data facilitates information exchange and information dissemination. The open architecture of the research framework can be incorporated into data sharing standards, which ensure the compliant GIS-T spatial data have the best possible accuracy and currency, and a suitable level of generalization. Compatibility and interoperability are guaranteed by different data sources and GIS-T applications when this common standard is met. Spatial data sets complying with a common standard can be specifically handled and supported by various computer platforms and are interoperable among different GIS applications. Such interoperability standardization is of crucial importance for future ITS applications, where extensive data exchange will take place at all levels of organizations (Latham et al., 1997).

REFERENCES

Alam, M., M. Melchiorre, K. Han, and A. Clayton (1998), *City of Winnipeg Traffic Information System Development Phase II*, University of Manitoba Transport Information Group, October 1998.

Anderson, J. and E. Mikhail (1998), *Surveying: Theory and Practice (Seventh Edition)*, WCB/McGraw-Hill.

Association for Geographic Information (AGI) (1999), *GIS Dictionary*, Web site: http://www.geo.ed.ac.uk/agidict/welcome.html

Barnes, S. (2000), *Mine Your Map Data*, CADLYST, September, 2000, pp. 66-71.

Bastiaansen, A. (1999), *Digital Maps and Their Vital Role in ITS Applications*, Proceedings of the 6th ITS Congress, Toronto, November 8-12, 1999.

Beskpalko, S., M. Wyman, and J. Sutton (1997), *The Need for a Formal GIS Transportation Model*, Proceedings of International Conference and Workshop on Interoperating Geographic Information Systems, Santa Barbara CA, December 3-6, 1997, Web site: http://www.ncgia.ucsb.edu/conf/interop97/program/papers/bespalko.html

Bureau of Transportation Statistics (BTS), U.S. Department of Transportation (USDOT) (1997), *Transportation in the United States: A Review*, Web site: http://www.bts.gov/programs/transtu/titus/titus97t.pdf

Department of the Interior (1918), *Manual of Instructions for the Survey of Dominion Lands*, Ottawa.

Dias, W. (2001), *GPS*, IBM GIS Dept., Web site: http://giswww.pok.ibm.com/gps/

Dick, C. and A. Clayton (2001), *Impact of New Technology on Grain Transportation in Western Canada*, Preprint CD-ROM, 80th Annual Meeting, Transportation Research Board.

Dueker, K. and J. A. Butler (2000), *GIS-T Data Sharing Issues*, Preprint CD-ROM, 79th Annual Meeting, Transportation Research Board.

Earth Resource Mapping (2001), *USA DEMs*, Web site: http://www.ermapper.com/datasales/usadems/usadems_body.htm

ESRI (2001), *ESRI Library--GIS Glossary*, Web site: http://www.esri.com/library/glossary/e_h.html#GIS

Federal Geographic Data Committee (FGDC) (2001), *FGDC Metadata*, Web site: http://www.fgdc.gov/metadata/metadata.html

George, M. (1991), *Transportation GIS Applications*, Proceedings of the 1991 Geographic Information Systems (GIS) for Transportation Symposium, Orlando, Florida, March 1991.

Goodchild, M., M. Egenhofer, and R. Fegeas (1997), *Report of International Conference and Workshop on Interoperating Geographic Information Systems*, 1997, Web site: http://www.ncgia.ucsb.edu/conf/interop97/report.html

Goodwin, C. (1996), *WHITE PAPER Location Referencing for ITS*, Oak Ridge National Laboratory, Web site: http://itsdelpoyment.ed.ornl.gov/spatial/

Han, K., E. Tan, K. Lew, J. Montufar, and A. Clayton (2000), *Techniques to Facilitate Interoperation between Different Spatial Data Sets in GIS-T*, 2000 Geographic

Information Systems for Transportation Symposium, Minneapolis, Minnesota, March 2000.

Heywood, I., S. Cornelius, and S. Carver (1998), *An Introduction to Geographical Information Systems*, Longman.

Intergraph (1998a), *MGE Segment Manager (MGSM) User's Guide*, Intergraph Corporation, Huntsville, Alabama, USA.

Intergraph (1998b), *Working with GeoMedia*, Intergraph Corporation, Huntsville, Alabama, USA.

Kennedy, J., A. Shalaby, and R. Cauwenberghe (2000), *Dynamic Segmentation of Pavement Surface Condition Data*, Proceedings of the 3rd Transportation Specialty Conference of the Canadian Society for Civil Engineering, London, Ontario, 7-10 June, 2000, pp. 41-47.

Latham, F., D. Siegel, and D. Xiong (1997), *Real-time Data Exchange and Interoperability*, Proceedings of International Conference and Workshop on Interoperating Geographic Information Systems, Santa Barbara CA, December 3-6, 1997.

Lee, S. (2000), *Serious Java Programming For The Wireless World*, GeoCommunity Spatial News, Web site: http://www.spatialnews.com/features/javagis/kvm.html

Lehtonen, M. and Lähesmaa, J. (1999), *Multimodal Transport Information System Based on Geographical Information System for Transportation*, Proceedings of the 6th ITS Congress, Toronto, November 8-12, 1999.

Lew, K., K. Han, J. Montufar, J. Drury, E. Tan, and A. Clayton (1999), *Development of a Road Network Management System for Rural Municipalities: Pilot Project*, Technical Report, University of Manitoba Transport Information Group, September 1999.

Lew, K. (2000), *The Foundation of a New Transportation Planning Model for the Manitoba Capital Region*, M. Sc. Thesis, University of Manitoba, December 2000.

Lopez, R. (2000), *Civil Users to Get Military GPS*, Flight International, 9-15 May 2000, pp. 13.

Manheim, M. (1979), *Fundamentals of Transportation Systems Analysis, Volume 1: Basic Concepts*, The MIT Press.

Microsoft (1998), *Microsoft Press Computer and Internet Dictionary*, 1998 Microsoft Corporation.

Minty, S., A. Clayton, and K. Han (2001), *The Use of Specialized Computer Software in Transportation--Interoperability Issues in GIS-T*, Proceeding of 2001 Canadian Institute of Transportation Engineers Conference, Calgary, May 2001.

Montufar J. and R. Sanderson (1999), *Analysis of Heavy Truck Safety in Manitoba and Saskatchewan Using GIS Capabilities*, Proceedings of Canadian Multidisciplinary Road Safety Conference, Halifax, Nova Scotia, May 1999.

Montufar, J., A. Clayton, and K. Mruss (1999), *Issues Relevant to the Strategy of Linking Databases for Use in Road Safety Research*, University of Manitoba Transport Information Group, June 1999.

Montufar, J., K. Han, K. Lew, A. Clayton, R. McGregor, R. Sims, D. Palsat (2000), *Study for the Harmonization of Spring Weight Restrictions and Winter Weight Premiums for Roads in the Prairie Region, Tasks 1 & 4*, EBA Engineering Consultants, December 2000.

National Center for Geographic Information & Analysis (NCGIA) (1997), *Interoperating Geographic Information Systems*, Web site: http://www.ncgia.ucsb.edu/conf/interop97/i20prop/i20prop.html

Noronha, V. (1997), *Interoperability Issues in Intelligent Transportation Systems: Testing the Cross Streets Profile*, Proceedings of International Conference and Workshop on Interoperating Geographic Information Systems, Santa Barbara CA, December 3-6, 1997, Web site: http://www.ncgia.ucsb.edu/conf/interop97/program/papers/noronha.html

Nystuen, J., A. Frank, and L. Frank (1997), *Assessing Topological Similarity of Spatial Networks*, Proceedings of International Conference and Workshop on Interoperating Geographic Information Systems, Santa Barbara CA, December 3-6, 1997.

Open GIS Consortium (OGC) (2001), *OGC*, Web site: http://www.opengis.org/

Pierowicz, J. and M. Roser (1999), *Testing of an On-board Unsignalized Intersection Violation Prevention System*, Proceedings of the 6th ITS Congress, Toronto, November 8-12, 1999.

Price, W., A. Clayton, J. Montufar, and L. Escobar (1999), *Oversize Trucking Policy on Manitoba Highways: Evaluation of the Friday Restriction*, University of Manitoba Transport Information Group, January 1999.

University of Manitoba Transport Information Group (UMTIG) (2000), *Traffic on Manitoba Highways 1999*, University of Manitoba Transport Information Group, June 2000.

U.S. Department of Transportation (USDOT) (1999), *Intelligent Transportation Systems Awareness CD-ROM Version 2.0*.

U.S. Department of Transportation (USDOT) (2001), Joint Program Office Intelligent Transportation Systems, *National Architecture for ITS*, Web site: http://www.its.dot.gov/arch/brochure.htm

Wyman, M. (2000), *Advanced GPS, GPS/GIS/LRS Integration, and the Dynamic Location Data Model*, Workshop for 2000 Geographic Information Systems for Transportation Symposium, Minneapolis, Minnesota, March 2000.

Xiong, D., S. Gordon, and O. Franzese (1999), *Interoperating Spatial Data for ITS: Problems and Alternative Solutions*, 1999 Geographic Information Systems for Transportation Symposium, San Diego, California, March 1999.

APPENDICES

APPENDIX A GIS SOFTWARE

As an integral component of GIS, GIS software assumes responsibilities of interacting with different hardware and software components of GIS, coordinating their work, interfacing with users, querying databases conventionally or spatially, and automating the workflow to facilitate the users. GIS software also takes care of the need for creation, maintenance, dissemination, and storage of spatial and attribute data.

There are many GIS software packages available. It is essential for users to know what GIS software can do for them, as well as the limit of its ability. A major consideration of this research is the functions that various GIS software provides to users. Five categories of GIS software are identified based on functions provided and level of users.

Group I is the full-feature GIS software. Examples are ARC/INFO, MGE (Modular GIS Environment), and GeoMedia Professional. They are targeted at higher-level users engaged in sophisticated analysis and the development of spatial data for lower-level GIS users.

They are often in a modular form to provide large selections of GIS-related functions. Each module or function is independent, which helps make these modules more efficient to run, more powerful in terms of data handling ability, and more flexible to fit into different situations. However, this characteristic also makes this type of software difficult to learn and to get familiar with.

This group of software can be seen as toolboxes containing powerful GIS tools which give users the most freedom to do analysis, data management and integration. However, they are relatively expensive, and require substantial computing power and GIS specialists to operate.

Group II is the most commonly used GIS software. It has the broadest user base. Examples are ArcView, MapInfo, GeoMedia, and Maptitude.

This group of software packages the most commonly used GIS functions, and eliminates less used ones to create integrated GIS environments for a broad spectrum of users. The integrated environment makes the software easy to learn and use, and yet versatile enough to allow users with ordinary computers to do routine GIS analysis and applications. They are fairly flexible in accessing different spatial data formats. This group also has better map generating and printing capability than Group I software. They often run on desktop computers and are more user-friendly than the first group.

Group III is specialized GIS software for individual application fields. For example, TransCAD for transportation applications, FRAMME for utility industry applications, ER Mapper for image processing applications.

This group of software do not have as many functions as Group I, and are less capable of dealing with spatial data and automated processes. But they have some built-in functions to handle special demands posed by different application fields. For example, TransCAD has integrated linear referencing, network routing, transportation planning, and logistic tools to support transportation planning and analysis. FRAMME provides functions such as planning electrical power lines, outage management, and dispatching emergency responding staff to help utility companies to manage their power network.

Group IV comes from CAD software, and only came to the scene for a short period of time. Examples are "AutoCAD Map" and "MicroStation Geogrphics". They are basically CAD software with extensions of common functions for GIS usage. They share the same Graphical User Interface (GUI) with their CAD counterpart, and heavily emphasize compatibility with existing large amounts of CAD drawing data.

Aside from their unmatched compatibility with drawings in CAD format, this group of software inherits strong editing capability from the original CAD software. This makes them ideal tools to handle GIS interoperability at the spatial data level. However, since they are new to the GIS field, the GIS-related functions they offer are rather common and general-purpose. They do not support GIS analysis as well as the above three groups. But they have the potential to become good GIS software in the future.

Group V is new and accompanies the rapidly growing field of Internet applications. Examples of Internet GIS servers are ArcIMS, GeoMedia Web Map, MapXtreme, and MapGuide. Examples of client GIS software are ArcPad and OnSite running on hand-held computers (Barnes, 2000).

Since the full-functional GIS is considered too complicated for client-end browsers or handheld devices, the workload is split between the server computer and the client computer. The map generation is done on the server side and resulting map is sent to the client computer through the Internet. On the client side, the viewer can browse through maps using an ordinary Web browser (i.e., Netscape and Internet Explorer) with certain downloadable plug-in features installed. Through a wireless Internet connection, handheld computers (e.g., Palm PDA and Windows CE PDA) can also load and view maps with very limited computing capability (Lee, 2000). In the near future, with the help of much higher speed Internet connection, map-viewing functions will be operated on a simple client device, such as cellular phones, which is often referred to as thin client due to its limited computing power.

APPENDIX B TRANSPORTATION APPLICATIONS OF GIS

Since GIS is a useful tool for comprehensive analysis and decision making, there are many GIS application fields in the transportation realm.

Transportation Planning

Transportation planning involves correlating factors such as land use, environmental concerns, economic development, and components of transportation systems. With its data integration ability, GIS has become a useful tool for the planning process. To further support transportation planning, several GIS are customized to provide sophisticated tools for planning/analysis/engineering, such as network modeling based on a GIS platform.

Transportation Asset Management

Transportation infrastructure includes a wide spectrum of facilities (highways, bridges, airports, harbors, canals, etc.). GIS help to store, manage, and visualize information about these transportation assets, which are spread over space and which interact over time. Sophisticated management, such as geographical distribution analysis, and network balancing can be supported through GIS functions.

Construction

GIS helps capture and manage the geographical information of an engineering project and keep the information up to date. Through the integration of data from different sources (e.g., aerial photos, space imagery, inventories, and surveys), construction plan, right-of-way management, and project management can be done more effectively with GIS.

Traffic Monitoring

GIS is often used to store, analyze, and display traffic information, monitor traffic operations, manage congestion, make counting plans, and also deliver traffic information for design and mapping purposes.

Safety Engineering and Management

GIS helps to uncover geographical distribution of accidents, locate potential hazard sections, intersections, and railway crossings.

Oversize/Overweight Permitting and Enforcement

Regulations and restrictions can be stored in GIS database, and conveniently retrieved, updated, and distributed. The built-in network analysis capability of GIS helps generate alternative routes and calculate fees. Enforcement officers can also use GIS to keep track of oversize/overweight vehicles.

Transit Service

Public transit uses GIS to integrate data like demographics, bus ridership, peak-hour traffic, etc. from various sources. This integration greatly helps transit planners make schedules, plan for new routes, and adjust fleet mix.

GIS will help improve the efficiency of the system and provide better customer service. GIS helps improve the communication between a dispatch office and vehicles and eliminates communication mistakes. The ability to monitor bus location and schedule status from a central dispatch office improves the quality of the service and facilitates more informed decisions on how to allocate resources and deliver service.

Emergency Response

For departments like E911, every minute counts. Coupled with GPS/DGPS, good up-to-date GIS saves time in automatic addressing and real-time network routing. GIS is essential for computer aided dispatch. A complete computer-aided dispatch system always has a sophisticated GIS component.

New technology advances come every day. Wireless location technology, like using cellular phone to locate a vehicle in real-time, and providing navigation information by voice interface and the proliferation of GPS devices offers a lot of opportunities for new GIS applications in transportation fields. Examples include vehicle navigation, on-board travel information systems, and fleet monitoring systems.

In incident response, GIS coupled with automatic vehicle location can help to bring emergency vehicles to incidents quickly and efficiently. GIS-based systems allow dispatchers to visually display the map and quickly find the closest available unit to respond to a call, provide the shortest route to get the site, view all vehicles as they travel or emergency routes, evaluate the route's efficiency and adjust directions to accommodate traffic conditions, and effectively coordinate emergency efforts with other agencies.

Private Sector

With the help of GIS, private transportation companies can plan their routes, track their equipment, and manage their fleet and infrastructure. As more and more geo-spatial data are made available to public, applications of GIS in the private transportation sector will continue to grow rapidly.

The recent (May 1, 2000) termination of the Selective Availability (SA) makes the GPS more accurate and more stable. This greatly helps the development of GPS/GIS in transportation areas. GPS+GIS can be used for route guidance, MAYDAY, and fleet tracking purposes (Pierowicz and Roser, 1999).

Automatic vehicle location uses global positioning systems to pinpoint the precise location of vehicles. This satellite-based technology provides real-time location, latitude and longitude coordinates, and direction of travel.

Intelligent Transportation Systems (ITS)

ITS are the integrated application of advanced sensor, computer electronics, and communications technologies and management strategies to provide traveler information and increase the safety and efficiency of the surface transportation system (USDOT, 1999).

ITS is all about information--the collection, sharing, processing, and redistribution of information--to move people and goods more efficiently. Information lets travelers make better decisions, and helps improve the efficiency and safety of the various elements of our surface transportation infrastructure: for example, transit systems, freeways, toll facilities, rail intersections, truck regulatory facilities/services, and rural roadways (USDOT, 2001).

With the development of ITS, a large number of sensors and monitors are being installed. Inbound data flow is enormous. To efficiently retrieve information from these data, a suitable GIS platform is needed. Through efficient visual representation, GIS can be a bridge between data and end-users. The success of many ITS applications hinges on the ability to communicate a location message unambiguously across dissimilar map bases.

Information sharing and communication are intensively employed in Intelligent transport systems (ITS) to increase the efficiency and level of service of the transport system in rural areas, reduce the need to travel long distances and combine modes of transport in order to produce efficient multi-modal transport chains. In urban areas, ITS enhances the capacity of congested networks and provides a means to manage traffic, thus maximizing flow throughput and minimizing environmental impacts.

Traveler Information System

GIS helps to find the precise location of incidents and consequent congestions, and identifies which alternative route traffic is to be diverted to. With good information about weather, road condition, and/or incidents, travelers can plan their route, determine when to take the trip, and in case of incidents, make an informed decision, either delay their trip or take an alternative route--resulting in less congestion which is easier to manage.

Incident-related congestion quickly worsens traffic conditions on highways, causing delay to travelers and commercial vehicles and increasing fuel consumption. Incident management systems combine technologies such as embedded sensors, video cameras, and variable message signs to help traffic control centers detect slowdowns caused by accidents and initiate a response promptly. With the help of GIS, they can also verify incident sites and determine the appropriate emergency resources required, dispatch resources and provide information to motorists.

Information Carrier

Another transportation field where GIS plays an important role is information dissemination. Using GIS, transportation information can be distributed throughout transportation agencies and to the public.

Real-time traffic information can be provided to drivers in the form of maps through advanced traveler information systems to help them navigate their trip, and route around certain congestion and/or subject to restrictions.

Travelers can connect to a service center through wireless connection, and request dynamic information integrated with a digital map (Bastiaansen, 1999).

The upcoming on-board driving assistant system will use GIS as an information carrier. One example is "map chunks" on demand. In this scenario, the user can download on demand the geo-spatial data for an area of interest. Typically, the downloaded data are stored on some hand held device or appliance, such as a Palm Pilot. The user then uses a local application to view and query the geo-spatial data. Newer generation products now also allow the user in the field to “red line” new information on the displayed map, and then upload the new data to the enterprise database.

APPENDIX C GIS-T BASE MAPS IN MANITOBA

The purpose of this document is to provide an overview of the GIS-T base maps developed in Manitoba.

Government transportation agencies (Province of Manitoba, City of Winnipeg, and RMs), universities, and industry have long recognized the value and importance of GIS and explored its application in transportation.

Many efforts have been made through the years. Several GIS-T have been developed and used to varying degrees. Due to improvements in mapping technology, base maps that are developed for new GIS-T are often more accurate and contain more detail (e.g., more shape nodes to better depict the roadway alignment, double centerlines for divided highways, and ramps for interchange).

However, these GIS-T have been developed separately, based on different, often partial spatial data sets. Even though the objects they are representing are the same (e.g., highway network and street network), the spatial data sets they work on (base maps) are structured, projected, and disaggregated differently, so that convenient data sharing is prohibited.

The specifications (i.e., accuracy, data structure, level of detail, and topology) for the base map of GIS-T depend on many factors: cost of acquiring spatial data (still expensive in Canada); purpose of the GIS-T (network planning, pavement management, or visual presentation); and application requirements (two centerline vs. AB/BA representing scheme).

The development of GIS-T usually goes through different stages, which typically involve establishing the basic transportation network base map. The next stage is attaching relevant feature class information layers to the base map; this includes other transportation facilities, social economical information, and geographical/mapping related information. The next step is attributing the base map and relevant feature class layer through the development of appropriate databases. Based on this fundamental information, transportation applications such as a planning model can be developed, and analysis and various studies can be conducted.

Since this development always take a long time, spatial data in the base map can become outdated within the life cycle of the GIS-T. For example, road construction, roadway realignment, and changes of authority can affect the basic network dramatically. Mapping technology also provides opportunities for more accurate, more affordable, and more detailed spatial data sources. To prevent the premature obsolescence of the spatial data, and to take advantage of new spatial data sources, interoperability of GIS-T is highly desirable.

While many GIS-T have been developed, certain data regarding the transportation system are possessed and maintained by the systems individually. To get a systematic view of every aspect of the subject, a decision maker, engineer, planner, or manager needs to gather GIS-T information from various sources and put it into context. Data sharing between GIS-T is increasingly important. Interoperability of the spatial data sets is of increasing concern.

Some efforts have been made to share GIS-T information in the province and the urban center and rural municipalities. The University of Manitoba Transport Information Group (UMTIG) have tried to sharing GIS-T data in the heavy truck accident analysis project, the RM road project, and the WWP and SWR project, and later joined the city street network with provincial highways for the transportation planning model in the Manitoba capital region project. In these projects, various GIS software is used and techniques are developed to address individual problems faced by UMTIG.

In hindsight, we conclude that it is not reasonable and not feasible to serve many purposes of GIS-T by one single base map. Interoperability is inevitable. We have to deal with several GIS-T at the same time. To make the most efficient use of every individual system and retain the data sharing capability are our goals. A systematic approach is ideal. Under a conceptual framework, several versions of the same map work in synergy, and fulfill the individual task they are designed to do. While they all have been derived from the same spatial data source, they have an inherent linkage between each other. Results and other types of information can flow freely back and forth between base maps. At the same time, techniques developed under the framework will help bring in useful information from other GIS-T sources. And these techniques will also help in the timely updating of the base map, and to take advantage of more accurate spatial data while keeping the existing attribute data intact or compatible.

The usages and problems regarding the data sharing ability of the existing GIS-T base map in Manitoba are discussed in detail as follows:

First Generation Highway Map

This map is the earliest GIS-T base map developed to cover the Manitoba provincial highway network. It is still being used by Manitoba Highway Traffic Information System as the main GIS platform. Its spatial data set has been customized to represent highway traffic flow and an associated traffic counting program. Attached attribute data contain highway functional class, roadway characteristics, and traffic statistics, including historical data. The spatial data set and attribute data are constantly being updated by UMTIG.

Some of the characteristics of this base map are:

- Since the base map was developed in the early age of GIS-T in Manitoba, the spatial data set was imported from the then existing CAD drawing used by Manitoba Highways' drafting department.
- Due to certain projection problems associated with the original CAD drawing, the spatial features in the base map are not correctly located. This mis-locationing creates problems when other geographical information sources are to be connected. For example, the first generation map cannot be projected with maps from adjacent jurisdictions. (as illustrated in Figure C.1)

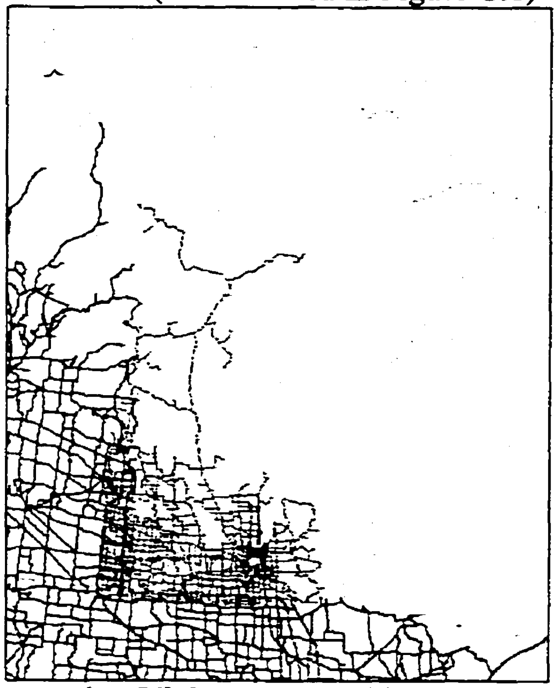


Figure C.1: The First Generation Highway Map with Maps from Adjacent Jurisdictions (Han et al., 2000)

- This base map is a center line map, i.e., the spatial data feature representing highways are lines tracing the center line of the roadway sections--no matter whether they are divided or undivided highways. These linear features are segmented according to highway control sections. To more accurately reflect traffic conditions, UMTIG enhanced the base map by segmenting the control section into sub-sections, called control sequence, which facilitate the assignment of traffic.
- Although this base map is constructed to support network-based abilities, such as routing and linear referencing, these abilities have not been pursued due to the fact that accuracy will suffer from the incorrectly located spatial data of the base map.
- This base map covers only provincial highways and some access roads under Highways' jurisdiction. Due to its projection problem, linking this base map to other GIS-T spatial data has proved difficult.

- In conjunction with this base map, associated spatial data sets representing rivers, lakes, towns, highway zones, traffic counting stations, and highway labels are also developed to support the presentation of traffic information.

Second Generation Highway Map

The next provincial highway base map after the first generation map was developed solely for GIS-T purposes. This base map is correctly projected to NAD 83 Datum. This ensured that the base map is compatible with maps from adjacent jurisdictions, and other correctly projected spatial data sources, e.g., land use, urban streets, and railway lines. The comparison of location accuracy of the second and the first generation map is shown in Figure C.2.

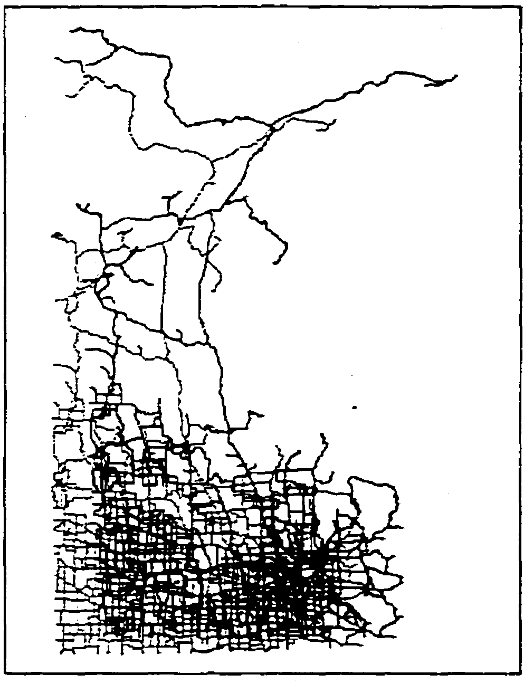


Figure C.2: The Second vs. the First Generation Highway Map (Han et al., 2000)

The characteristics of this second generation highway map are as follows:

- This version of the highway base map is a centerline map.
- The structure of the linear features in the base map is based on highway control sections. The segmentation of centerlines strictly follows the highway control sections. This practice facilitates certain applications, such as linear referencing and pavement asset management. However, network ability is sacrificed. Since highway control stations often run through several intersections with crossing highways, and do not stop, the topology of the linear features in the base map with this structure do not have a connecting point at intersections. This

connectivity problem prohibits true network applications, such as network routing, from being developed on this base map.

- This base map is linear referenced across the province. This feature gives users the ability to attach 1-dimensional data (mile post) to the GIS-T. For example, highway pavement rating and highway accidents can be attached along the highway.
- This base map also only covers provincial highways and some access roads under Highways jurisdiction. There is no spatial data representing the street networks inside urban areas. This eliminates the ability to conduct integrated analysis in areas where urban and rural traffic are both under concern.
- This base map is not updated after it was developed. Some of the newly constructed highways and highway realignments are not reflected in the spatial data set. And some control section assignment changes have not been updated.
- Since no associated spatial feature layers (such as label, river, and lake layers) are developed to support this base map, and layers from other sources are not fully compatible in terms of scale and level of detail, it is difficult to use this base map to produce large scale maps, which are very useful to present details of traffic flows on provincial highways.
- However, this base map can be use for certain purposes. For example, its linear referencing ability can work as a bridge to attach tabular information to geographical locations to be analyzed with other spatial data and information.

The Third Generation Highway Map

The third generation map is the latest endeavor of Manitoba Transportation and Government Services (MT&GS) to establish a common platform for all the GIS-T applications. In this version of the base map, highway alignments are updated, the accuracy of the linear features is enhanced, and more details of the highway network are added. However, due to the changes made in the structures of the spatial data set, this third generation base map is not compatible with either of the two preceding base maps. Existing information cannot be easily transferred to this new base map. That hinders the usage of the third generation map.

The characteristics of the third generation base map are as follows:

- This version of the highway base map has a higher level of detail than former versions. Divided highways are represented as double centerline, each line representing one side of roadway. Interchange ramps are planned to be added.
- Linear features of the spatial data are also based on highway control sections. To facilitate applications such as a pavement management system, sub-sections are introduced to long control sections. But this practice does not render fully network connectivity. And double center lines make routing more complicated and direction sensitive.

- Although the third generation map is quite current, its spatial data set still suffers from lack of flexibility in data sharing. The latest highway constructions are not reflected in the data set.
- The structure of this third generation base map is designed to facilitate applications that are centralized around facilities, such as the bridge asset management system. And its high level-of-detail helps the visual presentation of the information. However, for applications like transportation planning, routing, and traffic monitoring, the structure is found cumbersome, sometimes even prohibitive to be used.

NRCan Map Fabric

Natural Resources Canada (NRCan) is conducting an ambitious 10-year project to create a digital map for the entire province. Road fabric is a major portion of the on-going project. So far the map fabric covers most of Southern Manitoba, and the coverage is expanding to the north. Strictly speaking, NRCan road fabric is not a GIS-T base map. It is not constructed as a network, and attributing is solely based on air-photo readings. However, it can serve as a good source for spatial data in regards to various road networks, including provincial highways, urban streets, rural municipal roads, and railway lines.

The characteristics of the NRCan fabric are as follows:

- The NRCan fabric is very accurate in most places compared to the existing GIS-T base maps covering the province or the city. Only Highways' third generation map has similar accuracy.
- The fabric has a rich collection of spatial data. It contains not only provincial highways, but also urban streets, back lanes, cul-de-sacs, driveways, rural municipal roads, and railway lines. Details, such as interchange ramps, turning lanes, divided highways, and railway spurs are reflected in the fabric.
- However, the fabric does not cover the whole province. Large areas in the east and north of the province are not covered.
- Another problem of the fabric is that it is based on a 1996 air-photo set. Newly built or realigned roads have not been updated.
- The digitizing process of the fabric is not designed for transportation applications. Linear features are constructed without regard to network requirements. For example, digitized highway centerlines are segmented at the edge of air photos as opposed to highway intersections or highway control section. This makes it difficult to attach transportation information to these spatial data. Therefore, the original spatial data set has to be modified to accommodate transportation applications such as linear referencing.

City of Winnipeg SCL Map

The City of Winnipeg Street Center Line Map was developed in early years by LBIS based on city survey data and mainframe computer programs. It was later modernized to fit the current GIS environment. Modifications and improvements have been done through the years by city transportation division staff and UMTIG to meet the needs of various GIS-T applications. Currently, this data set is serving as the base map for the City of Winnipeg OnTrac accident database and the traffic information system. An early version of the base map is also used by the Winnipeg Transit for transit system planning and modeling.

The characteristics of the City of Winnipeg SCL map are as follows:

- The original base map was structured with a large number of very fine straight segments, each of which is depicted by two points. This structure was designed to uphold the simplicity of the original mainframe program. However, this overly segmented spatial data structure creates great difficulty and unnecessary burden for many GIS-T functions.
- To make the base map more GIS-T "friendly", UMTIG took the lead in modifying the structure, joining non-intersection end-nodes and making them shape nodes of the new linear features. The greatly simplified base map has received updating and further enhancements from UMTIG.
- Attribute information includes street name, functional class, one-way restrictions, and network linking information, such as crossing street at the start and end points of the street section, and rough direction of the street.
- Information such as traffic estimation, accidents, and transportation zoning system are attached to the base map.
- The base map also forms a fully homogeneous link-node type network with links being street sections and nodes being intersections.
- The inherent direction of the linear features is not uniform in the base map. This creates a topology problem when directional traffic count information is to be related to street sections.
- The base map covers only the streets under City of Winnipeg jurisdiction. Provincial highways and rural municipal roads connecting with city streets are not reflected in the data set. For example, a large portion of the perimeter highway (city by-pass highway) is not included; therefore, transportation analysis regarding the city and vicinity area can not be conducted using this map.

UMTIG Rural Municipal Road Map

This base map was developed during a pilot project to build a GIS-T covering the road network for 6 Rural Municipalities (RMs) in southern Manitoba. The road system involved in the project includes provincial highways and municipal roads. Spatial data from different sources, including Highways, NRCan, and UMTIG are integrated. And traffic related information are gathered and attached to the base map.

The characteristics of the UMTIG RM Road Map are as follows:

- The 6 RMs covered by base map are: Cartier, Gray, Portage la Prairie, North Norfolk, South Norfolk, and Victoria.
- The basic skeleton of the base map was developed from the theoretical grid system established by the Survey of Dominion Land in the early 20th century (Department of the Interior, 1918). This skeleton was then enhanced by NRCan fabric and GPS field survey conducted by UMTIG.

UMTIG Western Canada Railway Map

This railway base map is based on the BTS 1:1,000,000 railway map. The information sources include CP, CN, and UMTIG.

The characteristics of the UMTIG Western Canada Railway Map are as follows:

- Railway lines are classified as groups, i.e., main line, secondary line, and branch line.
- The base map is linear referenced by company, by subdivision, and by mile post.
- Information such as siding location and length, major spur into industrial park, grain elevators, capacity, load class, and operating speed are attached to the base map.
- A grain handling model for Western Canada is developed on this base map, and a grain flow map was created.
- However, due to restricted resources, this base map only enhances the Western Canada portion of the whole network, from the Manitoba/Ontario border to the Pacific coast.
- Accuracy of the original spatial data source is only suitable for large scale planning, and is not good enough to be used in collaboration with detailed spatial data sets such as city street network. This limits the ability of the base map being used in analysis concerning local areas, such as analysis of accidents at at-grade railway crossings, where railway line and local street network should be of similar scale.

In general, data sharing between these GIS-T base maps is in great demand, and current GIS-T base maps lack the flexibility to conduct data sharing efficiently. Transportation applications often require special features and functions which cannot be met by one single base map. It is recommended to create an interoperable GIS platform for transportation applications. The platform will use the best available spatial data, integrate current GIS-T base maps, and have multiple base maps to fulfill various tasks, such as network analysis, traffic monitoring, and asset management.

APPENDIX D STANDARDIZING TOPOLOGICAL DIRECTIONS IN A COMPLEX NETWORK

The Problem

Direction is an inherent characteristic of digital spatial data. For line features, the inherent direction is expressed in the sequence of nodes (vertices) making up the line.

The sequence of the nodes is often determined when the line or polygon features are being digitized in the first place. The directions become the inherent characteristic of the spatial data feature.

In the case of the Winnipeg SCL spatial data, the directions of the SCL features do not follow one uniform convention (as illustrated in Figure D.1).

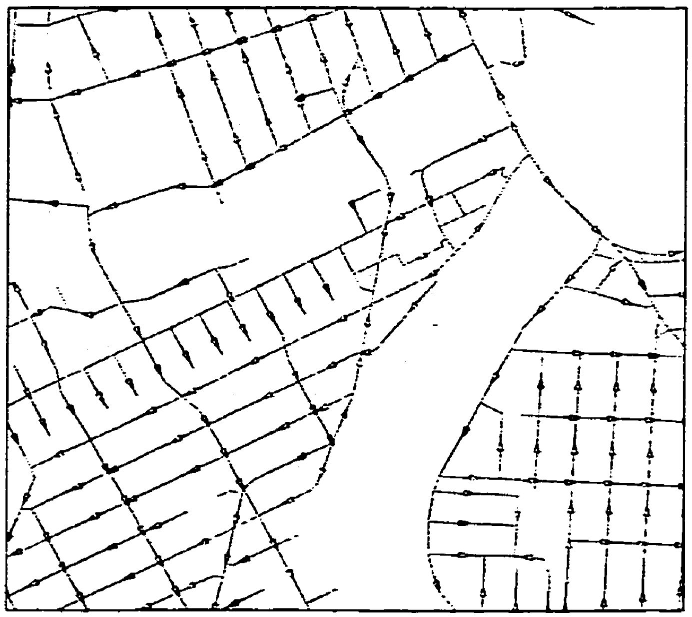


Figure D.1: Original (Haphazard) Directions of the SCL Features

Based on the above observation, it can be concluded that although certain conventions--e.g., going out perpendicularly from rivers--were applied in the digitizing process, exceptions are numerous and no uniform convention was followed through the city.

Traffic counts conducted at intersections are recorded in such a way that numbers of vehicles are organized by directions and movements. A typical intersection traffic count report is shown in Figure D.2. Without a uniform direction convention, the assignment of traffic count data requires a large amount of human interaction.

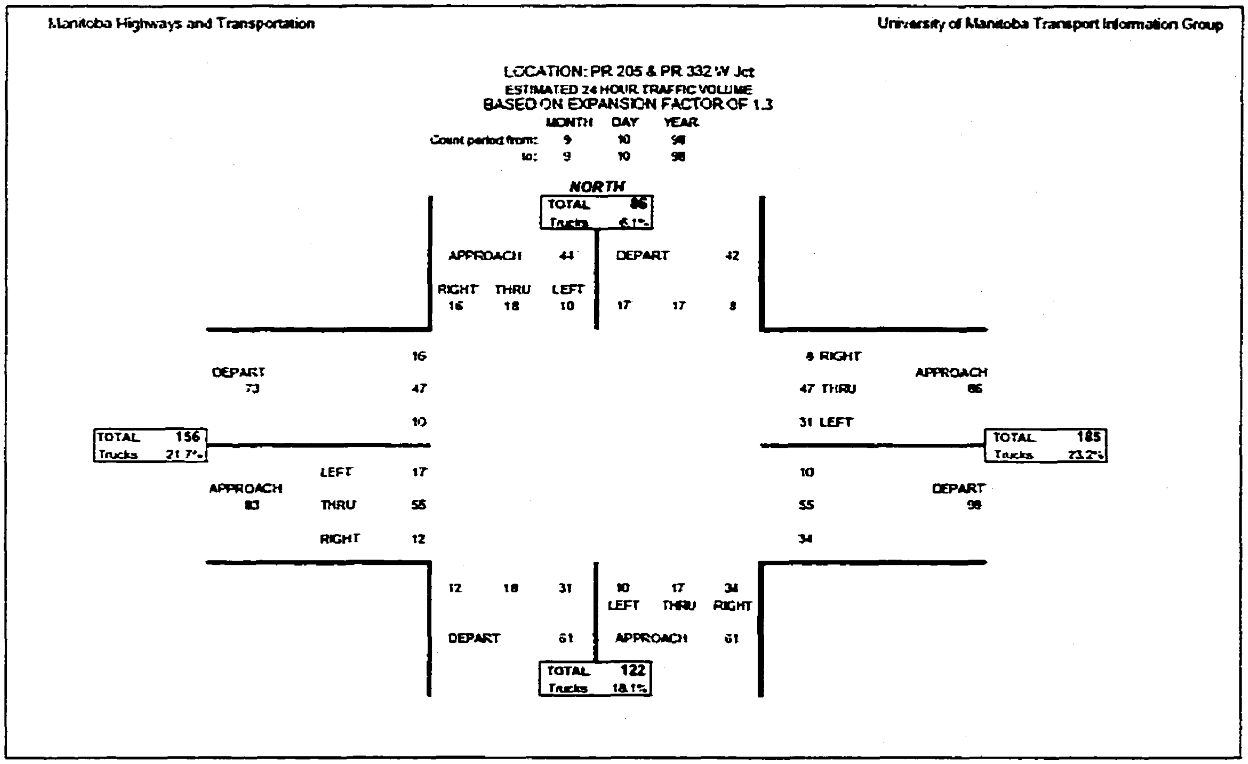


Figure D.2: Example of Intersection Traffic Count Reports

Once the directions of each leg of the intersection are determined, traffic counts can be summarized and assigned to the appropriate street sections. In GIS-T, the differentiation of directional information depends on the original topological direction of the line segments. It is important to know the original direction before proper directional information can be assigned correctly.

Therefore, establishing a uniform direction convention is key to automating traffic count assignments and improving GIS data sharing.

<u>Problem Solving Approach</u>

Since there are no readily available functions for direction modifications in common GIS software, the solution is to develop an in-house program and change the inherent topological directions of the line features. The sequence of the nodes making up the line feature is to be changed where the original direction does not comply with the convention.

The convention of the uniform direction is chosen based on several considerations:

- Compatibility with provincial highways directional convention--The provincial highway department assigns highway control sections following the convention of going from south to north, from west to east. The natural directions of the highway sections basically follow the same rule. In order to eliminate conflicting directions at the joining point at the city boundary, the direction convention used for the city street network follows a similar rule.

- Change of direction within the street section--It is quite common that a street section goes from south-north to east-west within a street segment. To address this problem, dominating directions are chosen over directions of individual sections. Since urban street segments are usually short, start point and end point are picked to represent the rough direction of the line segment.

- One-way streets--Streets having one-way restrictions in the city are treated differently from ordinary two-way streets. A special field in the attribute data is assigned to define the legal direction-of-travel. To simplify the one-way restriction attribution process, natural travelling directions are adopted in the case of interchange ramps.

- Directional consistency within the whole span of a street--Urban streets usually keep their name wherever directions do not change drastically. Within its whole span, the street is segmented by crossing streets. In certain situations, e.g., when a street runs close to 45 degrees northwest-southeast, consecutive sections of the same street may have conflicting directions according to uniform convention. To avoid this kind of problem, an additional conflict-resolving procedure is developed to follow the main procedure.

Programs and Procedures

To keep the integrity of the attribute information contained in the GIS, spatial data and attribute data are exported from GIS separately. The attribute information is stored in database to be linked back to spatial data later on. The spatial data of the street line features are exported into text format to be read and manipulated by computer programs.

Several computer programs and procedures are developed to fulfill the task of changing the line feature directions:

a) The first program reads the spatial data file in text format and puts the start point and end point of each line section into a database. It is followed by a database query process involving calculation of the approximate directions of each section and selection of the lines that need to be changed.

b) Based on the selection, the spatial data set is divided into two sub-sets, with the line features to be changed in one set, and the rest in another set. On the to-be-changed sub-set, a second program is applied to reverse the sequence of the shape nodes making up the each line feature.

c) The process then goes into the conflict-resolving phase, in which user intervention may be required. In the case of interchange ramps, natural directions are determined manually and line directions are changed accordingly.

d) In order to resolve confecting directions within the whole span of a street, a few more steps are required. Attribute information is reattached to all line features, and a database query is conducted based on street names and connectivity of each node (start point or end point). Under normal circumstance, if the directions of all the sections making up the whole street are consistent within the whole span of this same street, each node will only be a start-point once, and an end-point once. When the query result shows a node as a start-point or end-point more than once, that node is the conflicting point.

e) Based on the selection of the conflicting point, inconsistent segments can be selected by visual checking. The final step of the conflict-resolving procedure is to reverse the directions of these inconsistent sections using the previously mentioned second program.

f) After the conflict-resolving procedure, with all the line features possess correct directions, the attribute data set is attached to the spatial data set again to make an integrated GIS data set. This data set is the final result of the whole undertaking, which is shown in Figure D.3.

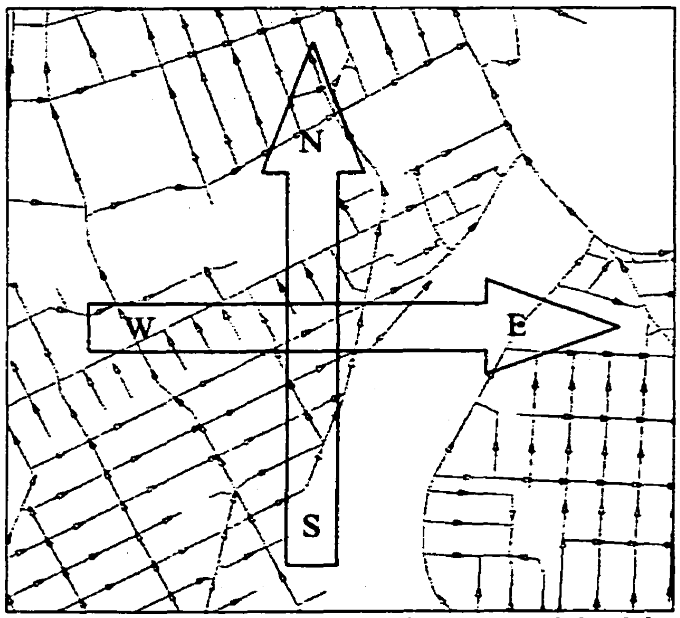


Figure D.3: Modified (Standardized) Directions of the SCL Features

Throughout the changing and conflict-resolving procedure, the unique ID's assigned to spatial data play the key role as the linking agent. These unique ID's are kept attached to both spatial and attribute data sets. This ensures the ability to restore the integrated GIS data set and the integrity of the final result.

APPENDIX E DETAILS OF IN-HOUSE PROGRAMS

The programs listed here are owned and copyrighted by the author and Professor Alan Clayton. They are not to be copied, modified, or re-compiled without the consent of the aforementioned. Certain portions of the program list are blanked out.

Endpoint.cpp

```
// Endpoint.cpp : Defines the entry point for the console application.
// 1st generation of matching program to find matching lines from two
// spatial data sets. Feburary 2000.

#include "stdafx.h"
#include <strstrea.h>
#include <fstream.h>
#include <string.h>
#include <stdlib.h>
#include <math.h>

#define MAX_LINES 20000

#define BUF_DIST 10

[redacted]

int main(int argc, char* argv[])
{
 char ch;
 int i, j, k, l, node_num, line_num1, line_num2, match_num;
 int match[MAX_LINES][2];
 long id[MAX_LINES], uid[MAX_LINES];
 float start_northing[MAX_LINES], start_easting[MAX_LINES];
 float end_northing[MAX_LINES], end_easting[MAX_LINES];
 float start_x[MAX_LINES], start_y[MAX_LINES];
 float end_x[MAX_LINES], end_y[MAX_LINES];
[redacted]

 // initialization
 i = 0;
 j = 0;
 k = 0;
 l = 0;
 node_num = 0;
 line_num1 = 0;
 line_num2 = 0;
 match_num = 0;
[redacted]
  id[i] = 0;
  uid[i] = 0;
  match[i][0] = 0;
  match[i][1] = 0;
  start_northing[i] = 0;
  start_easting[i] = 0;
  end_northing[i] = 0;
  end_easting[i] = 0;
  start_x[i] = 0;
  start_y[i] = 0;
  end_x[i] = 0;
  end_y[i] = 0;
 }
```

```
 if(argc != 3) {
  [redacted]
  return -1;
 }

 ifstream infl(argv[1]);
 if(!infl) {
  cout << "Cannot open file1.\n";
  return 1;
 }

 ofstream out("result.txt");
 if(!out) {
  cout << "Cannot open result file.\n";
  return 1;
 }

// Retrieve info from "infile1"
 j = 0;
 [redacted]
  // define string buffer
  istrstream ins1(inbuf, strlen(inbuf));
  // read line ID
  ins1.get(numbuf,101,',');
  id[j] = atol(numbuf);

  // read NUMBER of shape-nodes in the line
  ins1.get(ch); // skip the comma
  ins1.get(numbuf,101,',');
  [redacted]

  // read easting & northing of the START point
  ins1.get(ch); // skip the comma
  ins1.get(numbuf,101,',');
  start_easting[j] = atof(numbuf);

  ins1.get(ch); // skip the comma
  ins1.get(numbuf,101,',');
  start_northing[j] = atof(numbuf);

  [redacted]
   ins1.get(ch); // skip the comma
   ins1.get(numbuf,101,',');

   ins1.get(ch); // skip the comma
   ins1.get(numbuf,101,',');
  }

  // read easting & northing of the END point
  ins1.get(ch); // skip the comma
  ins1.get(numbuf,101,',');
  [redacted]

  ins1.get(ch); // skip the comma
  ins1.get(numbuf,101,',');
  end_northing[j] = atof(numbuf);

  j++;
 }

 [redacted]
 infl.close();


// Retrieve info from "infile2"
 [redacted]
 if(!inf2) {
  cout << "Cannot open file2.\n";
  return 1;
 }
```

```
j = 0;
[redacted]
 // define string buffer
 istrstream ins2(inbuf, strlen(inbuf));
 // read line ID
 ins2.get(numbuf,101,',');
 uid[j] = atol(numbuf);

 // read NUMBER of shape-nodes in the line
 ins2.get(ch); // skip the comma
 ins2.get(numbuf,101,',');
 [redacted]

 // read easting & northing of the START point
 ins2.get(ch); // skip the comma
 ins2.get(numbuf,101,',');
 start_x[j] = atof(numbuf);

 ins2.get(ch); // skip the comma
 ins2.get(numbuf,101,',');
 start_y[j] = atof(numbuf);

 [redacted]
   ins2.get(ch); // skip the comma
   ins2.get(numbuf,101,',');

   ins2.get(ch); // skip the comma
   ins2.get(numbuf,101,',');
 }

 // read easting & northing of the END point
 ins2.get(ch); // skip the comma
 ins2.get(numbuf,101,',');
 [redacted]

 ins2.get(ch); // skip the comma
 ins2.get(numbuf,101,',');
 end_y[j] = atof(numbuf);

 j++;
}

[redacted]
inf2.close();

// Find out lines have common start and end points

l = 0;

for (i=0; i<line_num1; i++)
  for (j=0; j<line_num2; j++)
    if (DIST(start_easting[i],start_northing[i],start_x[j],start_y[j]) <= BUF_DIST)
      // If start points match
      [redacted]
        // Then check if end points match
        match[l][0] = i;
        match[l][1] = j;
        l++;
      }
    else if (DIST(start_easting[i],start_northing[i],end_x[j],end_y[j]) <= BUF_DIST)
      // If start & end points match
      [redacted]
        // Then check if end points match
        match[l][0] = i;
        match[l][1] = j;
        l++;
      }

[redacted]
```

```
 for (i=0;i<match_num;i++) {
   out << id[match[i][0]] <<"," << uid[match[i][1]] << "," << start_easting[match[i][0]] << "," << start_northing[match[i][0]] << ","
<< end_x[match[i][0]] << "," << end_y[match[i][0]] << "," << start_easting[match[i][1]] << "," << start_northing[match[i][1]] << ","
[redacted]
 }
 out.close();

 return 0;
}
```

match.cpp

```
// match.cpp : Defines the entry point for the console application.
// 2nd generation of matching program to find matching lines from two
// spatial data sets. This program will find several sub segments
// which are being represented as one line in another data set
// May 2000.

#include "stdafx.h"
#include <strstrea.h>
#include <fstream.h>
#include <string.h>
#include <stdlib.h>
#include <math.h>

#define MAX_LINES 10000


[redacted]

int main(int argc, char* argv[])
{
 char ch;
 int i, j, k, l, node_num, line_num1, line_num2, match_num;
 int match[MAX_LINES][3];
 long id[MAX_LINES], uid[MAX_LINES];
 float start_northing[MAX_LINES], start_easting[MAX_LINES];
 float end_northing[MAX_LINES], end_easting[MAX_LINES];
 float start_x[MAX_LINES], start_y[MAX_LINES];
 float end_x[MAX_LINES], end_y[MAX_LINES];
 // Super long buffer to accommodate complicated lines
 [redacted]
 float buf_dist=10; //default value of the buffer distance


 // initialization
 i = 0;
 j = 0;
 k = 0;
 l = 0;
 node_num = 0;
 line_num1 = 0;
 line_num2 = 0;
 match_num = 0;
 [redacted]
  id[i] = 0;
  uid[i] = 0;
  match[i][0] = 0;
  match[i][1] = 0;
  match[i][2] = 0;
  start_northing[i] = 0;
  start_easting[i] = 0;
  end_northing[i] = 0;
  end_easting[i] = 0;
  start_x[i] = 0;
  start_y[i] = 0;
  end_x[i] = 0;
  end_y[i] = 0;
 }

 if (argc != 4) {
  cout<<"Usage: match <inputfile1> <inputfile2> <buf_dist>\n";
  return -1;
 }

 buf_dist = atof(argv[3]);

 ifstream inf1(argv[1]);
 if(!inf1) {
```

```
    cout << "Cannot open file1.\n";
    return 1;
  }

  ofstream out("result.txt");
  if(!out) {
    cout << "Cannot open result file.\n";
    return 1;
  }

// Retrieve info from "infile1"
  j = 0;
  [redacted]
    // define string buffer
    istrstream ins1(inbuf, strlen(inbuf));
    // read line ID
    ins1.get(numbuf,101,',');
    id[j] = atol(numbuf);

    // read NUMBER of shape-nodes in the line
    ins1.get(ch); // skip the comma
    ins1.get(numbuf,101,',');
    [redacted]

    // read easting & northing of the START point
    ins1.get(ch); // skip the comma
    ins1.get(numbuf,101,',');
    start_easting[j] = atof(numbuf);

    ins1.get(ch); // skip the comma
    ins1.get(numbuf,101,',');
    start_northing[j] = atof(numbuf);

    [redacted]
      ins1.get(ch); // skip the comma
      ins1.get(numbuf,101,',');

      ins1.get(ch); // skip the comma
      ins1.get(numbuf,101,',');

//***** memerize these points

    }

    // read easting & northing of the END point
    [redacted]
    ins1.get(numbuf,101,',');
    end_easting[j] = atof(numbuf);

    [redacted]
    ins1.get(numbuf,101,',');
    end_northing[j] = atof(numbuf);

    j++;
  }

  [redacted]
  infl.close();


// Retrieve info from "infile2"
  ifstream inf2(argv[2]);
  if(!inf2) {
    cout << "Cannot open file2.\n";
    return 1;
  }

  j = 0;
  [redacted]
    // define string buffer
```

```
istrstream ins2(inbuf, strlen(inbuf));
// read line ID
ins2.get(numbuf,101,',');
[redacted]

// read NUMBER of shape-nodes in the line
ins2.get(ch); // skip the comma
ins2.get(numbuf,101,',');
node_num = atoi(numbuf);

// read easting & northing of the START point
ins2.get(ch); // skip the comma
ins2.get(numbuf,101,',');
start_x[j] = atof(numbuf);

ins2.get(ch); // skip the comma
ins2.get(numbuf,101,',');
start_y[j] = atof(numbuf);

[redacted]
 ins2.get(ch); // skip the comma
 ins2.get(numbuf,101,',');

 ins2.get(ch); // skip the comma
 ins2.get(numbuf,101,',');

//***** memerize these points

}

// read easting & northing of the END point
[redacted]
ins2.get(numbuf,101,',');
end_x[j] = atof(numbuf);

ins2.get(ch); // skip the comma
ins2.get(numbuf,101,',');
end_y[j] = atof(numbuf);

j++;
}

[redacted]
inf2.close();

// Find out lines have common start and end points

//***** find ranges of the two lines, if they are not close enough
// then skip them. If they are close, find which one is longer. And
// then check if points from the shorter line are within buffer
// distance from the longer line. In that way, determine matching
// lines.


l = 0;

for (i=0; i<line_num1; i++)
 for (j=0; j<line_num2; j++)
  if (DIST(start_easting[i],start_northing[i],start_x[j],start_y[j]) <= buf_dist)
   // If start points match
   [redacted]
    // Then check if end points match
    match[l][0] = i;
    match[l][1] = j;
    l++;
   }
  else if (DIST(start_easting[i],start_northing[i],end_x[j],end_y[j]) <= buf_dist)
   // If start & end points match
   [redacted]
    // Then check if end points match
```

```
      match[l][0] = i;
      match[l][1] = j;
      match[l][2] = 1;
      l++;
    }

[redacted]
 for (i=0;i<match_num;i++) {
  out << id[match[i][0]] <<"," << uid[match[i][1]] << "," << match[i][2] << "\n";
 }
 out.close();

 return 0;
}
```

Midpoint.cpp

```
// Midpoint.cpp : Defines the entry point for the console application.
// 1st section of the proposed deriving centerline from double-line
// representation in spatial data. It calculates theoritical centerline
// using lines perpendicular to tangent at existing shape-nodes.
// October, 2000

#include "stdafx.h"
#include <strstrea.h>
#include <fstream.h>
#include <string.h>
#include <stdlib.h>
#include <math.h>
#include <iomanip.h>

[redacted]
#define MAX_NODES 1000

#define BUF_DIST 200

[redacted]

int main(int argc, char* argv[])
{
 char ch;
 int i, j, k, l, m, mm;

 int kk, ll; // temporary index in places of k and l
 int choice;
 double tmp0, tmp1, tmp2;
 int index;
 double mini;

 // Super long buffer to accommodate complicated lines
 [redacted]

 // coordinates to help calculating the theoretical midpoint
 double x1, y1, x2, y2, x3, y3, x4, y4;
 double xh, yh, xk, yk, xt, yt;

 // coordinates to help figure out the start point
 double start_x1, start_y1, start_x2, start_y2;
 [redacted]
 double end_x1, end_y1, end_x2, end_y2;

 // length used to calculate midpoints
 double dist_h, dist_k, dist_t, dist_x;

 // length used to figure out which point comes first
 double dist_1, dist_2, dist_3, dist_4;

 // variables to help reading lines from two input files and storing results
 long id1[MAX_LINES];
 int node_num1[MAX_LINES];
 double xx1[MAX_NODES], yy1[MAX_NODES];
 long id2[MAX_LINES];
 int node_num2[MAX_LINES];
 double xx2[MAX_NODES], yy2[MAX_NODES];
 long id3[MAX_LINES];
 int node_num3[MAX_LINES];
 double xx3[MAX_NODES], yy3[MAX_NODES];


 // initialization
 i = 0;
 j = 0;
 k = 0;
 l = 0;
```

```
m = 0;
mm = 0;
kk = 0;
ll = 0;
choice = 0;
tmp0 = 0;
tmp1 = 0;
tmp2 = 0;
[redacted]
mini = 0;
x1 = 0; y1 = 0;
x2 = 0; y2 = 0;
x3 = 0; y3 = 0;
x4 = 0; y4 = 0;
xh = 0; yh = 0;
xk = 0; yk = 0;
xt = 0; yt = 0;
start_x1 = 0; start_y1 = 0;
start_x2 = 0; start_y2 = 0;
mid_x1 = 0; mid_y1 = 0;
mid_x2 = 0; mid_y2 = 0;
end_x1 = 0; end_y1 = 0;
end_x2 = 0; end_y2 = 0;
dist_h = 0; dist_k = 0; dist_t = 0; dist_x = 0;
dist_1 = 0; dist_2 = 0; dist_3 = 0; dist_4 = 0;

for (i=0; i<MAX_LINES; i++) {
  id1[i] = 0; node_num1[i] = 0;
  id2[i] = 0; node_num2[i] = 0;
  id3[i] = 0; node_num3[i] = 0;
}
[redacted]
  xx1[i] = 0; yy1[i] = 0;
  xx2[i] = 0; yy2[i] = 0;
  xx3[i] = 0; yy3[i] = 0;
}

// Start openning the files
if (argc != 4) {
  cout<<"Usage: Midpoint <inputfile1> <inputfile2> <outputfile>\n";
  return -1;
}

ifstream inf1(argv[1]);
if(!inf1) {
  cout << "Cannot open INPUT file1.\n";
  return 1;
}

ifstream inf2(argv[2]);
if(!inf2) {
  cout << "Cannot open INPUT file2.\n";
  return 1;
}

ofstream out(argv[3]);
if(!out) {
  cout << "Cannot open OUTPUT file.\n";
  return 1;
}

// Retrieve info from "infile1" and "infile2", and calculate the mid-point
j = 0;

// read a line from "infile1"
[redacted]
  // define string buffer
  istrstream ins1(inbuf, strlen(inbuf));
  // read line ID
  ins1.get(numbuf,101,',');
```

```
id1[j] = atol(numbuf);

// read NUMBER of shape-nodes in the line
ins1.get(ch); // skip the comma
ins1.get(numbuf,101,',');
node_num1[j] = atoi(numbuf);

// read easting & northing readings of the shape-nodes
[redacted]
  ins1.get(ch); // skip the comma
  ins1.get(numbuf,101,',');
  xx1[i] = atof(numbuf);

  ins1.get(ch); // skip the comma
  ins1.get(numbuf,101,',');
  yy1[i] = atof(numbuf);
}
start_x1 = xx1[0];
start_y1 = yy1[0];
end_x1 = xx1[node_num1[j]-1];
end_y1 = yy1[node_num1[j]-1];
mid_x1 = (start_x1 + end_x1)/2;
mid_y1 = (start_y1 + end_y1)/2;


// read a line from "infile2"
[redacted]

// define string buffer
istrstream ins2(inbuf, strlen(inbuf));
// read line ID
ins2.get(numbuf,101,',');
id2[j] = atol(numbuf);

// read NUMBER of shape-nodes in the line
ins2.get(ch); // skip the comma
ins2.get(numbuf,101,',');
node_num2[j] = atoi(numbuf);

// read easting & northing readings of the shape-nodes
[redacted]
  ins2.get(ch); // skip the comma
  ins2.get(numbuf,101,',');
  xx2[i] = atof(numbuf);

  ins2.get(ch); // skip the comma
  ins2.get(numbuf,101,',');
  yy2[i] = atof(numbuf);
}
start_x2 = xx2[0];
start_y2 = yy2[0];
end_x2 = xx2[node_num1[j]-1];
end_y2 = yy2[node_num1[j]-1];
mid_x2 = (start_x2 + end_x2)/2;
mid_y2 = (start_y2 + end_y2)/2;


// calculate the theoretical center point from
// the point in concern with the following huge formulae:

// For start point from line TWO:

// Assign base points
x1 = xx1[0];   y1 = yy1[0];
x2 = xx1[1];   y2 = yy1[1];
x3 = xx2[0];   y3 = yy2[0];
x4 = xx2[1];   y4 = yy2[1];

[redacted]
```

```
[redacted]
dist_h = DIST(x3,y3,xh,yh);

[redacted]
dist_k = DIST(x3,y3,xk,yk);

dist_t = dist_h/(1 + dist_h/dist_k);

dist_x = DIST(x3,y3,x4,y4);
xt = x3 - dist_t*(y4-y3)/dist_x;
yt = y3 + dist_t*(x4-x3)/dist_x;

if (DIST(xt,yt,xk,yk)> dist_k ) {
  xt = x3 + dist_t*(y4-y3)/dist_x;
  yt = y3 - dist_t*(x4-x3)/dist_x;
}

// Help figure out which point is closer to start point
dist_1 = DIST(xt,yt,mid_x1,mid_y1);
dist_2 = DIST(xt,yt,mid_x2,mid_y2);
if (dist_1 > dist_2) dist_1 = dist_2;

// Register the new mid-point
id3[j] = j+1;
node_num3[j] = 1;
xx3[0] = xt;
yy3[0] = yt;

// For start point from line ONE:

// Re-initialize
x1 = 0; y1 = 0;
x2 = 0; y2 = 0;
x3 = 0; y3 = 0;
x4 = 0; y4 = 0;
xh = 0; yh = 0;
xk = 0; yk = 0;
xt = 0; yt = 0;
dist_h = 0; dist_k = 0; dist_t = 0; dist_x = 0;

// Assign base points
x1 = xx2[0];   y1 = yy2[0];
x2 = xx2[1];   y2 = yy2[1];
x3 = xx1[0];   y3 = yy1[0];
x4 = xx1[1];   y4 = yy1[1];

[redacted]
dist_h = DIST(x3,y3,xh,yh);

[redacted]
dist_k = DIST(x3,y3,xk,yk);

dist_t = dist_h/(1 + dist_h/dist_k);

dist_x = DIST(x3,y3,x4,y4);
xt = x3 - dist_t*(y4-y3)/dist_x;
yt = y3 + dist_t*(x4-x3)/dist_x;

if (DIST(xt,yt,xk,yk)> dist_k ) {
  xt = x3 + dist_t*(y4-y3)/dist_x;
  yt = y3 - dist_t*(x4-x3)/dist_x;
}

// Help figure out which point is closer to start point
```

```
dist_4 = DIST(xt,yt,mid_x1,mid_y1);
dist_3 = DIST(xt,yt,mid_x2,mid_y2);
if (dist_3 > dist_4) dist_3 = dist_4;

// Register the new mid-point, if necessary
if (dist_1 > dist_3) choice = 2;
else {
 choice = 1;
 xx3[0] = xt;
 yy3[0] = yt;
}

i = 0; // control index for center line
k = 0; // control index for line ONE
l = 0; // control index for line TWO

if (choice == 1) k++;
else l++;


[redacted]
 // figure out next point;
 // choice changed accordingly;
 if (k == node_num1[j]) choice = 2;
 else if (l == node_num2[j]) choice = 1;
 else if (DIST(xx3[i], yy3[i], xx1[k], yy1[k]) < DIST(xx3[i], yy3[i], xx2 [l], yy2[l])) choice = 1;
 else choice = 2;

 // determine the matching segment
 if (choice == 1) { // find the nearest node point on the opposite line
  mini = DIST(xx1[k], yy1[k], xx2[0], yy2[0]);
  index = 0;
  for (mm = 1; mm < node_num2[j]; mm++) {
   tmp0 = DIST(xx1[k], yy1[k], xx2[mm], yy2[mm]);
   [redacted]
    mini = tmp0;
    index = mm;
   }
  }
  if (index > 0 && index < node_num2[j]-1) { // pick up the right line segment
   x1 = xx2[index];   y1 = yy2[index];
   x2 = xx2[index+1];   y2 = yy2[index+1];
   x3 = xx1[k];   y3 = yy1[k];

[redacted]

   if ((xh >= xx2[index] && xh <= xx2[index+1]) || (xh >= xx2[index+1] && xh <= xx2[index])) {
    ll = index;
    kk= k;
    x1 = xx2[ll]; y1 = yy2[ll]; x2 = xx2[ll+1]; y2 = yy2[ll+1];
    x3 = xx1[kk]; y3 = yy1[kk]; x4 = xx1[kk+1]; y4 = yy1[kk+1];
   }
   else {
    [redacted]
    kk= k;
    x1 = xx2[ll+1]; y1 = yy2[ll+1]; x2 = xx2[ll]; y2 = yy2[ll];
    x3 = xx1[kk]; y3 = yy1[kk]; x4 = xx1[kk-1]; y4 = yy1[kk-1];
   }
  }
  else if (index == 0) {
   ll = 0;
   kk= k;
   x1 = xx2[ll]; y1 = yy2[ll]; x2 = xx2[ll+1]; y2 = yy2[ll+1];
   x3 = xx1[kk]; y3 = yy1[kk]; x4 = xx1[kk+1]; y4 = yy1[kk+1];
  }
  else if (index == node_num2[j]-1) {
   [redacted]
   kk= k;
```

```
      x1 = xx2[ll+1]; y1 = yy2[ll+1]; x2 = xx2[ll]; y2 = yy2[ll];
      x3 = xx1[kk]; y3 = yy1[kk]; x4 = xx1[kk-1]; y4 = yy1[kk-1];
    }
  }
  else { // choice == 2
   [redacted]
   index = 0;
   for (mm = 1; mm < node_num1[j]; mm++) {
     tmp0 = DIST(xx2[l], yy2[l], xx1[mm], yy1[mm]);
     if (tmp0 < mini) {
       mini = tmp0;
       index = mm;
     }
   }
   if (index > 0 && index < node_num1[j]-1) {
     x1 = xx1[index];    y1 = yy1[index];
     x2 = xx1[index+1];   y2 = yy1[index+1];
     x3 = xx2[l];    y3 = yy2[l];

[redacted]

     if ((xh >= xx1[index] && xh <= xx1[index+1]) || (xh >= xx1[index+1] && xh <= xx1[index])) {
       kk= index;
       ll= l;
       x1 = xx1[kk]; y1 = yy1[kk]; x2 = xx1[kk+1]; y2 = yy1[kk+1];
       x3 = xx2[ll]; y3 = yy2[ll]; x4 = xx2[ll+1]; y4 = yy2[ll+1];
     }
     else {
       kk = index-1;
       ll= l;
       x1 = xx1[kk+1]; y1 = yy1[kk+1]; x2 = xx1[kk]; y2 = yy1[kk];
       x3 = xx2[ll]; y3 = yy2[ll]; x4 = xx2[ll-1]; y4 = yy2[ll-1];
    }
   }
   [redacted]
    kk = 0;
    ll= l;
    x1 = xx1[kk]; y1 = yy1[kk]; x2 = xx1[kk+1]; y2 = yy1[kk+1];
    x3 = xx2[ll]; y3 = yy2[ll]; x4 = xx2[ll+1]; y4 = yy2[ll+1];
   }
   else if (index == node_num1[j]-1) {
    kk = node_num1[j]-2;
    ll= l;
    x1 = xx1[kk+1]; y1 = yy1[kk+1]; x2 = xx1[kk]; y2 = yy1[kk];
    x3 = xx2[ll]; y3 = yy2[ll]; x4 = xx2[ll-1]; y4 = yy2[ll-1];
   }
  }

 // caculate midpoint;
[redacted]
 dist_h = DIST(x3,y3,xh,yh);

[redacted]
 dist_k = DIST(x3,y3,xk,yk);

 dist_t = dist_h/(1 + dist_h/dist_k);

 dist_x = DIST(x3,y3,x4,y4);
 xt = x3 - dist_t*(y4-y3)/dist_x;
 yt = y3 + dist_t*(x4-x3)/dist_x;

 if (DIST(xt,yt,xk,yk)> dist_k ) {
   xt = x3 + dist_t*(y4-y3)/dist_x;
   yt = y3 - dist_t*(x4-x3)/dist_x;
```

```
    }

    // Register the new mid-point
[redacted]
     tmp1 = xx3[i];
     tmp2 = yy3[i];
     xx3[i] = xt;
     yy3[i] = yt;
     xt = tmp1;
     yt = tmp2;
    }

    if (DIST(xt,yt,xx3[i],yy3[i]) > 1) { // if the two points are not too close, i.e., 1 meter
     i++;
     node_num3[j] = i+1;
     xx3[i] = xt;
     yy3[i] = yt;
    }
    else {
     xx3[i] = (xx3[i] + xt) / 2;
     yy3[i] = (yy3[i] + yt) / 2;
    }

    if (choice == 1) {
     k++;
    }
    else { // choice == 2
     l++;
    }
   }


  // Output the result
  out.setf(ios::fixed, ios::floatfield);
  out.setf(ios::showpoint);
/*
  out << id1[j] << "," << node_num1[j];
  for (m=0; m<node_num1[j]; m++) out << "," << setprecision(6) << xx1[m] << "," << setprecision(6) << yy1[m];
  out << "\n";

  out << id2[j] << "," << node_num2[j];
  for (m=0; m<node_num2[j]; m++) out << "," << setprecision(6) << xx2[m] << "," << setprecision(6) << yy2[m];
  out << "\n";
*/
  out << id3[j] << "," << node_num3[j];
  for (m=0; m<node_num3[j]; m++) out << "," << setprecision(6) << xx3[m] << "," << setprecision(6) << yy3[m];
  out << "\n";

 [redacted]
 }
 inf1.close();
 inf2.close();

 out.close();

 return 0;

}
```

reverse.cpp

```
// Reverse.cpp : Based on Endpoint.cpp, reverse direction of the line
// segments. July 12, 2000.

// Endpoint.cpp : Defines the entry point for the console application.
// 1st generation of matching program to find matching lines from two
// spatial data sets. Feburary 2000.

#include <strstrea.h>
#include <fstream.h>
#include <string.h>
#include <stdlib.h>
#include <math.h>

#define MAX_NODES 2000

int main(int argc, char* argv[])
{
 char ch;
 int i, node_num;
 long id;
 char node_northing[MAX_NODES][20], node_easting[MAX_NODES][20];
 // Super long buffer to accommodate complicated lines
 [redacted]


 // initialization
 i = 0;
 node_num = 0;
 id = 0;

 if (argc != 3) {
  cout<<"Usage: Reverse <inputfile> <outputfile>\n";
  return -1;
 }

 ifstream inf(argv[1]);
 if(!inf) {
  cout << "Cannot open input file.\n";
  return 1;
 }

 ofstream out(argv[2]);
 if(!out) {
  cout << "Cannot open output file.\n";
  return 1;
 }

// Retrieve info from "infile"
 [redacted]
  // define string buffer
  istrstream ins(inbuf, strlen(inbuf));
  // read line ID
  ins.get(numbuf,101,',');
  id = atol(numbuf);

  // read NUMBER of shape-nodes in the line
  ins.get(ch); // skip the comma
  ins.get(numbuf,101,',');
  node_num = atoi(numbuf);

  [redacted]
   ins.get(ch); // skip the comma
   ins.get(numbuf,101,',');
   strcpy (node_easting[i],numbuf);

   ins.get(ch); // skip the comma
   ins.get(numbuf,101,',');
```

```
    strcpy (node_northing[i],numbuf);
   }


  out << id <<"," << node_num;
  for (i=node_num-1; i>=0; i--) {
   out << "," << node_easting[i] << "," << node_northing[i];
  }
  out << "\n";
 }

 inf.close();
 out.close();

 return 0;
}
```

st_end.cpp

```
// st_end.cpp : Defines the entry point for the console application.
// Based on Reverse.cpp. Gives out start and end points' coordinates
// with id's for further studys.

// Reverse.cpp : Based on Endpoint.cpp, reverse direction of the line
// segments. July 12, 2000.

// Endpoint.cpp : Defines the entry point for the console application.
// 1st generation of matching program to find matching lines from two
// spatial data sets. Feburary 2000.

#include <strstrea.h>
#include <fstream.h>
#include <string.h>
#include <stdlib.h>
#include <math.h>

#define MAX_NODES 2000

int main(int argc, char* argv[])
{
 char ch;
 int i, node_num;
 long id;
 char node_northing[MAX_NODES][20], node_easting[MAX_NODES][20];
 // Super long buffer to accommodate complicated lines
 [redacted]

 // initialization
 i = 0;
 node_num = 0;
 id = 0;

 if (argc != 3) {
  cout<<"Usage: st_end <inputfile> <outputfile>\n";
  return -1;
 }

 ifstream inf(argv[1]);
 if(!inf) {
  cout << "Cannot open input file.\n";
  return 1;
 }

 ofstream out(argv[2]);
 if(!out) {
  cout << "Cannot open output file.\n";
  return 1;
 }

// Retrieve info from "infile"
 [redacted]
  // define string buffer
  istrstream ins(inbuf, strlen(inbuf));
  // read line ID
  ins.get(numbuf,101,',');
  id = atol(numbuf);

  // read NUMBER of shape-nodes in the line
  ins.get(ch); // skip the comma
  ins.get(numbuf,101,',');
  node_num = atoi(numbuf);

  for (i=0; i<node_num; i++) {
   ins.get(ch); // skip the comma
   ins.get(numbuf,101,',');
```

```
    strcpy (node_easting[i],numbuf);

    ins.get(ch); // skip the comma
    ins.get(numbuf,101,',');
    strcpy (node_northing[i],numbuf);
   }

  out << id <<"," << node_num;
  out << "," << node_easting[0] << "," << node_northing[0];
  out << "," << node_easting[node_num-1] << "," << node_northing[node_num-1];
  out << "\n";
 }

 inf.close();
 out.close();

 return 0;
}
```