7Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
SWUTC/03/167426-1		

4. Title and Subtitle GIS-BASED ASSESSMENT OF HIGHWAY NETWORK, CRASH, AND TRAFFIC VOLUME DATA IN THE LAREDO-NUEVO LAREDO AREA	5. Report Date October 2003 6. Performing Organization Code
7. Author(s) Cesar Quiroga, Robert Pina, and Michael Perez	8. Performing Organization Report No. Research Report 167426-1
9. Performing Organization Name and Address Texas Transportation Institute The Texas A&M University System College Station, Texas 77843-3135	10. Work Unit No. (TRAIS) 11. Contract or Grant No. 10727
12. Sponsoring Agency Name and Address Southwest Region University Transportation Center Texas Transportation Institute The Texas A&M University System College Station, Texas 77843-3135	13. Type of Report and Period Covered 14. Sponsoring Agency Code

15. Supplementary Notes

Supported by general revenues from the State of Texas.

16. Abstract

The Texas-Mexico Border Region is undergoing extremely fast, radical changes, and it is critical for planners, engineers, and decision makers to develop a clear understanding of the various processes that drive, and are affected by, such growth. A number of government agencies are responsible for collecting and processing transportation-related data in the border region. Unfortunately, the available data tend to be scattered among various organizations and rarely include substantial amounts of information from the Mexican side of the border.

There is a need to provide an integrated approach to the issue of transportation data collection and analysis for the Texas-Mexico Border Region. This report describes a prototype geographic information system (GIS)-based framework for transportation data with a goal to better understand the characteristics of the transportation system along border areas. The report focuses on highway network data, crash data, and traffic volume data in the Laredo – Nuevo Laredo area. The report examines existing transportation data sources at various jurisdictional levels and evaluates the degree to which the data can be integrated. The report also summarizes an assessment of transportation data on the Mexican side of the border that resulted from a collaborative effort with the Instituto Mexicano del Transporte (IMT).

17. Key Words	18. Distribution Statement
Texas-Mexico Border, GIS, Transportation Data, Highway Network, Crash Data, Traffic Volumes, Databases	No restrictions. This document is available to the public through NTIS: National Technical Information Service 5285 Port Royal Road Springfield, Virginia 22161

19. Security Classif.(of this report)	20. Security Classif.(of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	74	

Form DOT F 1700.7 (8-72) **Reproduction of completed page authorized**

GIS-BASED ASSESSMENT OF HIGHWAY NETWORK, CRASH, AND TRAFFIC VOLUME DATA IN THE LAREDO-NUEVO LAREDO AREA

by

Cesar Quiroga, Ph.D., P.E.
Associate Research Engineer
Texas Transportation Institute

Robert Pina
Programmer/Analyst I
Texas Transportation Institute

And

Michael Perez
Student Programmer
Texas Transportation Institute

Report Number SWUTC/03/167426-1

Sponsored by the
Southwest Region University Transportation Center
Texas Transportation Institute
Texas A&M University System
College Station, TX 77843-3135

October 2003

DISCLAIMER

The contents of this document reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official view or policies of the Texas Transportation Institute (TTI), the Texas A&M University System, the University Transportation Centers Program, or the U.S. Department of Transportation. This document does not constitute a standard, specification, or regulation, nor is it intended for construction, bidding, or permit purposes.

NOTICE

The United States Government and the State of Texas do not endorse products or manufacturers. Trade or manufacturers' names appear herein solely because they are considered essential to the object of this report.

ACKNOWLEDGMENTS

The authors recognize that support for this research was provided by a grant from the U.S. Department of Transportation, University Transportation Centers Program to the Southwest Region University Transportation Center, which is funded 50% with general revenue funds from the State of Texas. The researchers would like to gratefully acknowledge the availability of data and assistance provided by the Texas Department of Transportation (TxDOT), the Texas Department of Public Safety (TxDPS), the Laredo Police Department, and the Texas Natural Resource Information System (TNRIS). The researchers are particularly grateful to Jo Ann Garcia at the TxDOT Laredo District for her feedback and for allowing the use of specific traffic volume and crash data figures and analysis in the report. The researchers are also grateful for Jorge Acha's effort at the Instituto Mexicano del Transporte (IMT) in setting up a similar project to assess transportation data sources on the Mexican side of the border. The researchers also acknowledge the assistance provided by Lisa Dye, project monitor with the Federal Highway Administration.

ABSTRACT

The Texas-Mexico Border Region is undergoing extremely fast, radical changes, and it is critical for planners, engineers, and decision makers to develop a clear understanding of the various processes that drive, and are affected by, such growth. A number of government agencies are responsible for collecting and processing transportation-related data in the border region. Unfortunately, the available data tend to be scattered among various organizations and rarely include substantial amounts of information from the Mexican side of the border. There is a need to provide an integrated approach to the issue of transportation data collection and analysis for the Texas-Mexico Border Region. This report describes a prototype geographic information system (GIS)-based framework for transportation data with a goal to better understand the characteristics of the transportation system along border areas. The report focuses on highway network data, crash data, and traffic volume data in the Laredo – Nuevo Laredo area. The report examines existing transportation data sources at various jurisdictional levels and evaluates the degree to which the data can be integrated. The report also summarizes an assessment of transportation data on the Mexican side of the border that resulted from a collaborative effort with the Instituto Mexicano del Transporte (IMT).

EXECUTIVE SUMMARY

Urban areas along the Texas-Mexico border have experienced phenomenal growth rates in recent years. Transborder trade and transportation volumes are also growing rapidly as well. Despite recent economic slow downs and the implementation of tighter security measures at the border, most estimates point to a continued growth in the region in the foreseeable future. It is therefore critical for planners, engineers, and decision makers to develop a clear understanding of the various processes that drive, and are affected by, such growth. A number of government agencies are responsible for collecting and processing transportation-related data in the border region. Unfortunately, the available data tend to be scattered among various organizations and rarely include substantial amounts of information from the Mexican side of the border. There is a need to provide an integrated approach to the issue of transportation data collection and analysis for the Texas-Mexico Border Region. This research developed a prototype GIS-based framework for transportation data with a goal to better understand the characteristics of the transportation system along border areas.

The research examined existing transportation data sources at various jurisdictional levels and evaluated the degree to which these data sources could be integrated in a GIS environment. While some data were already available in GIS format, e.g., road networks, land cover, and land use, most of the data available were aggregated and not in a format suitable for inclusion in a GIS. Some other data, mainly volume data and crash data, which are critical data elements for most transportation analyses, contained some georeferencing elements; however, they were not in a GIS format. The researchers detected a lack of documentation in the transportation community with respect to procedures for documenting and integrating existing traffic volume and crash data into a GIS format. The decision was therefore made to focus on the development of those procedures.

Geocoding crash data varied depending on whether the crash occurred on a state highway or a local (jurisdiction) street. In the case of state highways, geocoding was possible by using control section and milepoint data. In the case of local streets, geocoding was possible by using street name and block number data. The researchers geocoded traffic volume data provided by TxDOT. Three types of traffic volume data were available: (a) data from permanent count stations around the state, including average daily traffic volumes and vehicle classification data; (b) data from count stations on state highways in the Laredo District; and (c) data from count stations on local (jurisdiction) streets in Webb County. In general, traffic volume data were not georeferenced, which meant that the researchers had to follow a manual approach for converting the data to a GIS format.

The report also summarizes an assessment of transportation data on the Mexican side of the border that resulted from a collaborative effort with IMT. IMT researchers evaluated existing data at the federal, state, and local levels in Mexico and produced a document that summarized their findings. They also focused on existing highway network, traffic volume, and crash data sources. IMT researchers faced a number of challenges derived from the lack of common data frameworks and standards for minimum data content and positional accuracy. They concluded, however, that the prototype system they developed for the Mexican side laid the foundation for more integrated approaches to transportation data management for the border region and more effective interagency collaboration.

After converting each data source to a GIS format, the final step was integrating the data to make sure all layers aligned properly. This process also involved integrating the data into a single geodatabase, which included dealing with coordinate system integration issues and base map discrepancies. In some cases, the researchers observed noticeable offsets when projecting data from one coordinate system to another and had to manually correct the offsets to ensure proper data overlaying. Also noticeable were the differences between the road base maps used to geocode data. The differences were not uniform, clearly indicating that a simple map translation and/or rotation transformation would not solve the problem. It would be ideal if the same road base map could be used for all data sources. Unfortunately, this is not feasible at present because of the dependency between some data sources and their underlying road base maps. As new data collection and data management systems with the capability to record latitude-longitude data in the field become available, the need to depend on road base maps to generate GIS features will decrease.

The research focused more on system architecture issues and design rather than applications, although the report did include a couple of application examples to illustrate some of the capabilities of the approach discussed here. In general, the research addressed two fundamental issues affecting the availability and use of traffic volume and crash data: georeferencing and integration. The research addressed these issues by developing a GIS-based prototype and by laying the foundation for the integration of additional data sources as these data sources become available. As a prototype, the system only handles a few data types. However, it is sufficiently generic and expansions appear feasible because of the GIS-based relational database structure that was chosen for its development.

Although the prototype system incorporates data elements from both Texas and Mexico, clearly a lot more needs to be accomplished to develop a truly binational transportation data management system. Transportation agencies on both sides of the border frequently interact to deal with traffic operation issues affecting crossborder traffic. If developed and properly implemented, a system based on the prototype discussed here could help make the crossborder interaction more effective, therefore increasing the chances that strategies that are intended to meet the needs on both sides of the border are successful. With the advent of Internet-based interactive mapping technologies that enable engineers and planners at remote locations to view maps and attribute data using commonly used web browsers, crossborder interaction could become even more effective. To illustrate this approach, the researchers implemented a web site that shows the different GIS data layers developed as part of the research. The web site can be accessed at http://imr.tamu.edu.

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

Page

LIST OF ACRONYMS, ABBREVIATIONS, AND TERMS

AADT	Annual average daily traffic
AADTT	Annual average daily traffic for trucks
ADT	Average daily traffic
ATR	Automatic traffic recorder
BLOB	Binary large object
CAD	Computer aided design
CSDGM	Content Standard for Digital Geospatial Metadata
ESRI	Environmental Systems Research Institute
FGDC	Federal Geographic Data Committee
FHWA	Federal Highway Administration
FK	Foreign key
GIS	Geographic information system
IMT	Instituto Mexicano del Transporte
LOS	Level of service
MPRME	Mile Point to Reference Marker Equivalence
MSA	Metropolitan statistical area
NMSU	New Mexico State University
NMDOT	New Mexico Department of Transportation
OCR	Optical character recognition
PDO	Property damage only
PDF	Portable document format
PFP	Policía Federal Preventiva
PK	Primary key
SCT	Secretaría de Comunicaciones y Transportes
SIGET	Sistema de Información Geoestadística para el Transporte
STARS	Statewide Traffic Analysis and Reporting System
TIFF	Tag image file format
TIGER	Topologically Integrated Geographic Encoding and Referencing
TMC	Turning movement count
TNRIS	Texas Natural Resources Information System
TxDOT	Texas Department of Transportation
TxDPS	Texas Department of Public Safety

CHAPTER 1. INTRODUCTION

Urban areas along the Texas-Mexico border (Figure 1) have experienced phenomenal growth rates in recent years. According to the U.S. Census Bureau (2001), the McAllen-Edinburg-Mission, Laredo, and Brownsville-Harlingen-San Benito Metropolitan Statistical Areas (MSAs) ranked 4th, 9th, and 28th, respectively, among 280 MSAs around the country in population growth rates between 1990 and 2000. Transborder trade and transportation volumes are also growing rapidly as well. For example, between 1987 and 2002, total loaded truck crossings at the Laredo inland port grew 13% per year on average (Laredo Development Foundation, 2003). During the same period, total loaded rail car crossings in Laredo grew 10% per year on average. Laredo alone accounts for 38% of all transborder shipments between the U.S. and Mexico and 51% of all total trade through Texas. Despite economic slow downs and the implementation of tighter security measures at the border, most estimates point to a continued growth in the region in the foreseeable future. It is therefore critical for planners, engineers, and decision makers to develop a clear understanding of the various processes that drive, and are affected by, such growth.

Figure 1. Texas-Mexico Border Region

A number of government agencies are responsible for collecting and processing transportation-related data in the border region. Unfortunately, the available data tend to be scattered among various organizations and rarely include substantial amounts of information from the Mexican side of the border.

There is a need to provide an integrated approach to the issue of transportation data collection and analysis for the Texas-Mexico Border Region. This report describes a prototype GIS-based framework for transportation data with a goal to better understand the characteristics of the transportation system along border areas. The report focuses on highway network data, crash data, and traffic volume data in the Laredo – Nuevo Laredo area. The report examines existing transportation data sources at various jurisdictional levels and evaluates the degree to which these data sources can be integrated in a GIS environment. The research also summarizes an assessment of transportation data on the Mexican side of the border that resulted from a collaborative effort with the Instituto Mexicano del Transporte (IMT).

This report is organized in chapters as follows:

- Chapter 1 is this introductory chapter.
- Chapter 2 describes transportation data sources and the database architecture.
- Chapter 3 describes the process for collecting and integrating crash data.
- Chapter 4 describes the process for collecting and integrating volume data.
- Chapter 5 describes data on the Mexican side of the border.
- Chapter 6 describes data integration issues.
- Chapter 7 summarizes the main research findings.

CHAPTER 2. DATA SOURCES AND DATABASE ARCHITECTURE

DATA SOURCES

A number of local, state, and federal government agencies collect and process transportation-related data in the Texas-Mexico Border Region. In addition to government agencies, an increasing number of commercial organizations also collect and process transportation-related data, e.g., road base maps and high-resolution imagery. This section provides a brief summary of some of the agencies and the type of data they collect, with a focus on data that can be represented in a GIS format. To narrow down the search, the analysis focuses on data available in the Laredo – Nuevo Laredo area. This chapter focuses mostly on data collected on the U.S. side, although, to the extent possible, it also includes a description about data available on the Mexican side. Chapter 5 provides a more in-depth description of data in the Nuevo Laredo area, which resulted from a collaborative effort with IMT.

Binational Border Transportation Planning & Program Process

In 1994, the U.S. and Mexico signed a Memorandum of Understanding creating a Joint Working Committee to coordinate planning and programming efforts of intermodal projects along the U.S.-Mexico border. In 1998, the Joint Working Committee completed a planning and programming study that evaluated state and national transportation planning processes in the U.S. and Mexico, reviewed available data on border transportation infrastructure and goods movement, and recommended an ongoing, binational planning and programming process (Binational Border Transportation Planning and Programming Study, 1998). The study resulted in a series of reports documenting the planning and programming process along with GIS data and a planned and programmed border transportation project database. Data in GIS format, most of which are already available from other sources, include highway networks, railway networks, state boundaries, county boundaries, ports, airports, truck/rail transfer facilities, and border crossings.

It may be worth noting that New Mexico State University (NMSU) in conjunction with the New Mexico Department of Transportation (NMDOT) is developing a binational transportation GIS for the Joint Working Committee (Czerniak, R., 2003). Currently, a pilot database includes transportation facilities—roads, rail, airports, and ports of entry—within 60 miles of the border in the states of Chihuahua and New Mexico. As part of a second phase, the database will be expanded to the entire border and will include layers such as satellite images and land use within one mile of the ports of entry, marine ports, and attribute data such as road functional classes, traffic volumes, capacities, and levels of service (LOS) for major corridors.

Laredo Development Foundation

The Laredo Development Foundation (2003) is a private, non-profit corporation dedicated to marketing Laredo for economic development opportunities. It focuses on industrial attraction to Laredo, workforce development, assistance to small business start-ups, as well as expansion and retention of existing industry. Some of the data available at the Foundation's web site include U.S. exports through Texas border ports, Laredo-Nuevo Laredo transportation system, Port of Laredo primary routes, Laredo regional highway system, U.S.-Mexico trade, crossborder loaded trucks and cars, airline passengers and air cargo, MSA population and employment, and electronic and telephone connections. The data are aggregated and not in a format suitable for inclusion in a GIS.

Laredo Police Department

Like other law enforcement agencies in the U.S., one of the functions of the Laredo Police Department is the enforcement of traffic laws. In this capacity, the Department responds to crash scenes, conducts related investigations, prepares accident reports, and sends copies of the reports to the Texas Department of Public Safety (TxDPS, 2003) for processing and compilation into the official state crash database. The Department also maintains a separate database (in Polaris format) of all crashes that occur within its jurisdiction based on information derived from the accident reports prepared by police officers in the field. As a subsequent section describes, crash data are disaggregated at the individual crash level and include references to crash locations. However, it would be necessary to conduct a substantial amount of processing to make the data suitable for inclusion in a GIS.

Texas Center for Border Economic and Enterprise Development

The Texas Center for Border Economic and Enterprise Development (2003) is a consortium of three universities in Texas. One of the focus areas of the Center includes the dissemination of U.S.-Mexico border socio-economic data. The database includes a number of aggregated data elements such as U.S. import/export data, border business data, border crossing data, and Texas regional indicators. Most of the data are aggregated, which limits the possibilities for integration into a GIS environment. For example, the border crossing data are port-aggregated data that include inbound and outbound monthly volumes of vehicle, pedestrian, truck, and rail crossings for all inland ports along the U.S.-Mexico border. However, a port is actually composed of several border crossings, each one having different locations and operational characteristics. This is the case in Laredo, where the Port of Laredo includes International Bridges #1 and #2 (which handle passenger car traffic and are located just south of downtown Laredo), the railroad bridge (which is also located south of downtown Laredo), the World Trade Bridge (which only handles commercial vehicle traffic and is located in northwest Laredo), and the Colombia/Solidarity Bridge (which handles mostly commercial vehicle traffic and is located some 25 miles west of Laredo).

Texas State Data Center

The Texas State Data Center (2003) is one of 52 centers around the country that compose the State Data Center System. The Center provides a state level liaison to the U.S. Census Bureau. It also provides population estimates and projections for inter-census years. The Center publishes thematic maps (in portable document format—PDF), statistics, trends, and reports, many of which are U.S. Census Bureau products. It also provides links to downloadable Census 2000 Topologically Integrated Geographic Encoding and Referencing (TIGER)/Line files.

Texas Department of Public Safety (TxDPS)

Among other functions, TxDPS is responsible for supervising traffic on rural highways in Texas, for supervising and regulating commercial traffic in the state, and for managing regulatory programs in driver licensing, motor vehicle inspection, and safety responsibility. TxDPS is the official repository of crash data in the state. In this capacity, TxDPS processes and compiles crash data from thousands of accident reports sent to the agency every year throughout the state (TxDPS, 2003). While not fully GIS-compatible, the TxDPS database does include data elements that enable geocoding of crash data. As a subsequent section describes, the procedure to geocode crash data varies depending on whether the crash occurred on a state highway or a local (jurisdiction) street.

Texas Department of Transportation (TxDOT)

TxDOT (2003) is the official agency in charging of developing and maintaining state highways—some 80,000 miles—in Texas. TxDOT generates enormous amounts of data, both at the system wide level and at the local project level. At the system wide level, TxDOT maintains a GIS-based inventory of state highway features, including basic geometric information such as number of lanes, lane width, and pavement conditions. TxDOT also maintains a road base map covering all state and local roads in the state. This road base map provides the basis for the transportation urban files that the Texas Natural Resources Information System (TNRIS) catalogues and publishes on their web site (TNRIS, 2003). At the local project level, TxDOT develops and maintains substantial amounts of data including construction and as-built drawings, construction and maintenance records, right-of-way maps, and signing. Many of the drawings are in Microstation dgn format and are georeferenced using the State Plane coordinate system.

In addition to the road base map and the inventory of highway features, TxDOT collects traffic volume data on state and local jurisdiction highways. Volume data on state highways are the result of an annual counting effort that involves the deployment of 24-hour counters as well as a few permanent count stations that are located at strategic locations throughout the state. Some of the stations provide vehicle classification data in addition to regular traffic volume data. Volume data on local jurisdiction roads is the result of sporadic data collection efforts, mainly on major and minor arterials, major collectors, as well as minor collectors in the immediate vicinity of arterials and major collectors. Both state highway volume data and local jurisdiction volume data are available in non-georeferenced Microstation dgn format. It may be worth noting that TxDOT is developing a Statewide Traffic Analysis and Reporting System (STARS) for handling historical and projected volume data in a GIS environment, but this system will only handle traffic data collected on state highways.

Through its International Relations Office, TxDOT also maintains an inventory of bridges and border crossings along the Texas-Mexico border (TxDOT, 2002). The inventory, which TxDOT updates every year, provides descriptive information, planned improvements, as well as connecting roadway and general traffic data for each bridge.

Texas Natural Resources Information System (TNRIS)

TNRIS (2003) is a centralized data warehousing system for maps, aerial photos, and digital natural resource data in Texas. TNRIS manages a variety of GIS data including administrative areas (county boundaries, legislative districts, zip codes), aerial photography, land use and land cover, soils, elevations, waste sites, water supply, aquifers, streams, wetlands, and transportation (airports, roads, waterways). As mentioned previously, TxDOT develops and maintains the transportation urban files published by TNRIS.

U.S. Census Bureau

The U.S. Census Bureau (2003) is a federal agency in charge of providing data about trends in population and the economy in the U.S. Examples of data published by the U.S. Census Bureau include Census 2000 data trends, 1997 economic census, business patterns, import and export activity, housing, and population certifications. The Census Bureau also publishes a substantial amount of thematic maps (e.g., census tract outline maps, county block maps, urbanized area outline maps, and voting district outline maps). It also develops and maintains the TIGER/Line files, which represent the system and associated digital database that support the mapping needs for the decennial census and other programs. TIGER/Line files include geographic features such as roads, railroads, rivers, lakes, legal boundaries, and census statistical boundaries covering the entire U.S. The database also contains attribute data about these features such as location (in latitude and longitude), name, feature type, address range for most streets, and geographic relationship to other features.

DATA NEEDS

While some data are already available in GIS format, e.g., road networks, land cover, and land use, most of the data available are aggregated and not in a format suitable for inclusion in a GIS. Some other data, mainly volume data and crash data, which are critical data elements for most transportation analyses, contain some georeferencing elements; however, they are not in an explicit GIS format. Some of the agencies that develop and maintain those data elements have started programs to develop GIS-based inventories of volume and crash data. However, it will be several years before those programs are in place. In any case, the researchers detected a lack of documentation in the transportation community at large with respect to procedures for documenting and integrating existing traffic volume and crash data into a GIS format. The decision was therefore made to focus on the development of those procedures.

It may be worth noting that the research did not address border crossings. A number of organizations compile and publish border crossing trends and statistics. Unfortunately, most of the data available are aggregated by port (i.e., data are not readily available at the individual border crossing level). TxDOT (2002) publishes an annual report describing individual bridges; however, it does not include traffic volume data. Because the report only provides general descriptive information about each bridge, it was decided not to focus the research effort on building a database schema just to accommodate that piece of information. If individual bridge traffic data become available, the database schema described below could be very easily expanded to accommodate the new data source.

DATABASE SCHEMA

There are several types of database models, including hierarchical, network, relational, object, and object relational (Elmasri and Navathe, 1989). The relational database model represents data as a collection of relations between tables, where a table represents a collection of related data values (which can be in a variety of formats such as text, number, and date/time). Some database packages, e.g., Oracle and Microsoft Access, also support data types that enable the storage of objects such as text files, images, or computer aided design (CAD) drawings. GIS applications such as ESRI's ArcGIS now support this "object relational" database model through the use of "geodatabases" that enable the storage of all GIS data, both graphical and tabular, in a single relational database file.

The researchers organized the database around data layers. A layer defines the elements and procedures related to the display of the geographic data it references. The researchers used three types of layers: background layers, road network layers, and data event layers (Figure 2). The background layers were essentially digital images representing aerial photographs that showed landmarks and other features on the ground. The road network layers were layers that represented road features. Shape files included geometric features (polylines in ArcGIS jargon) and attribute data. Some road network files contained polylineM features that contained distance "measures," which were useful for locating features on a map based on route and cumulative distance data. Data event layers were the top-level layers used to display the location of traffic volume data and crash data in relation to the underlying road network layers.

Figure 3 shows a summarized view of the database schema. The database schema includes five basic groups of tables: road network data, intersection data, document data, traffic volume data, and crash data. It may be worth noting that the database schema in Figure 3 is only a simplified diagram of the complete database. For simplicity, this report only describes the tables and relations that pertain to basic geographic features, and not the indexes and other related tables that ArcGIS automatically generates when it creates new geodatabases.

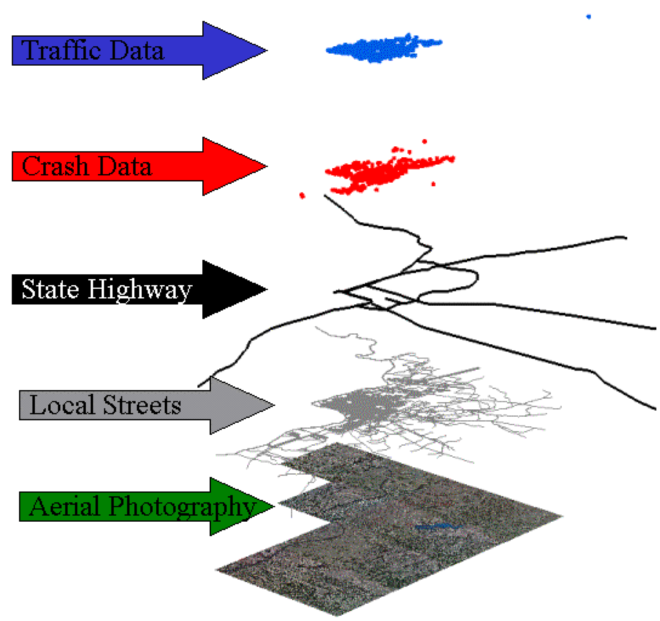


Figure 2. Map Layers

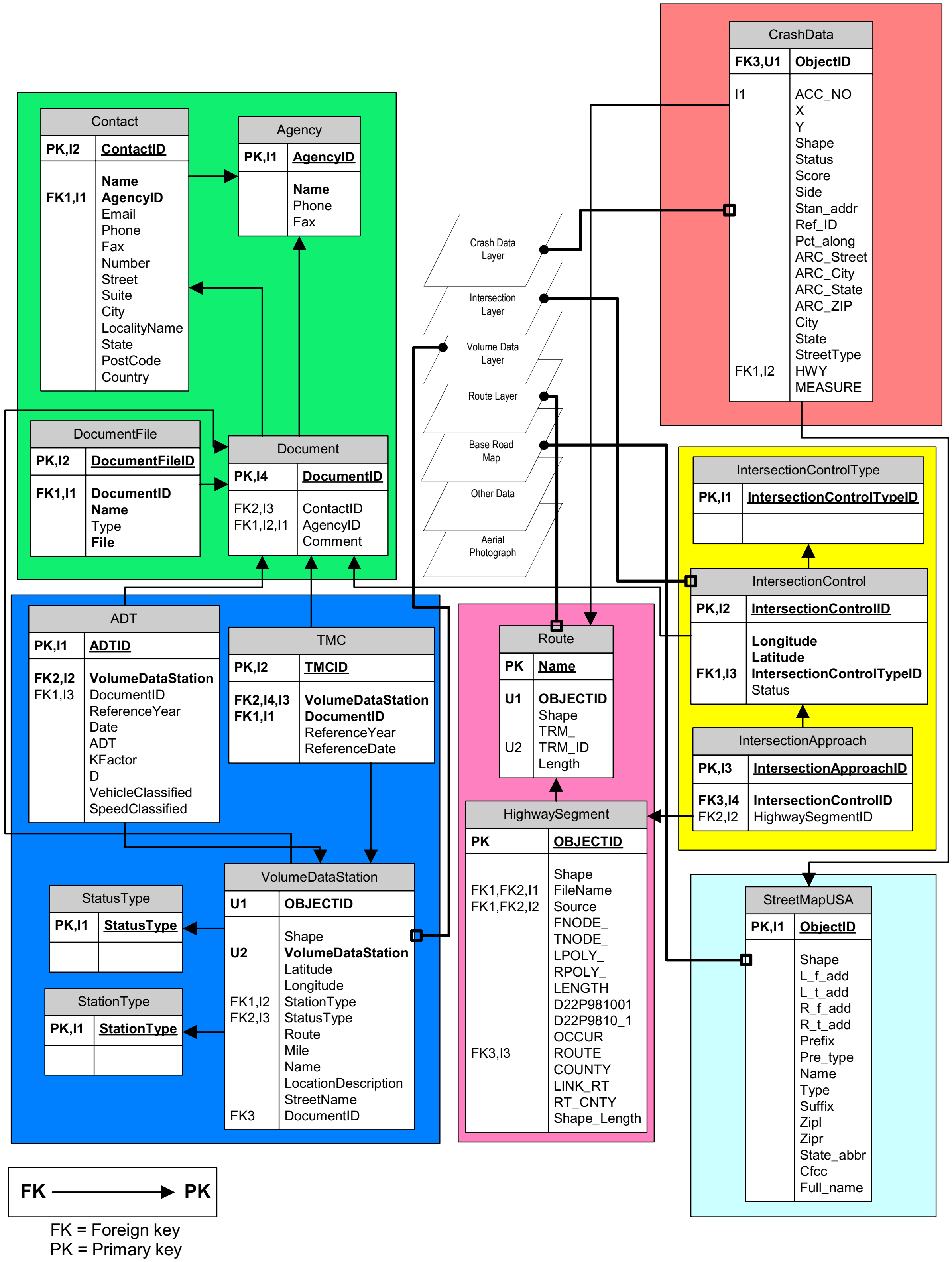


Figure 3. Database Schema

ROAD NETWORK DATA

Different entities manage state highway data and local (jurisdiction) street data. TxDOT manages state highway data. It also maintains a road base map covering all state and local roads in the state. Local jurisdictions usually manage their own local street databases. Some of the state highways TxDOT manages are in urban areas. As a result, there is some overlap between the state database and the local jurisdiction street databases. In addition to these databases, a number of commercial databases exist. For example, ESRI (2003a) distributes a series of street maps—called StreetMap USA—that enable the geolocation of features based on local street name and block number data. Other vendors, such as Navegation Technologies (2003) and Tele Atlas (2003), also produce electronic street map databases that enable geolocation and routing.

In addition to the road base map, TxDOT maintains a route-based inventory of state highway features. Every state highway has a route representation in the GIS (through a polylineM feature) that results from connecting adjacent highway segments and by associating cumulative distances—or measures—to vertices along the route. Figure 4a shows a schematic representation of state highway segments and routes. A brief description of the tables follows:

- HighwaySegment: this table contains basic segment data.
- Route: this table contains information needed to store routes including ID, Name, Direction, and Length.
- StreetMap USA: the researchers used the ESRI StreetMap USA local street database to geocode crash data that were not located on state highways. StreetMap, which is based on U.S. Census Bureau TIGER/Line files, includes layers such as roads, major landmarks, and zip codes. For simplicity, the researchers only used the local street component of the StreetMap USA database. As Figure 4 shows, geolocation of events is possible by using street name and block number data included in the StreetMap USA database. This dataset includes street name and block number data (from and to data on both sides of the street to indicate beginning and ending address ranges within individual blocks) to enable geocoding. For example, Figure 4b shows that the left and right from and to address indicators between 500 and 600 Main Street are given by L_from_add = 501, L_to_add = 599, R_from_add = 500, and R_to_add = 598.

INTERSECTION DATA

For completeness, the researchers integrated basic intersection data into the database schema. The intersection schema is a database representation of an intersection, its approaches, and its traffic control (Figure 5). The schema is modular and enables future expansions to include detailed intersection-related data such as lane configuration, signal timing phasing, and geometric characteristics. A brief description of the tables used follows:

- Intersection: this table is the database representation of an intersection containing ID, Long, Lat, IntersectionControlTypeID, DeploymentDate, Status, DocumentID.
- IntersectionControlType: this table is a lookup table for the types of intersection controls.

- IntersectionApproach: this table links the intersection to all the segments that approach the intersection.
- Document: this table stores intersection document data (see Document Data section).

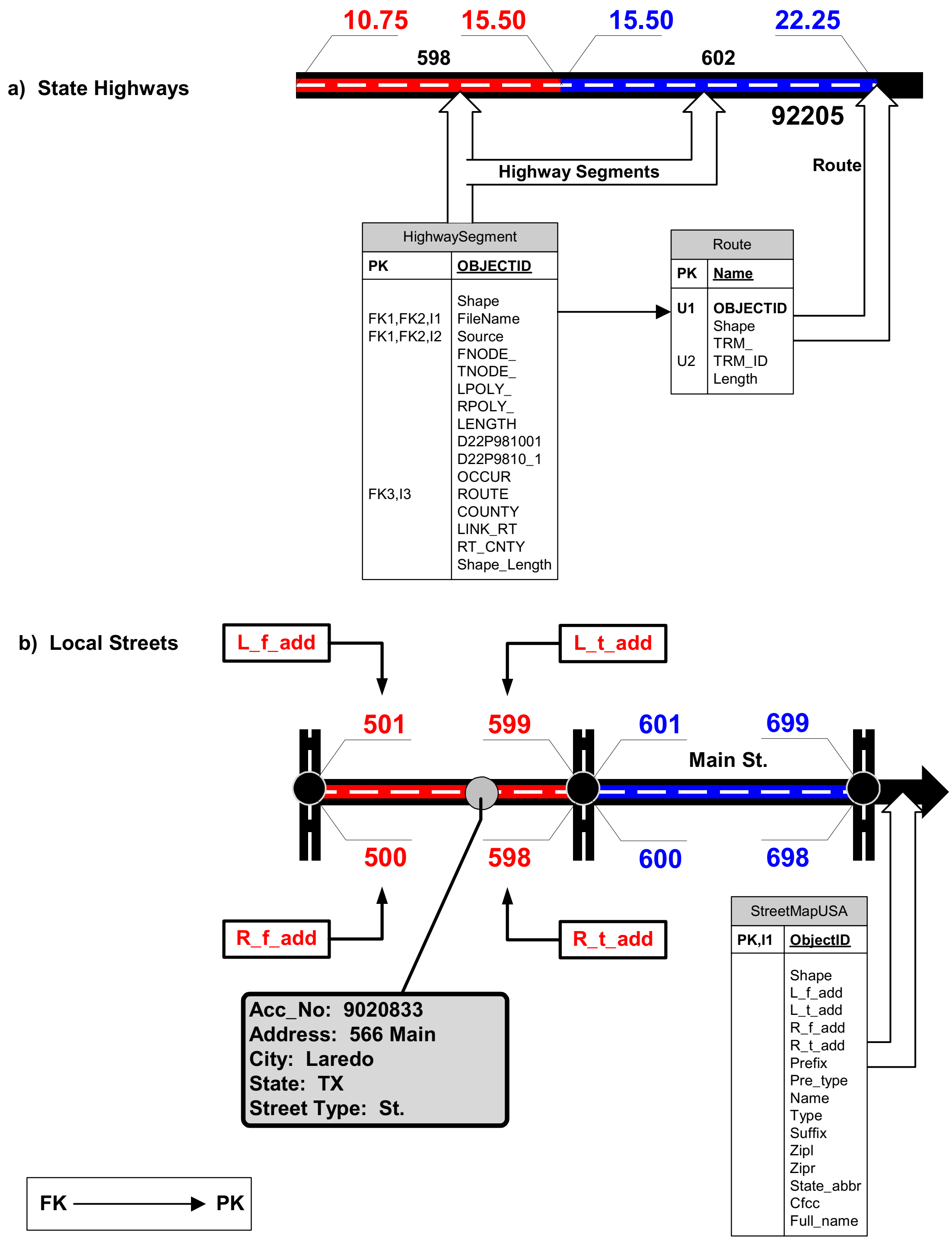


Figure 4. Database Representation of State Highways and Local Streets

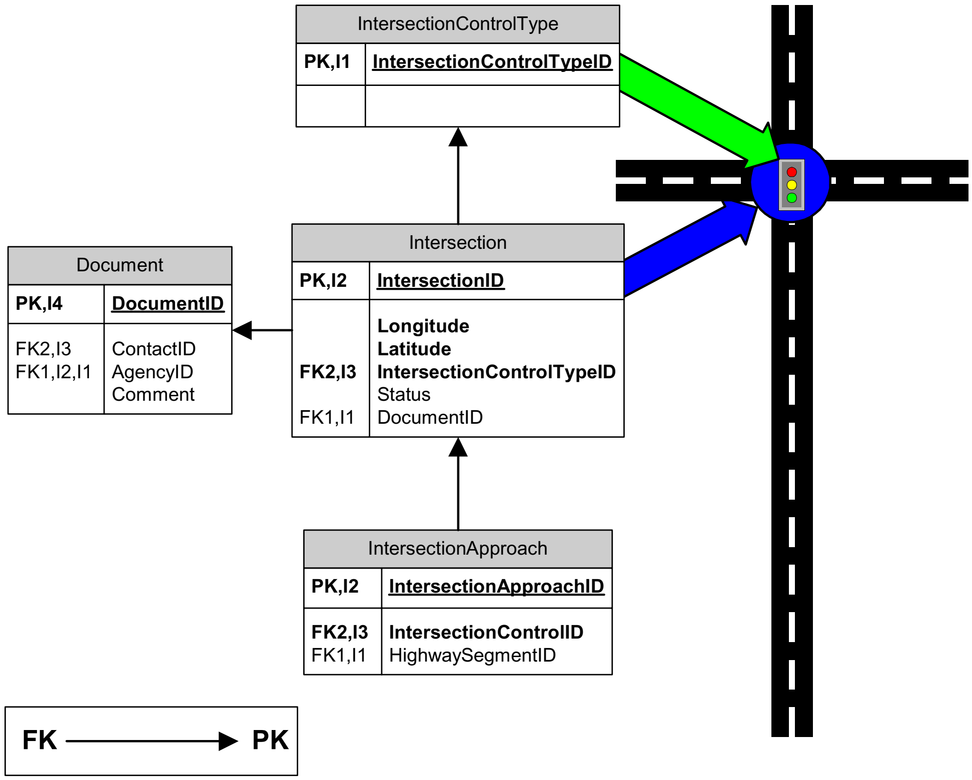


Figure 5. Database Representation of Intersection Data

DOCUMENT DATA

Document data describing electronic data files play a critical role in the database design. In general, a deliverable is any piece of electronic data stored in its original format that is associated with other data elements in the database. As Figure 6 shows, the database represents deliverable documentation using four tables:

- Document: this table is the main table for storing the deliverable data.
- DocumentFile: this table stores an electronic copy of the data file along with its name and type.
- Contact: this table stores data about a specific person that is responsible for the deliverable.
- Agency: this table stores data about the agency responsible for the deliverable.

It may be worth noting that this representation of document data is the foundation for a more comprehensive metadata representation that is compatible with current digital geospatial metadata standards (FGDC, 2003).

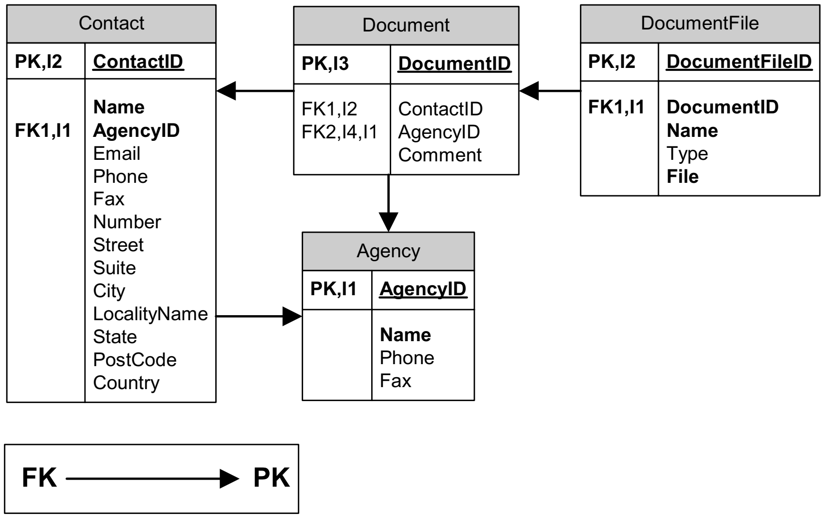


Figure 6. Database Representation of Document Data

TRAFFIC VOLUME DATA

A traffic volume data station refers to a location where traffic volume counts are collected. For the purpose of this research, two types of locations are of interest: segment count stations (which are normally located on sections of roadway outside the areas of influence of intersections) and intersection count stations. While intersection count stations logically involve all the approaches to an intersection, this research, for simplicity, treats intersection count stations as point features.

In the GIS, a geodatabase handles traffic volume data stations. Figure 7 shows a simplified version of the database schema. This schema includes five tables, as follows:

- VolumeDataStation: this table is a geodatabase table that contains both traffic volume station basic attribute data and the GIS representation of the station as a binary large object (BLOB).
- StationType: this table is a lookup table that stores possible station types (average daily traffic—ADT, turning movement count—TMC). The StationType field determines whether the volume data refers to ADT data or TMC data.
- StatusType: this table is a lookup table that stores station status types (mobile, permanent).
- ADT: this table stores ADT and associated data.
- TMC: this table stores basic metadata associated with the TMC data.

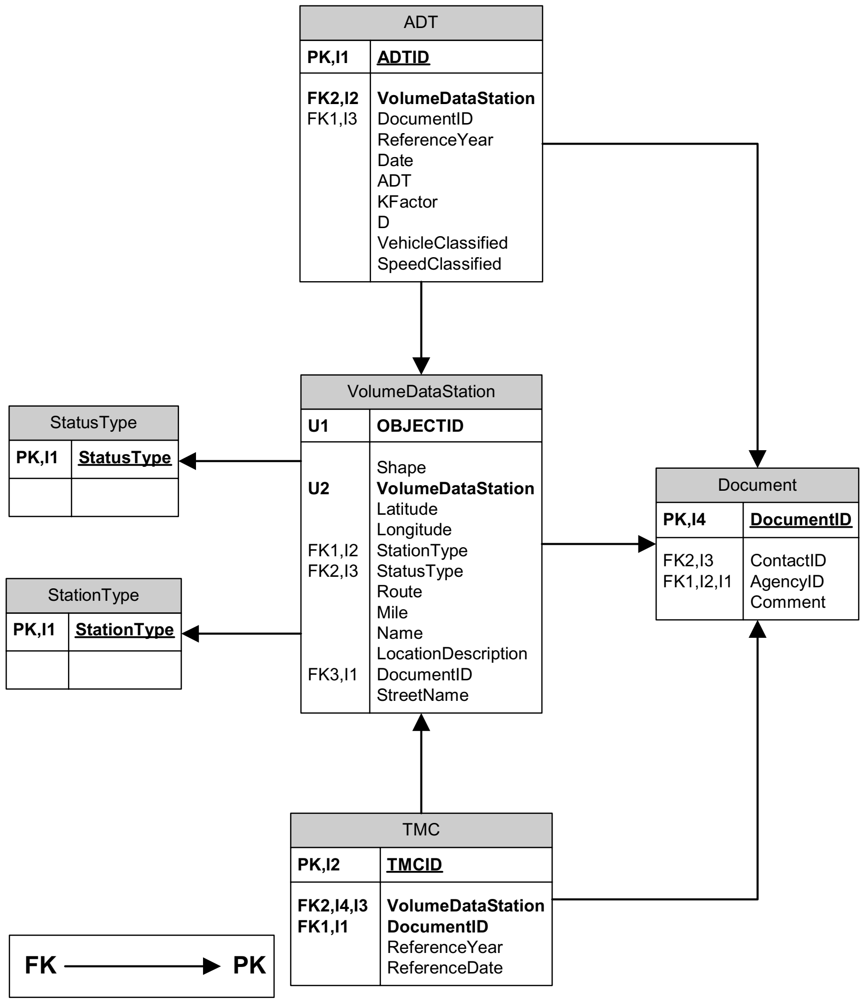
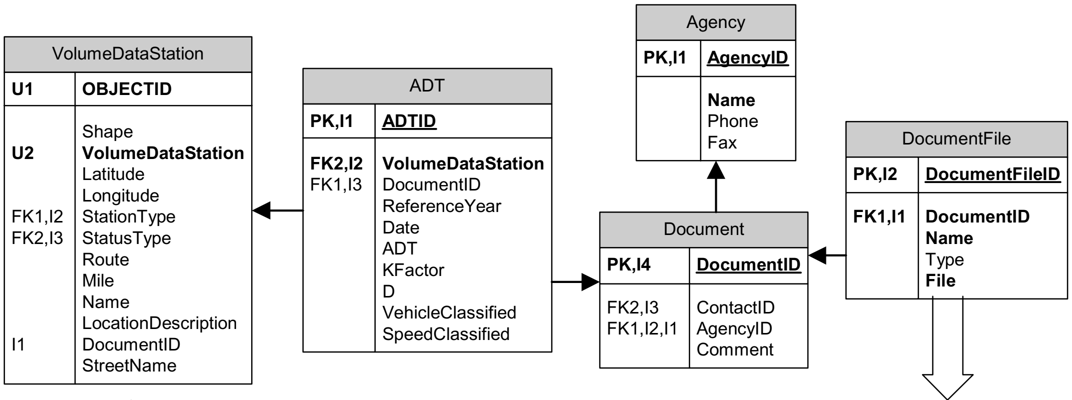


Figure 7. Database Representation of Volume Data

ADT Data

ADT data, usually obtained with the help of automatic traffic recorders (ATRs), represent 24-hr traffic volumes through a road section. The database handles two types of data recorders: permanent, which are recorders that have a fixed location and normally collect traffic data over long periods of time, and mobile, which are designed to be moved from location to location. Figure 8 shows a simplified version of the ADT database schema. This schema includes five tables, as follows:

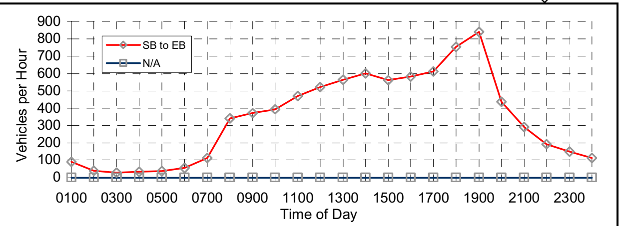
- ADT: this table stores ADT and associated data.
- VolumeDataStation: this table is a geodatabase table that contains traffic volume station basic attribute data.
- Document: this table stores ADT document data (see Document Data section).
- DocumentFile: this table stores an electronic copy of the ADT data file along with its name and type (see Document Data section).
- Agency: this table stores data about the agency responsible for the ADT deliverable (see Document Data section).



Texas Transportation Institute

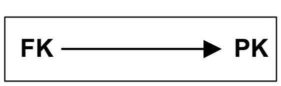
Automatic Traffic Recorder Counts

Project No. :	408150
Station No. :	1900
Counter No. :	9929
City/State:	Laredo, TX
Date:	11/28/2001
Day of Week:	Wednesday
Site:	SB Mines Rd. to EB Loop 20 Connecter
Notes:	



Direction	SB to EB		N/A	
End Time	15 Min	1 Hr	15 Min	1 Hr
0015	47			
0030	23			
0045	16			
0100	3	89		0
0115	7			
0130	10			
0145	14			
0200	8	39		0
0215	7			
0230	10			
0245	6			
0300	5	28		0
0315	6			
0330	5			
0345	16			
0400	6	33		0
0415	11			
0430	8			
0445	9			
0500	9	37		0
0515	9			
0530	19			
0545	15			
0600	12	55		0
0615	17			
0630	17			
0645	27			
0700	50	111		0
0715	65			
0730	70			
0745	99			
0800	107	341		0
0815	108			
0830	102			
0845	64			
0900	98	372		0
0915	83			
0930	106			
0945	102			
1000	103	394		0
1015	89			
1030	115			
1045	131			
1100	135	470		0
1115	128			
1130	143			
1145	111			
1200	140	522		0

Direction	SB to EB		N/A	
End Time	15 Min	1 Hr	15 Min	1Hr
1215	142			
1230	132			
1245	151			
1300	139	564		0
1315	133			
1330	151			
1345	181			
1400	137	602		0
1415	146			
1430	142			
1445	130			
1500	143	561		0
1515	149			
1530	152			
1545	130			
1600	152	583		0
1615	149			
1630	170			
1645	138			
1700	155	612		0
1715	190			
1730	186			
1745	198			
1800	179	753		0
1815	265			
1830	243			
1845	204			
1900	128	840		0
1915	129			
1930	99			
1945	110			
2000	100	438		0
2015	77			
2030	85			
2045	67			
2100	62	291		0
2115	68			
2130	45			
2145	43			
2200	36	192		0
2215	33			
2230	50			
2245	43			
2300	24	150		0
2315	36			
2330	30			
2345	24			
2400	22	112		0
24-Hour Directional Total:	8,189		0	
24-Hour Bidirectional Total:			8,189	



Note: As an example, table DocumentFile shows an Excel file documenting ADT count data obtained by TTI for the TxDOT Laredo District in 2001.

Figure 8. ADT Data Integration with Document Data

TMC Data

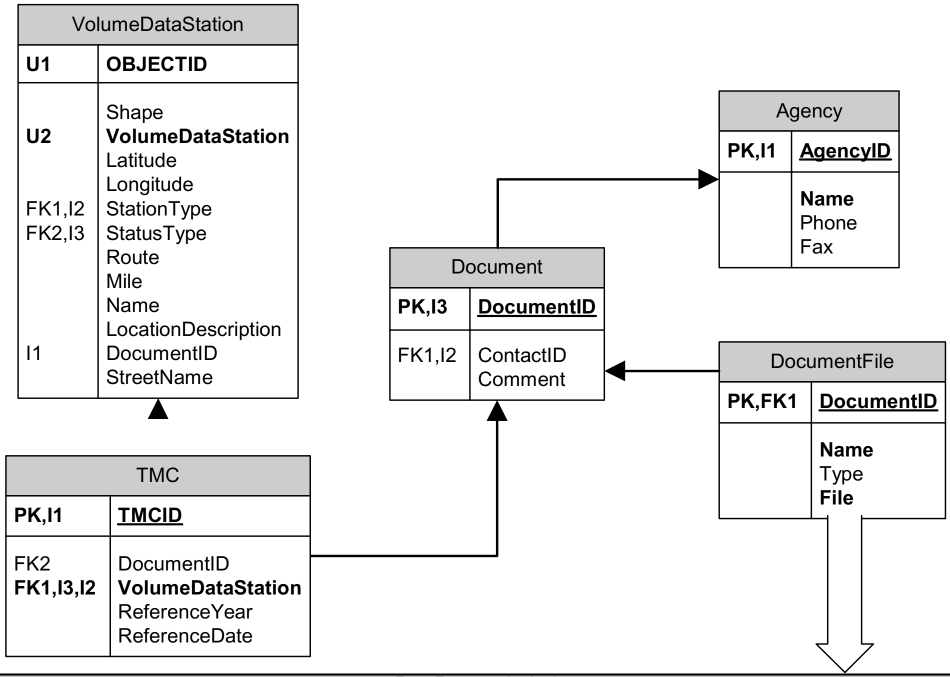
Unlike ADT data, which normally involve one or two (usually opposite) directions of travel, TMC data involve as many directions as traffic movements the intersection allows (which might include more than one mode of transportation). Figure 9 shows the database schema used in this research. The schema includes five tables, as described below:

- TMC: this table stores basic metadata associated with the TMC data, including a reference year and date, along with the VolumeDataStationID that was used to create the relationship with the GIS.
- VolumeDataStation: this table is a geodatabase table that contains traffic volume station basic attribute data.
- Document: this table is the main table for storing the deliverable data (see Document Data section).
- DocumentFile: this table stores the actual TMC counts are stored in their deliverable format as an object (see Document Data section).
- Agency: this table stores data about the agency responsible for the TMC deliverable (see Document Data section).

CRASH DATA

The researchers used sample crash data from the TxDPS state crash database. While not fully GIS-compatible, the TxDPS database does include data elements that enable geocoding of crash data. The procedure to geocode crash data varies depending on whether the crash occurred on a state highway or a local street. Figure 10 shows the database representation of the crash data. For simplicity, Figure 10 only shows fields from the state crash database that are relevant to the geocoding of crash data. However, through the crash number field (ACC_NO) it is possible to obtain additional information about individual crashes using the state database crash, driver, and casualty tables. A brief description of the basic tables follows:

- CrashEvents: a hybrid table that combines the original DPS data and the
- DPSCrashData: the source dataset compiled by the Texas Department of Public Safety.
- StreetMap USA: ESRI's StreetMap USA commercial database, which allows geocoding.
- StreetCodes: this table is a lookup table that contains a StreetCode, Street name, and StreetType for local streets in the Laredo area. This table is needed because crash data in TxDPS database use a five-digit street code representation, whereas the StreetMap USA database uses complete street names.
- HighwaySegment: this table contains basic segment data.
- Route: this table contains information needed to store routes this includes ID, Name, Direction, Length.



Texas Transportation Institute

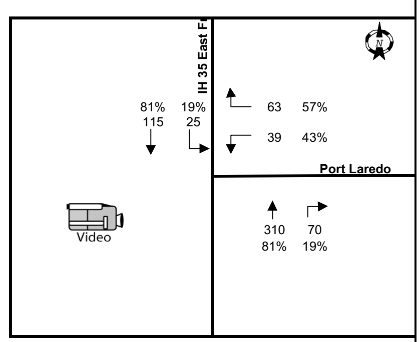
Location:	Laredo, Texas
Project & Project #:	408150
North-South street:	IH 35 East Frontage Road
East-West street:	Port Laredo
Time period (1-AM, 2-NOON, 3-PM):	3 3:00 - 5:00 PM
Date recorded:	November 26, 2002
Viewed by:	Ken Giusti, Jr.
Date viewed:	December 4, 2002

Time	*Northbound*						*Southbound*						*Eastbound*						*Westbound*					
Movement	*left*		*thru*		*right*		*left*		*thru*		*right*		*left*		*thru*		*right*		*left*		*thru*		*right*	
Vehicle Type	*C*	*T*	*C*	*T*	*C*	*T*	*C*	*T*	*C*	*T*	*C*	*T*	*C*	*T*	*C*	*T*	*C*	*T*	*C*	*T*	*C*	*T*	*C*	*T*
3:00 PM 3:15 PM			23	15	6	3	4	0	9	10									3	2			4	8
3:15 PM 3:30 PM			23	22	8	3	3	1	5	4									2	2			3	4
3:30 PM 3:45 PM			23	19	8	1	3	3	6	4									3	3			2	1
3:45 PM 4:00 PM			26	14	5	0	1	0	11	2									1	1			3	1
4:00 PM 4:15 PM			30	11	8	6	2	2	12	5									4	4			10	4
4:15 PM 4:30 PM			17	20	4	6	4	0	14	8									1	3			3	3
4:30 PM 4:45 PM			15	17	3	2	0	0	11	2									2	3			3	6
4:45 PM 5:00 PM			20	15	4	3	2	0	5	7									2	3			3	5
Total	0	0	177	133	46	24	19	6	73	42	0	0	0	0	0	0	0	0	18	21	0	0	31	32
Peak Total	0	0	96	64	25	13	10	5	43	19	0	0	0	0	0	0	0	0	9	11	0	0	18	9
Peak Movement Total	0		160		38		15		62		0		0		0		0		20		0		27	
Peak Turn Percent	0%		81%		19%		19%		81%		0%		#DIV/0!		#DIV/0!		#DIV/0!		43%		0%		57%	
Peak Approach Total	198						77						0						47					

Peak Hour 3:30 PM - 4:30 PM
Percent Trucks 41%

Time	*U-Turns*			
Approach:				
Vehicle Type	*C*	*T*	*C*	*T*
3:00 PM 3:15 PM				
3:15 PM 3:30 PM				
3:30 PM 3:45 PM				
3:45 PM 4:00 PM				
4:00 PM 4:15 PM				
4:15 PM 4:30 PM				
4:30 PM 4:45 PM				
4:45 PM 5:00 PM				
Total	0	0	0	0
Peak Total	0	0	0	0
Peak Movement Total	0		0	
Peak Turn Percent	0%		0%	

**Note: Train activity Northbound 3:45:15 pm - 4:00:30 pm*
Northbound 4:15:00 pm- 4:18:25 pm



FK ⟶ PK

Note: As an example, table DocumentFile shows an Excel file documenting TMC count data obtained by TTI for the TxDOT Laredo District in 2002.

Figure 9. Integration of TMC Data with Document Data

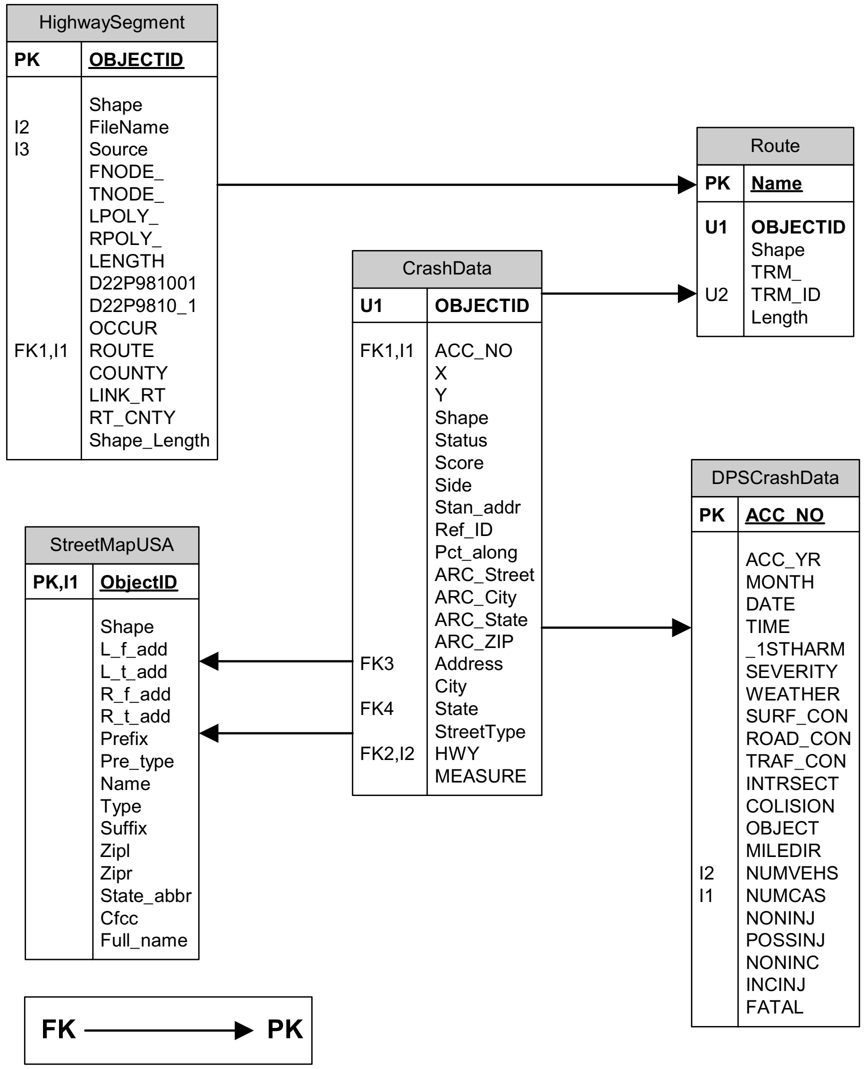


Figure 10. Database Representation of Crash Data

CHAPTER 3. COLLECTING AND INTEGRATING CRASH DATA

SOURCES OF CRASH DATA

A number of crash data sources exist in Texas, including the TxDPS state crash database, local police department crash databases, and hospital crash databases. By and large, these crash data sources are not georeferenced and are not compatible with each other. The basis for the TxDPS crash database, which the researchers used for this project, is the thousands of Texas "Peace Officer's Accident Report" forms (Form ST-3) that need to be sent to TxDPS whenever a crash exceeds some minimum thresholds established by law. The threshold for including property damage only (PDO) crashes in the TxDPS database has been increased several times (TxDPS, 1996). Between 1975 and 1995, the dollar damage threshold for PDO crashes (based on the investigating officers' estimate) was raised three times—from $25 per crash in 1975 to $250 worth of damage to the property of any one person in 1978 and to $500 per person in 1990. Starting July 1, 1995, only those PDO crashes that involved one or more vehicles being towed from the scene were entered into the state database. The threshold for authorizing a police officer to investigate a crash was increased to $1,000 in 2001 (House Bill 2230, 2001). The time threshold for assuming a fatality is associated with a crash in the TxDPS database has also changed: from within 90 days of the crash date in 1978 to within 30 days of the crash date in 1983 (TxDPS, 1996).

CRASH DATA LOCATION CODES

The procedure for geocoding crash data varies depending on whether the crash occurs on a state highway or a local (jurisdiction) street. In the case of local streets, geocoding is possible by using street name and block number data. In the TxDPS database, two fields, PRIM and BLOCK, represent street name and block number, respectively. If the crash occurs at an intersection, or the crash is intersection-related, the database might also contain an entry under a second street name field called SEC. In reality, the PRIM field contains a street name code, which means that, for geocoding purposes, it is necessary to also have a lookup table containing the equivalence between street name codes and street names.

In the case of state highways, geocoding is possible by using control section and milepoint data. In the TxDPS database, field CONTSEC1 represents a control section, and field MILE1 represents milepoint. If the crash occurs at the intersection of two state highways, the database might also contain entries for two additional control section and milepoint fields (CONTSEC2 and MILE2), which represent the secondary control section on which the crash occurred and the milepoint along this control section, respectively.

It may be worth noting that TxDOT uses both a control section-distance approach and a reference marker-distance approach for linearly referencing objects or events along the state highway network. With the control section-distance approach, the state highway network is divided into controls and sections and objects/events are located by determining their relative distance with respect to the beginning of the specific section. Practically all construction projects in the state are tied to the control section-distance model, and TxDPS uses this model to locate crash data on state highways.

With the reference marker-distance approach, the state highway network is divided into routes and objects/events are located by determining their relative distance from one or more reference markers that are physically located at strategic locations on all state highways. TxDOT uses the reference marker-distance approach for highway inventory purposes and for maintaining this inventory in a GIS format. To convert data from the control section-mile point system to the reference marker-distance system, TxDOT uses an equivalence lookup table called Mile Point to Reference Marker Equivalence (MPRME). TxDOT updates this table regularly.

CRASH DATA ON LOCAL STREETS

Table 1 shows a sample of records and fields from the TxDPS database with crashes on local streets. As mentioned previously, the researchers used the PRIM and BLOCK fields to geocode crash data on local streets. However, as Table 1 shows, the PRIM field contains a street name code, not a street name. The researchers requested from the Laredo Police Department a copy of a lookup table to translate the PRIM field code to a street name. Because the document was in paper format and some 44 pages long, the researchers scanned the pages and processed the resulting images through optical character recognition (OCR) software. After some additional post-processing to remove inconsistencies from the image-to-character conversion process, the result was a tab-delimited text file containing three fields: StreetName, StreetType, and StreetCode (Table 2).

Table 1. Local Street Source Crash Dataset

ACC_NO	CNTYCITY	COUNTY	THDCNTY	PRIM	DISTRICT	OTHERFAC	BLOCK
9295664	24001	240	240	SP500	002	00	1300
9018253	24001	240	240	SP500	003	11	1300
9024323	24001	240	240	SP500	002	00	1300
9125808	24001	240	240	SP500	002	00	1400
9001524	24001	240	240	SP500	002	00	1400
9151126	24001	240	240	SP500	002	00	1400
9183709	24001	240	240	SP500	004	28	1800

Note: CNTYCITY = 24001 corresponds to Laredo in the TxDPS crash database (240 corresponds to Webb County)

Table 2. Street Code Lookup Table

StreetName	StreetType	StreetCode
Spring Creek	Dr.	SP200
Spring	Rd.	SP300
Spring Valley	Cr.	SP400
Springfield	Ave.	SP500
Spruce	Ln.	SP600
St. Croix	Dr.	ST100
St. Pierre	Ln.	ST200
St. Thomas	Ct.	ST300

The researchers then imported the text file into Access and developed the following query that translated the PRIM field to a street name and, in combination with the BLOCK field, produced a street address (Table 3):

```
SELECT DPSCrashData1999.ACC_NO, [Block] & " " & [StreetName] AS Address, "Laredo" AS City, "TX"
AS State, [Street Codes].StreetType

FROM DPSCrashData1999 INNER JOIN [Street Codes] ON DPSCrashData1999.PRIM =[Street
Codes].StreetCode

WHERE IsNull(BLOCK) = False And IsNull(PRIM) = False And IsNull(MILE1) = True And
IsNull(ContSect1) = True And County = 240;
```

Table 3. Resulting Crash Street Address Dataset

ACC_NO	Address	City	State	StreetType
9140333	1300 Meadow	Laredo	TX	Ave.
9097835	1300 New York	Laredo	TX	Ave.
9024323	1300 Springfield	Laredo	TX	Ave.
9295664	1300 Springfield	Laredo	TX	Ave.
9018253	1300 Springfield	Laredo	TX	Ave.
9142188	1300 Tapeyste	Laredo	TX	Ave.
9301173	1400 Boston	Laredo	TX	St.
9028164	1400 Bustamante	Laredo	TX	St.
9243177	1400 Cedar	Laredo	TX	Ave.
9177632	1400 Cedar	Laredo	TX	Ave.
9265043	1400 Clark	Laredo	TX	Blvd.
9069582	1400 Coke	Laredo	TX	St.

Geocoding Process

Geocoding defines the process for creating geometric representations of geographic features out of location descriptions. As Figure 11 shows, geocoding local street crash data entailed five steps (ESRI, 2003b):

- Address standardization: This step involved manipulating the address data to a form acceptable to ArcGIS 8.2. The previous section described this step. ArcGIS supports a variety of geocoding "services" to take into consideration different ways to parse address components (e.g., prefix direction, prefix street type, street number, suffix street type, suffix direction, state, and zip code). The researchers used the StreetMap USA geocoding service, which explains the parsing shown in Table 3.
- Querying for similar standardized address components: This step involved querying features in the StreetMap USA dataset for similar standardized address components.
- Score assignment: This step involved computing composite scores to addresses by using individual address component matching scores and scoring weights based on match and unmatched probabilities (m and u, respectively, in Table 4). The geocoding process calculates matching scores for each address component (e.g., street name, city, state, and zipcode) depending on how closely the components match. In general, the better the match, the higher the individual score. The algorithm uses two types of score weights: if there is a match, the algorithm calculates the weight as the $\log_2$ (base 2) of the ratio of m to u. For example, if there is a match in street name, the corresponding weight is $\log_2$ (0.9/0.01), or 6.49. If there is a mismatch, the algorithm calculates the weight as the $\log_2$ of the ratio of 1-m to 1-u (the result is a negative weight). For example, if there is a

mismatch in street name, the corresponding weight is $\log_2$ [(1-0.9)/(1-0.01)], or -3.31. In general, agreements add to the composite score and disagreements subtract from it.

- Address matching: This step involved selecting possible match candidates based on a set of minimum pre-defined scoring thresholds (spelling, minimum candidate, and minimum match). After setting the thresholds, ArcGIS completed the address matching step automatically (although interactive matching is also possible). A short description of the thresholds follows:
    - Spelling threshold: This threshold controls how much variation in spelling the geocoding service allows. A lower setting allows less likely candidates while a high setting restricts candidates to exact spelling matches. For example, a spelling threshold of 80 (from 0 to 100) would reject the address “1400 Del Mar” (the closest street name in the StreetMap database is “Del Norte”). However, a spelling threshold of 20 would accept the address “1400 Del Mar” as a candidate match for “Del Norte.”
    - Minimum candidate threshold: This threshold determines how restrictive the algorithm should be in considering or rejecting a potential candidate. The algorithm rejects potential candidates with a score less than the threshold. For example, the algorithm would reject a candidate with a score of 50 if the minimum candidate threshold is 60. However, the algorithm would consider the candidate if the minimum candidate threshold is lowered to 40. Therefore, a lower setting is less restrictive in allowing for candidacy and a higher setting only selects candidates with very high scores.
    - Minimum match threshold: This threshold determines how well addressed a candidate must be to be considered a match. The algorithm rejects potential matches if they yield a score lower than the threshold. For example, a minimum match threshold of 60 would reject the address “1400 Del Mar” if its assigned score is lower than that accepted by the minimum match threshold. By comparison, a minimum match threshold of 20 would accept the address “1400 Del Mar” as a potential match.
- Feature creation: This step involved the creation of GIS point features representing crash locations based on the results from the address matching step.

Table 4. Default StreetMap Geocoding Match and Unmatched Probabilities (adapted from ESRI, 2002)

Match keys	Variables in the reference file	Match probability (m)	Unmatched probability (u)
ZP	LeftZip and RightZip	0.9	0.01
CT	LeftCity, RightCity	0.5	0.5
SA	State	0.8	0.1
SN	StreetName	0.9	0.01
PD	PreDir	0.8	0.1
PT	PreType	0.7	0.1
ST	StreetType	0.85	0.1
SD	SufDir	0.85	0.1
HN	FromLeft, ToLeft, FromRight, ToRight	0.999	0.05

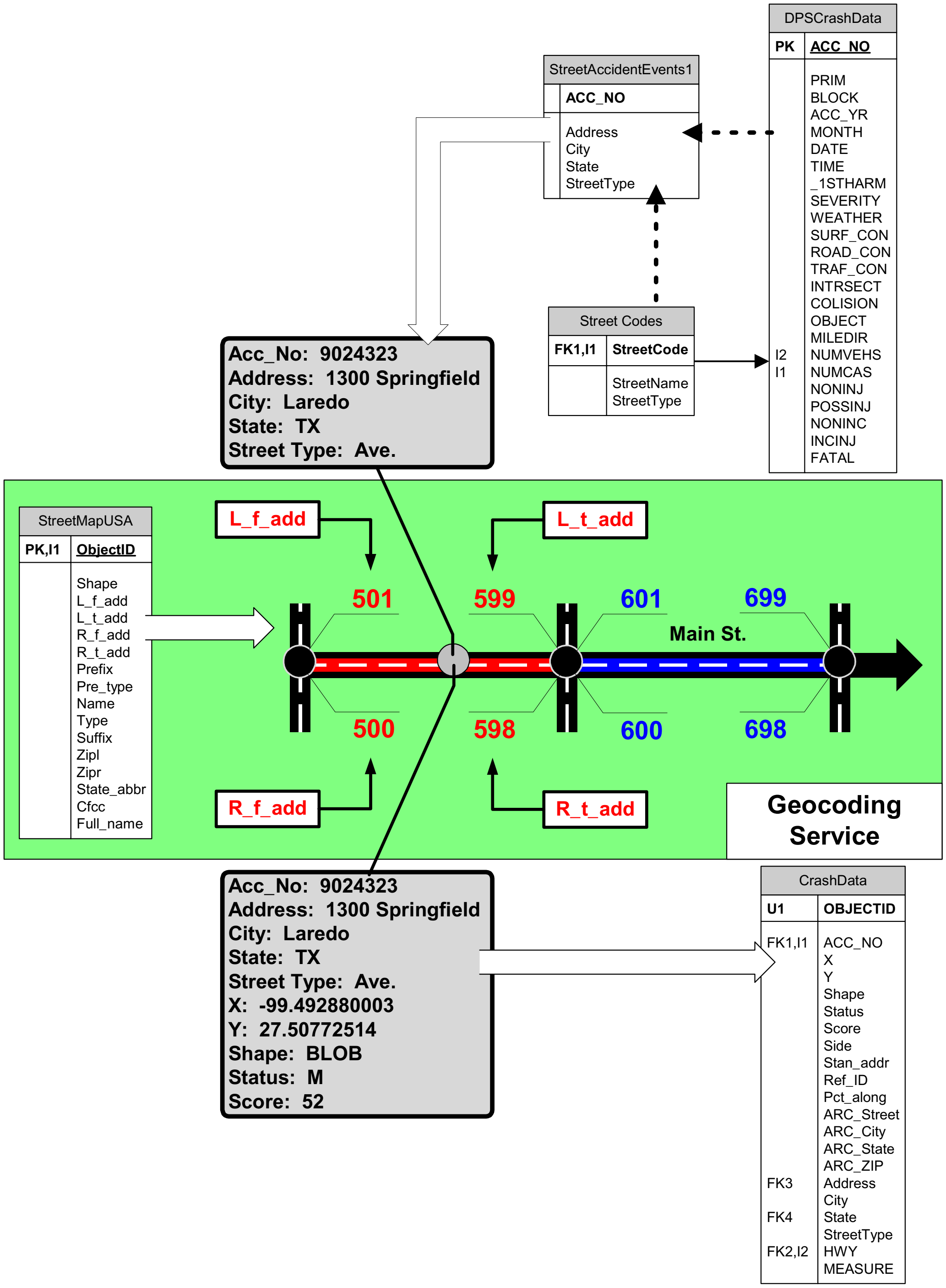


Figure 11. General Geocoding Process

Results

Figure 12 shows a sample of points depicting the location of crashes in local streets in Laredo. For simplicity, the researchers used the default m and u probability values shown in Table 4. The researchers also used the following scoring thresholds: spelling threshold (60), minimum candidate threshold (10), and minimum match threshold (40). In total, the researchers processed

1,050 year 1999 local street crash records. The geocoding process resulted in one matched record with a composite score higher than 80 (0.1% percent), 627 matched records with a composite score between 40 and 80 (59.7%), and 422 unmatched records (40.2%). Normally, scores of 80 or higher are considered good scores. However, the crash dataset did not include zip code data. As the following section shows with more detail, not having zip code data can result in scores that are up to 30 points lower than those with zip code data.

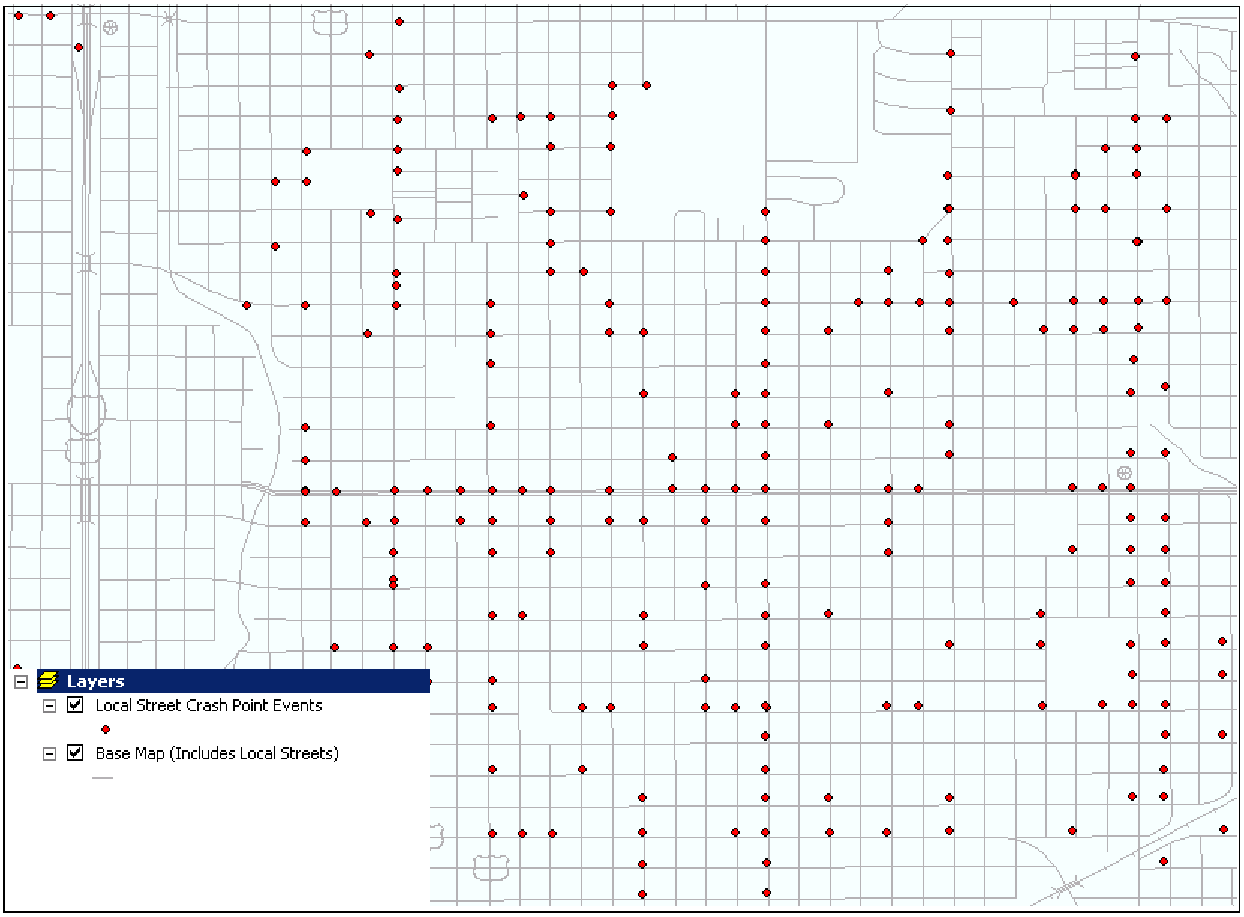


Figure 12. Geocoded Local Street Crash Data

Sensitivity Analysis

The researchers conducted a sensitivity analysis to better understand the effects of the StreetMap USA address component scoring and thresholds (spelling, minimum candidate, and minimum match) on the geocoding of crash data in the Laredo area. For the sensitivity analysis, the researchers used the following reference values: spelling threshold (60), minimum candidate threshold (10), and minimum match threshold (40).

<u>Address Component Scoring</u>

Many factors influence the total score assigned to an address, including whether there are missing characters in the street name, whether there are misspelled characters, and whether zip code data are included. To better understand the nature of these effects, the researchers fed the algorithm with simulated variations of the address "1400 Calle Del Norte, Laredo TX, 78041."

Table 5 shows the results of the simulation. In general, the researchers observed the following trends (although, in some cases, results were not linear):

- Street name misspelled: -10 points per misspelled character
- Missing street name character: -5 points per missing character
- No prefix ("Calle" is a prefix): -14 points
- Misspelled prefix: -65 points
- No zip code: -30 points
- No street number: -22 points

Spelling threshold

As mentioned previously, the spelling threshold controls how much variation in spelling the geocoding service allows. A lower setting allows less likely candidates while a high setting restricts candidates to exact spelling matches.

Figure 13 summarizes the results of the analysis. The percentage of matched crash records did not substantially change, except when the spelling threshold was set to 100. At first sight, this could be an indication that the spelling threshold is a "weak" threshold (therefore potentially not very useful) in that, regardless of value, the effect on the number of matched records remains essentially the same. The possibility also exists that the crash records contained addresses with relatively few spelling errors because the researchers used a lookup table for street names that involved a considerable amount of processing, checking, and validation. To further investigate this result, the researchers conducted additional tests on the effect of the spelling threshold, each time varying the value of the minimum match threshold (see section below). In all cases, the results were similar, i.e., the percentage of matched records remained the same as the spelling threshold changed (except when the threshold was set to 100).

Interestingly, there was an inverse correlation between the effect of the spelling threshold and the effect of the minimum match threshold. For example, when the minimum match threshold was set to 10, the percentage of matched records was 88. Likewise, when the minimum match threshold was set to 50, the percentage of matched records was 43. This observation tends to confirm the suspicion that the spelling threshold is a "weak" threshold. Obviously, a more definite conclusion would only be possible with additional data sets.

Table 5. Effect of Changing Street Address Components on Matching Scores

Crash Address	City	State	Zip Code	Score
Reference crash address ("Calle" is a prefix)				
1400 Calle Del Norte	Laredo	TX	78041	100
Street name misspelled				
1400 Calle Del Norto	Laredo	TX	78041	90
1400 Calle Del Norre	Laredo	TX	78041	90
1400 Calle Del Narto	Laredo	TX	78041	84
1400 Calle Del Norfo	Laredo	TX	78041	80
No prefix				
1400 Del Norte	Laredo	TX	78041	86
No prefix, street name misspelled				
1400 Del Nort	Laredo	TX	78041	81
1400 Del Nor	Laredo	TX	78041	76
1400 Del	Laredo	TX	78041	56
1400 De	Laredo	TX	78041	0
1400 Del Mar	Laredo	TX	78041	56
1400 Del Main	Laredo	TX	78041	56
1400 Norte Del	Laredo	TX	78041	0
1400 etron led	Laredo	TX	78041	0
No zip code				
1400 Calle Del Norte	Laredo	TX		70
No prefix, no zip code, street name misspelled				
1400 Del Norte	Laredo	TX		56
1400 Del Nort	Laredo	TX		51
1400 Del Nor	Laredo	TX		46
1400 Del	Laredo	TX		26
1400 Del Mar	Laredo	TX		26
1400 Del Main	Laredo	TX		26
No prefix, zip code misspelled				
1400 Del Norte	Laredo	TX	7804	77
1400 Del Norte	Laredo	TX	780	68
1400 Del Norte	Laredo	TX	41	56
1400 Del Norte	Laredo	TX	78040	68
Prefix misspelled				
1400 Call Del Norte	Laredo	TX	78041	36
1400 Cal Del Norte	Laredo	TX	78041	36
1400 Ca Del Norte	Laredo	TX	78041	34
1400 Coll Del Norte	Laredo	TX	78041	34
1400 Colle Del Norte	Laredo	TX	78041	34
No street number				
Calle Del Norte	Laredo	TX	78041	78
No street number, street name misspelled				
Calle Del Norto	Laredo	TX	78041	68
Calle Del Norre	Laredo	TX	78041	68
Calle Del Narto	Laredo	TX	78041	62
Calle Del Norfo	Laredo	TX	78041	58

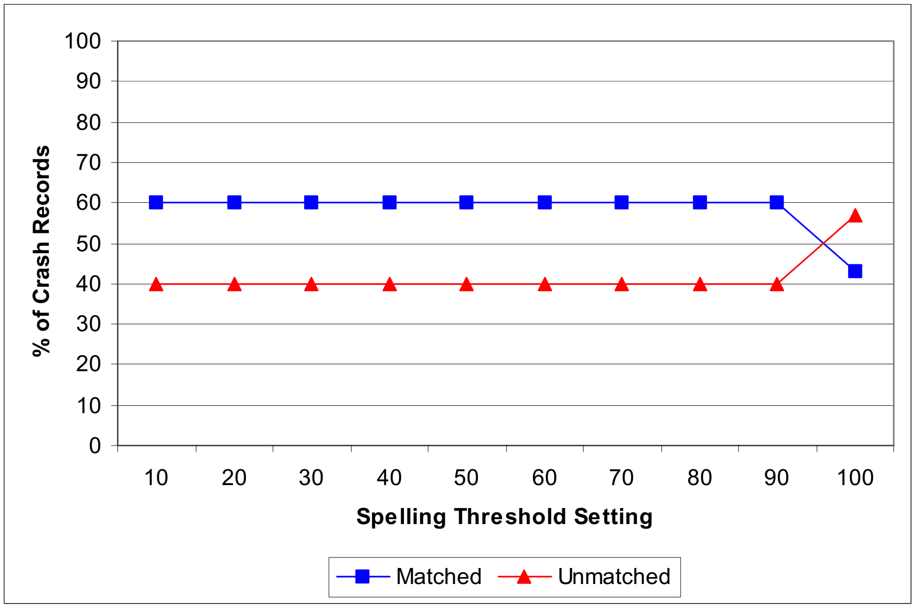


Figure 13. Sensitivity Analysis – Spelling Threshold

Minimum Candidate Threshold

The minimum candidate threshold determines how restrictive the algorithm should be in considering or rejecting a potential candidate. The algorithm rejects potential candidates with a score less than the threshold. Figure 14 summarizes the results of the analysis. Since almost all the possible candidates scored below 60, once the minimum candidate threshold was set to 60, the result was that almost none of the addresses were selected as candidates therefore resulting in very few matches.

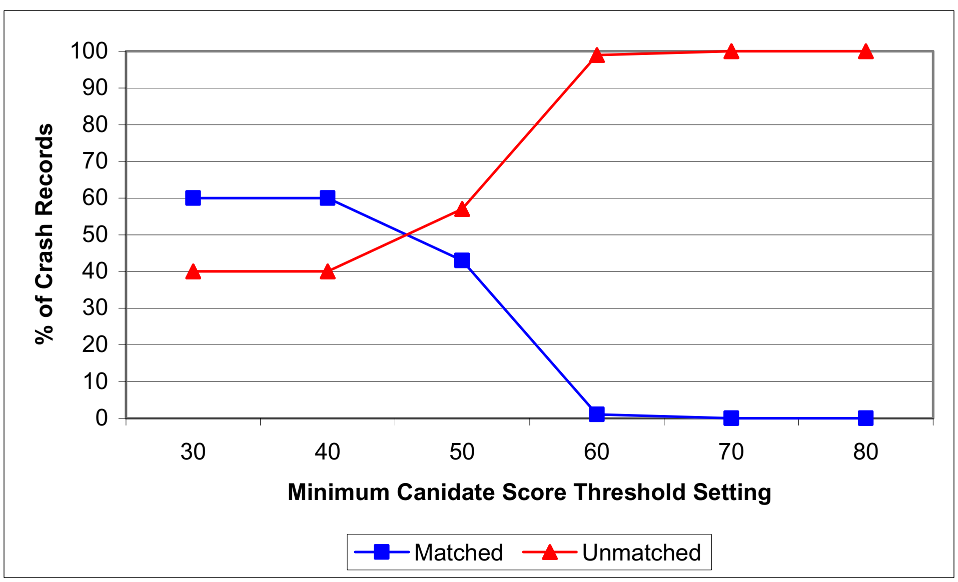


Figure 14. Sensitivity Analysis – Minimum Candidate Threshold

Minimum Match Threshold

The minimum match threshold determines how well addressed a candidate must be to be considered a match. The algorithm rejects potential matches if they yield a score lower than the threshold. Figure 15 summarizes the results of the analysis. As the minimum match threshold increased, those with lower scores began to weed out and only those addresses with scores that were higher than the minimum match threshold were identified as matches.

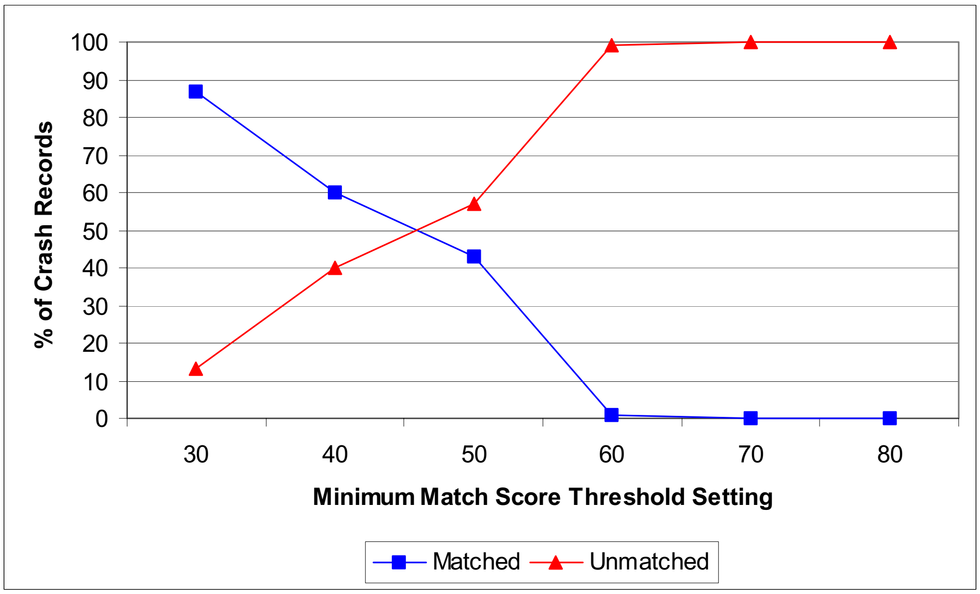


Figure 15. Sensitivity Analysis – Minimum Match Threshold

CRASH DATA ON STATE HIGHWAYS

As with the local street crash data, the researchers used state highway crash data from the TxDPS database. Table 6 shows a sample of records, with CONTSEC1 and MILE1 representing the control section and distance along the control section where the crash occurred. Notice in Table 6 that the accuracy associated with MILE1 measures is 1/10 of a mile (e.g., “36” represents 3.6 miles along control section 1805).

As mentioned previously, TxDOT uses the reference marker-distance model for highway inventory purposes and for maintaining this inventory in a GIS format. To convert data from the control section-mile point model to the reference marker-distance model, TxDOT uses an equivalence lookup table called Mile Point to Reference Marker Equivalence (MPRME). Figure 16 illustrates the conversion process and Table 7 shows a few sample records from the MPRME table.

Table 6. State Highway Source Crash Dataset

ACC_NO	CNTYCITY	COUNTY	THDCNTY	PRIM	MILE1	MILE2	DISTRICT	OTHERFAC	BLOCK	CONTSEC1
9054143	24000	240	240		46		014	38		3709
9054894	24000	240	240		261		014	00		1803
9054895	24000	240	240		93		014	30		3801
9057446	24000	240	240		36		014	00		1805
9057447	24000	240	240		212		014	00		54202
9060370	24000	240	240		4		014	00		3709
9067078	24000	240	240		30		014	00		3709
9070589	24000	240	240		176		014	30		1803
9072893	24000	240	240		74		014	3-		1805

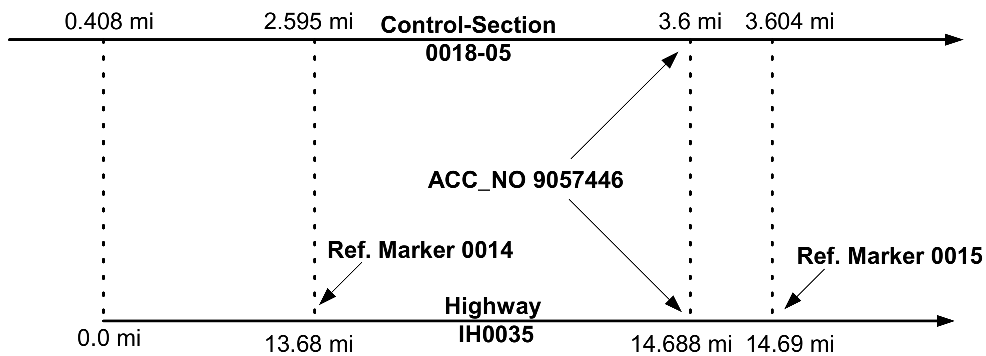


Figure 16. Control Section to Reference Marker Equivalence (Distances in Miles)

Table 7. Sample Records from the MPRME Table.

DI	CNT	CS	BMP	EMP	LEN	HWY	BRMKR	BSIGN	BDISP	ERMKR	ESIGN	EDISP
22	240	1805	0	0.605	0.605	IH0035	0011	+	0.403	0012	+	0
22	240	1805	0.605	1.576	0.971	IH0035	0012	+	0	0013	+	0
22	240	1805	1.576	2.595	1.019	IH0035	0013	+	0	0014	+	0
22	240	1805	2.595	3.604	1.009	IH0035	0014	+	0	0015	+	0
22	240	1805	3.604	4.61	1.006	IH0035	0015	+	0	0016	+	0
22	240	1805	4.61	5.592	0.982	IH0035	0016	+	0	0017	+	0
22	240	1805	5.592	6.621	1.029	IH0035	0017	+	0	0018	+	0
22	240	1805	6.621	7.6	0.979	IH0035	0018	+	0	0019	+	0
22	240	1805	7.6	8.274	0.674	IH0035	0019	+	0	0019	+	0.674

Figure 16 shows a crash occurred at 3.6 miles along control section 1805. The procedure for converting this location to an event located at 14.688 miles along IH-35 between reference markers 14 and 15 follows:

1. Table 7 translates a milepoint along a control-section to a distance relative to the nearest reference marker on the primary highway. For example, crash No. 9057446 occurred 3.6 miles along control section 1805 (Table 6). According to Table 7, the crash occurred between control section milepoints 2.595 and 3.604 on control section 1805, i.e., between reference markers 14 and 15 on IH-35.

2. Table 8 shows the location of a sample of reference markers along IH-35. This table is part of TxDOT's inventory of highway features. Table 8 shows that the location of reference marker 14 is 13.683 along IH-35.
3. Table 9 shows the result of adding the distance between the beginning milepoint and the accident (3.6 – 2.595 = 1.005) to 13.683 to obtain 14.688 miles, which is the location of crash No. 9057446 on IH-35. Table 9 also shows the location associated with other crashes on IH-35.

Table 8. TxDOT Reference Markers

ID	DIST	CNTY	CNTRL	HWY	MKR	DFO
44695	22	240	1805	IH0035	0011	11.088
44696	22	240	1805	IH0035	0012	11.693
44697	22	240	1805	IH0035	0013	12.664
44698	22	240	1805	IH0035	0014	13.683
44699	22	240	1805	IH0035	0015	14.692
44700	22	240	1805	IH0035	0016	15.698
44701	22	240	1805	IH0035	0017	16.68

Table 9. Sample of Crashes on IH-35

ACC_NO	HWY	Measure
9248780	IH0035	12.488
9311641	IH0035	13.388
9228274	IH0035	13.988
9023204	IH0035	14.388
9057446	IH0035	14.688
9034760	IH0035	15.188
9002837	IH0035	17.688
9281739	IH0035	18.288
9072893	IH0035	18.488

The researchers automated this procedure by using the following query (Figure 17 shows the database tables):

```
SELECT ACC_NO, HWY,(DFO + (MILE1 / 10) - BMP) AS Measure

FROM (DPSCrashData1999 INNER JOIN [MPMRE May 2002] ON [DPSCrashData1999].[CONTSEC1] = [MPMRE May
2002].[CS])INNER JOIN [Unique Reference Markers] ON [MPMRE May 2002].[CS] = [Unique Reference
Markers].[CNTRL]

WHERE IsNull(BLOCK) = True And IsNull(PRIM) = True And IsNull(MILE1) = False And IsNull(CONTSEC1)
= False And BMP <= (MILE1/10) And EMP >= (MILE1/10) And MKR = BRMKR And County = 240;
```

Once the crash data were in a format consistent with the TxDOT route highway network files, the researchers used the route event tool in ArcGIS to generate crash point features. Figure 18 shows those features overlaying the state highway network in the Laredo area. In general, as opposed to the crash data on local streets, the researchers did not encounter major difficulties during the crash point feature generation process.

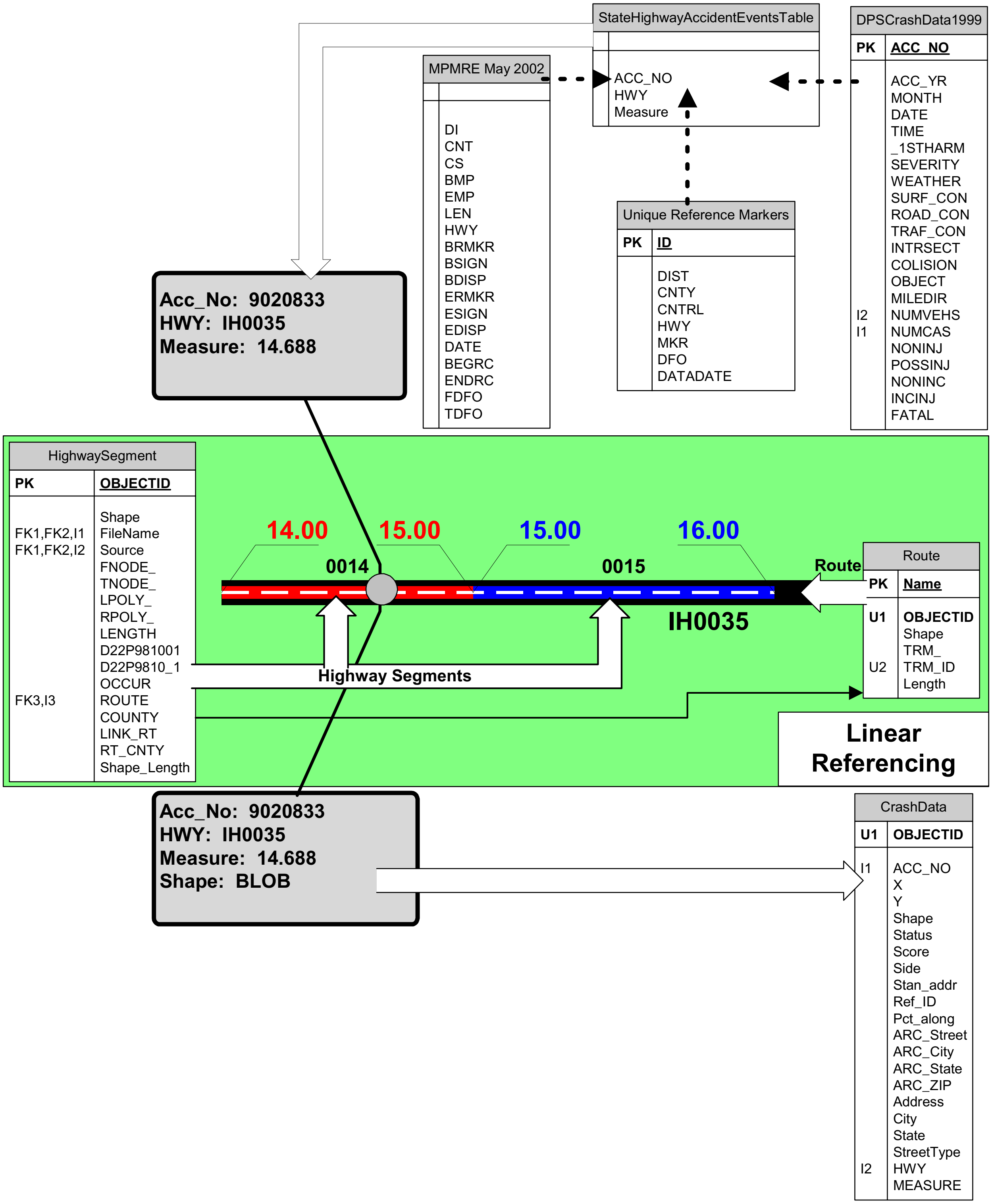


Figure 17. Procedure for Locating Crash Data on State Highways

Figure 18. State Highway Crash Data as GIS Features

DATA INTEGRATION

Since the TxDPS state crash database included both crash data on state highways and local streets, the final step was to integrate the two data sets into one data set. This procedure was straightforward by using the ArcGIS geoprocessing tool because both the state highway and local street data sets had the same database structure. Figure 19 shows the result of the merging process (notice that the state highway crash data and the local street crash data are now both contained in the CrashData layer).

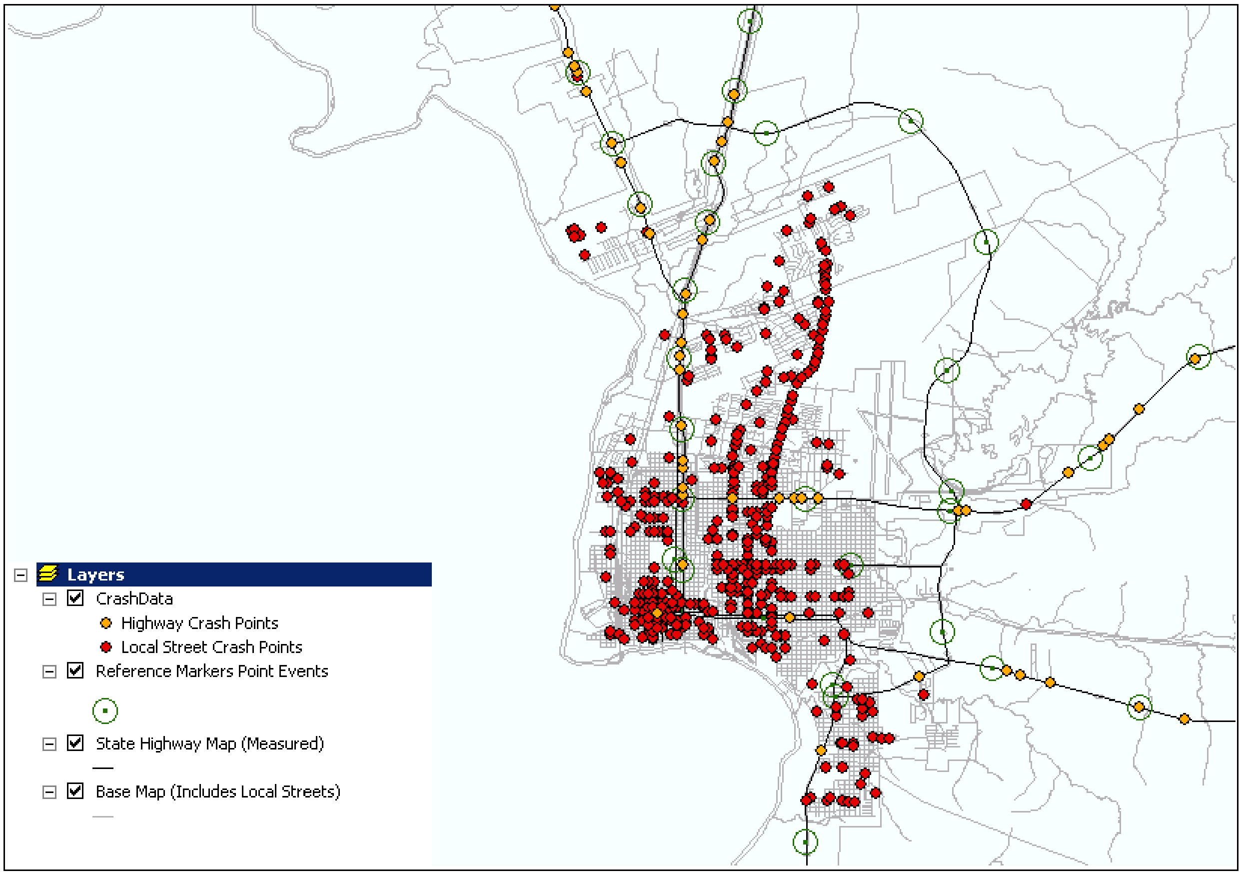


Figure 19. All Crash Data as GIS Features

CHAPTER 4. COLLECTING AND INTEGRATING VOLUME DATA

SOURCES OF VOLUME DATA

A number of traffic volume data sources are available in the Laredo area. For the project, the researchers used the following data sources:

- 2000 ATR Vehicle Classification & Border Trend Report (TxDOT, 2000a): This report contained summarized data for all ATR stations around the state, including average hourly volumes, annual average daily traffic (AADT) data, and vehicle classification data. For identifying the location of individual stations, the report included route names and approximate distances along the routes where the stations were located. In the Laredo District area, the report included data for three ATR stations. One of these stations (S219) is located close to the Laredo urban area. Figure 20 shows sample average hourly volumes and AADT for this station, which is located on IH 35, some 2.3 miles south of US 83.

Annual Average Hourly Volumes
By Days of the Week

HOUR	SUN.	MON.	TUE.	WED.	THR.	FRI.	SAT.
12-AM	290	289	252	265	275	301	411
01-02	216	224	194	211	220	258	357
02-03	193	195	167	176	201	236	330
03-04	162	177	147	163	183	217	284
04-05	157	199	166	184	201	219	283
05-06	174	270	230	248	272	286	325
06-07	228	424	386	384	412	434	427
07-08	309	584	545	534	568	607	569
08-09	413	629	590	604	636	671	720
09-10	524	629	621	633	669	710	818
10-11	625	704	666	693	724	775	884
11-12	707	745	709	712	739	831	902
12-PM	775	775	724	757	765	873	893
01-02	818	774	718	743	773	872	879
02-03	847	785	744	759	780	892	869
03-04	884	806	750	789	798	913	850
04-05	925	838	774	807	818	969	840
05-06	941	842	791	809	853	980	812
06-07	939	785	718	746	785	1,000	762
07-08	875	685	611	633	670	878	698
08-09	763	568	535	558	584	809	609
09-10	630	479	460	477	514	690	516
10-11	495	391	391	405	432	595	433
11-12	378	316	333	330	361	502	350
TOTAL	13,268	13,113	12,222	12,620	13,233	15,518	14,821
% of AADT	97.9	96.8	90.2	93.2	97.7	114.6	109.4

Annual Average Week Total -	94,795
AADT -	13,546

Figure 20. Sample Year 2000 Average Traffic Volume for TxDOT ATR Station S219 (adapted from TxDOT, 2000a)

- District Highway Traffic Map of Laredo District (TxDOT, 2000b): This map contained year 2000 AADT data on state highways in the TxDOT Laredo District (Figure 21). The researchers received a copy of the map in tag image file format (TIFF)—the original map format was Microstation .dgn. As Figure 21 shows, the map was actually a sketch showing approximate route alignments and count station locations.

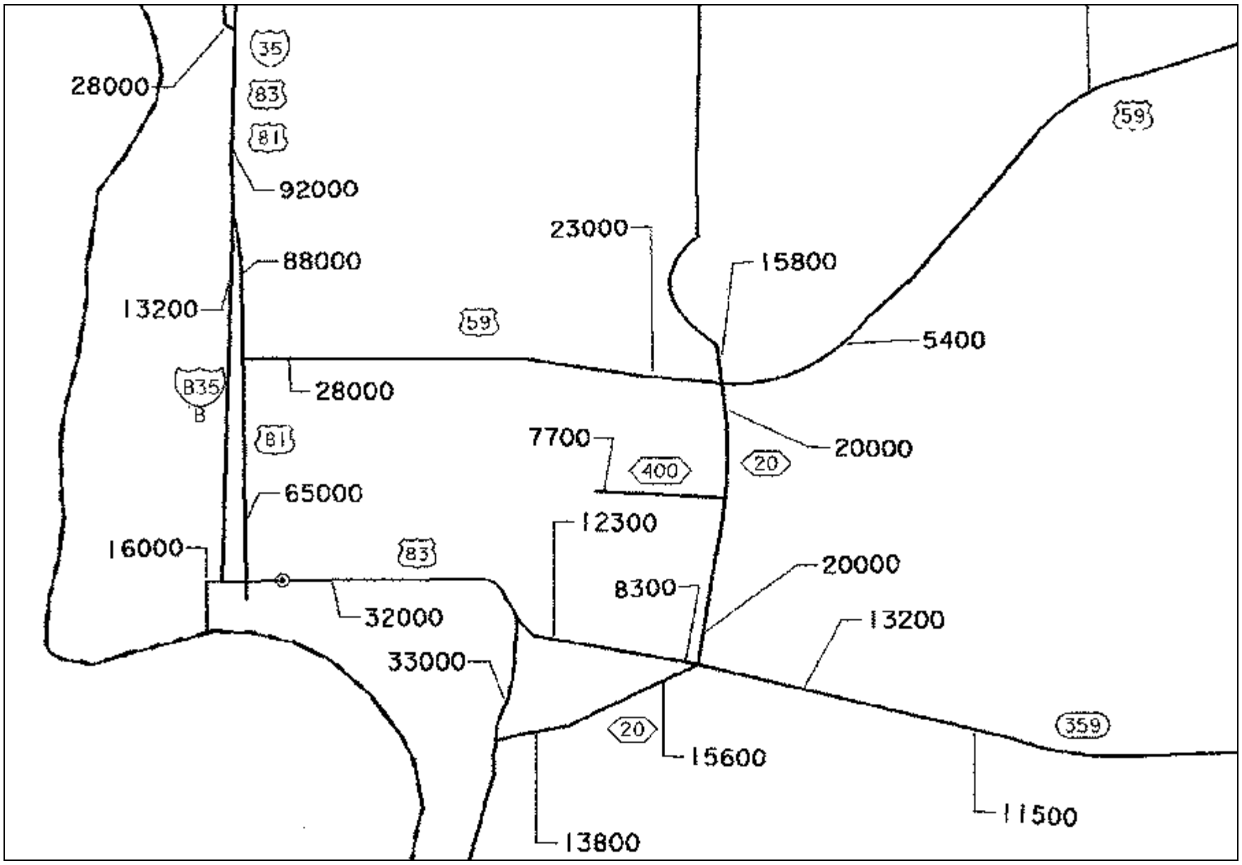


Figure 21. Year 2000 AADT on State Highways in Laredo (adapted from TxDOT, 2000a)

- Urban Study Map of Laredo Webb County Texas (TxDOT, 1998): This map contained ADT data for the local (jurisdiction) street network in Webb County (Figure 22). TxDOT provided this map both on paper and TIFF formats. The map was not georeferenced, which prevented its automatic conversion to a GIS environment. It may be worth noting that, as opposed to the AADT data on the state highway network, which TxDOT collects every year, local jurisdiction street ADT data are collected much less frequently. The last year for which TxDOT had ADT collected on local streets was 1998. Later this year, TxDOT is planning a new data collection effort.

The data represented in these data sources were in different formats, which implied a different approach for bringing the associated data into the GIS.

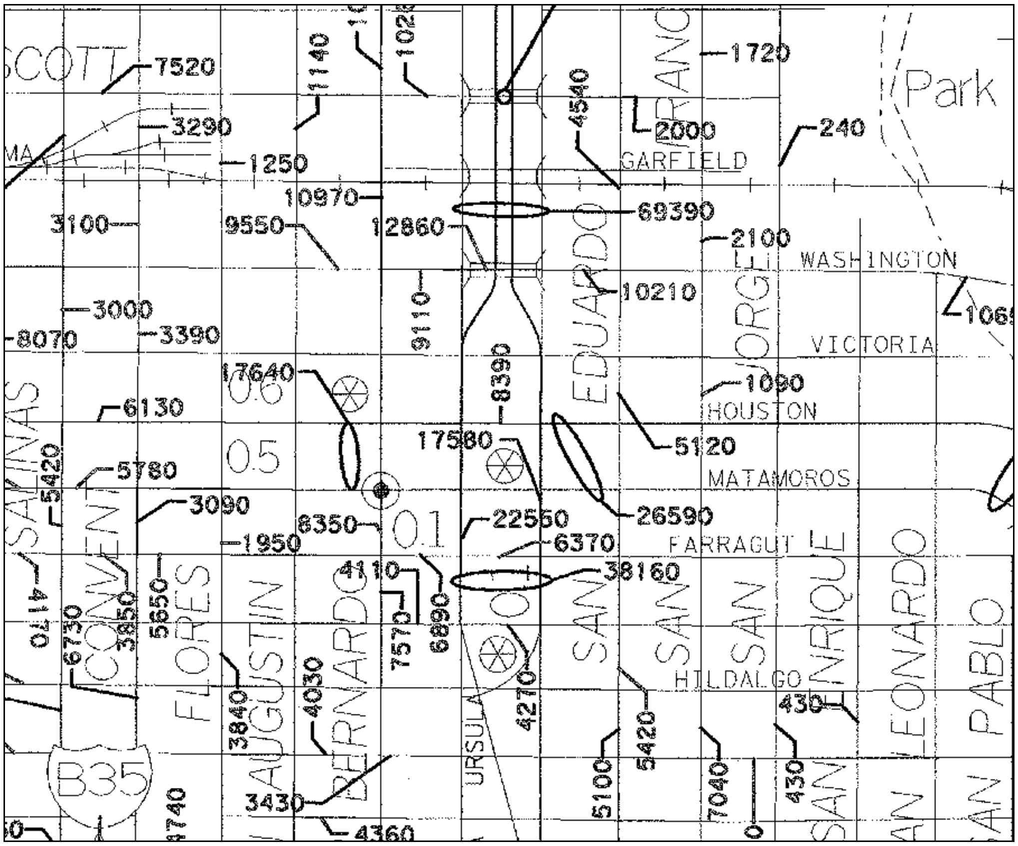


Figure 22. Year 2000 ADT Data on Local (Jurisdiction) Streets in Laredo (TxDOT, 1998)

GIS INTEGRATION

ATR Vehicle Classification & Border Trend Report

The 2000 ATR Vehicle Classification & Border Trend Report reported AADTs, annual average hourly volumes, high hours, average daily traffic volumes by month, day of week, and season. Of the three ATR stations in the Laredo District, only one station (S219) was close to the Laredo urban area. The location associated with this station ("IH 35, 2.3 MI S OF US 83, LAREDO") was only descriptive and did not use standard state highway route codes or GIS-based measures. As a result, it was not possible to use an automated procedure to locate the station in the GIS, forcing the researchers to manually locate the volume data station using the approximate location information and cross-referencing it with the state highway map. After finding an approximate location for the volume data station, the researchers snapped the volume data station point feature to the IH-35 route.

ADT Traffic Maps

The ADT traffic maps were also not georeferenced. As a result, it was necessary to follow a manual procedure to generate GIS features for the volume stations. First, the researchers located volume data points on the digital TxDOT road base map. To help with organization and reduce the chance of duplicate stations, the researchers partitioned the map into areas of concentration. Once approximate locations for the volume data stations were found, the researchers snapped volume data station point features to their associated traffic segments. Although ArcGIS's support for database relational constraints was not strong, using Access as the backbone database the researchers were able to enforce database integrity. By creating relational constraints in Access between the VolumeDataStation geodatabase table and the remainder of the database schema (Figure 23), the researchers made the GIS plotting of the ADT counts consistent with the rest of the database.

Volume Data Station Table

After creating traffic volume data station point features in the GIS, the researchers integrated the metadata associated with those features into the database. The fields processed in the VolumeDataStation table are as follows:

- OBJECTID: This field stores a unique identifier (automatically generated).
- Shape: This field stores the GIS object (automatically generated).
- VolumeDataStation: This field stores a unique identifier for the station (automatically generated).
- Latitude: This field stores the latitude value in decimal degrees associated with the volume data station (derived using a GIS script).
- Longitude: This field stores the longitude value in decimal degrees associated with the volume data station (derived using a GIS script).
- StatusType: This field describes the type of data that are collected by the volume data station. Acceptable values are occasional, one time, or permanent.
- StationType: This field describes the type of data that is collected by that station. Acceptable values are average daily traffic (adt) or turning movement counter (tmc), which are located in the StationType lookup table.
- Route: This field stores the route associated with the volume data station. In the case of station S219, the route was provided. In general, however, this was not the case. The researchers had to query the route map to identify the route with which the volume data station was associated.
- Mile: This field contains the measure in miles along the route where the volume data station is located.
- Name: This field contains the "official" volume station name, for example S219. This data element was not available in the case of the count stations on the local jurisdiction street network.
- LocationDescription: This filed contains a verbal description of where the volume data station is located.
- StreetName: This field contains the name of the street where the volume data station is located. This data element was not readily available and, as a result, it was necessary to manually query the base map.

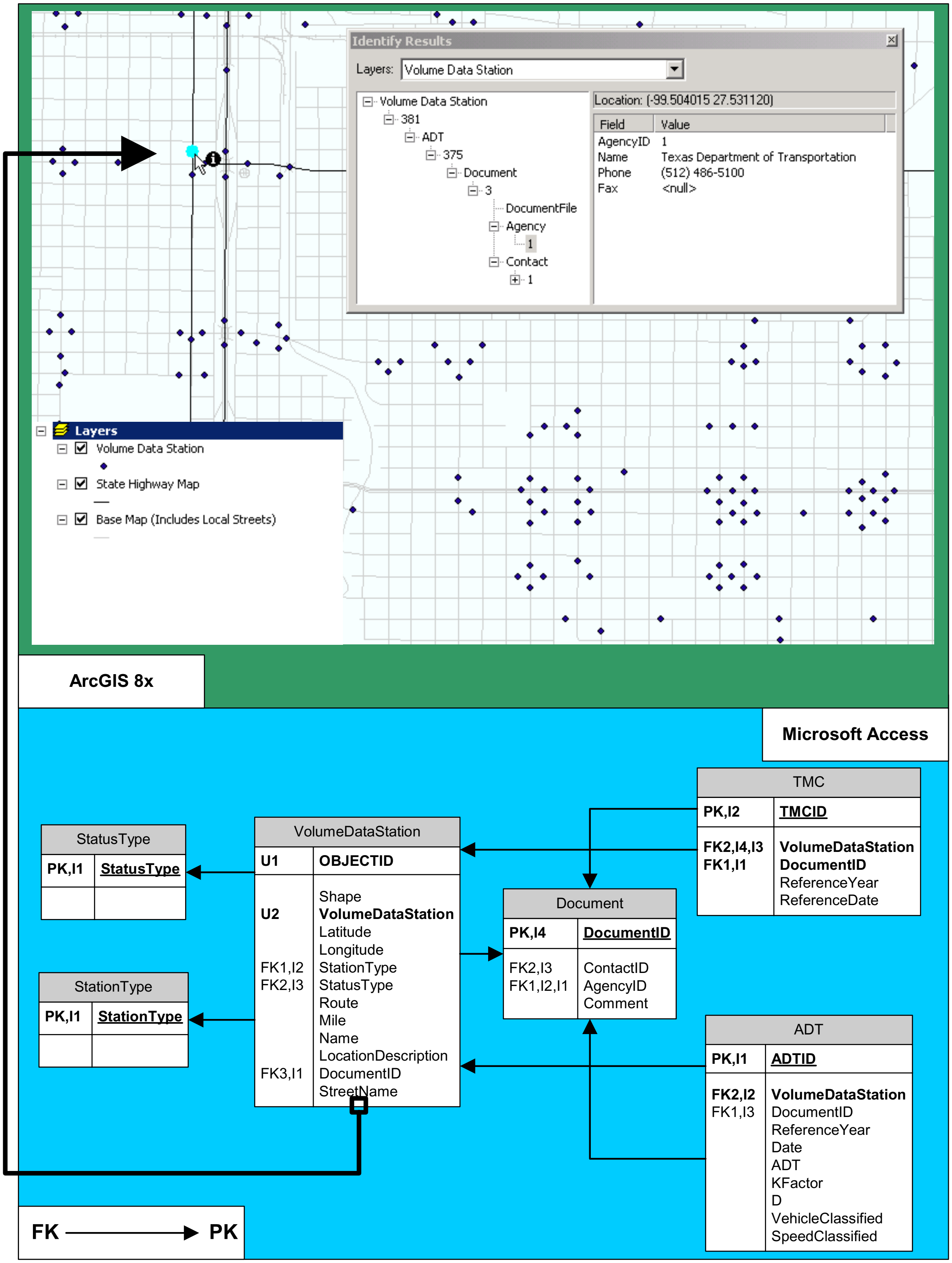


Figure 23. Volume Data Database Schema and GIS Interaction

ADT Table

After generating attributes in the volume data station table, the researchers populated the following ADT table elements:

- ADTID: This field stores a unique identifier (automatically generated).
- ReferenceYear: This field contains the reference year associated with ADT data.
- ADT: This field contains the ADT value (normally a call out from the data point on the original map).

Supporting Documents

The last step was to integrate supporting document data. The researchers accomplished this by entering document metadata in the database and by storing an electronic copy of the document in the document file table.

RESULTS

The final product was a GIS-based map containing the ADT values associated with the data provided by TxDOT along with the state highway and local jurisdiction street networks in the Laredo area (Figure 24). All ADT data are stored in the backbone Access database and displayed using ArcGIS. When queried through the map interface, GIS features display all related data associated with the features, including attribute data as well as metadata and a copy of the original document.

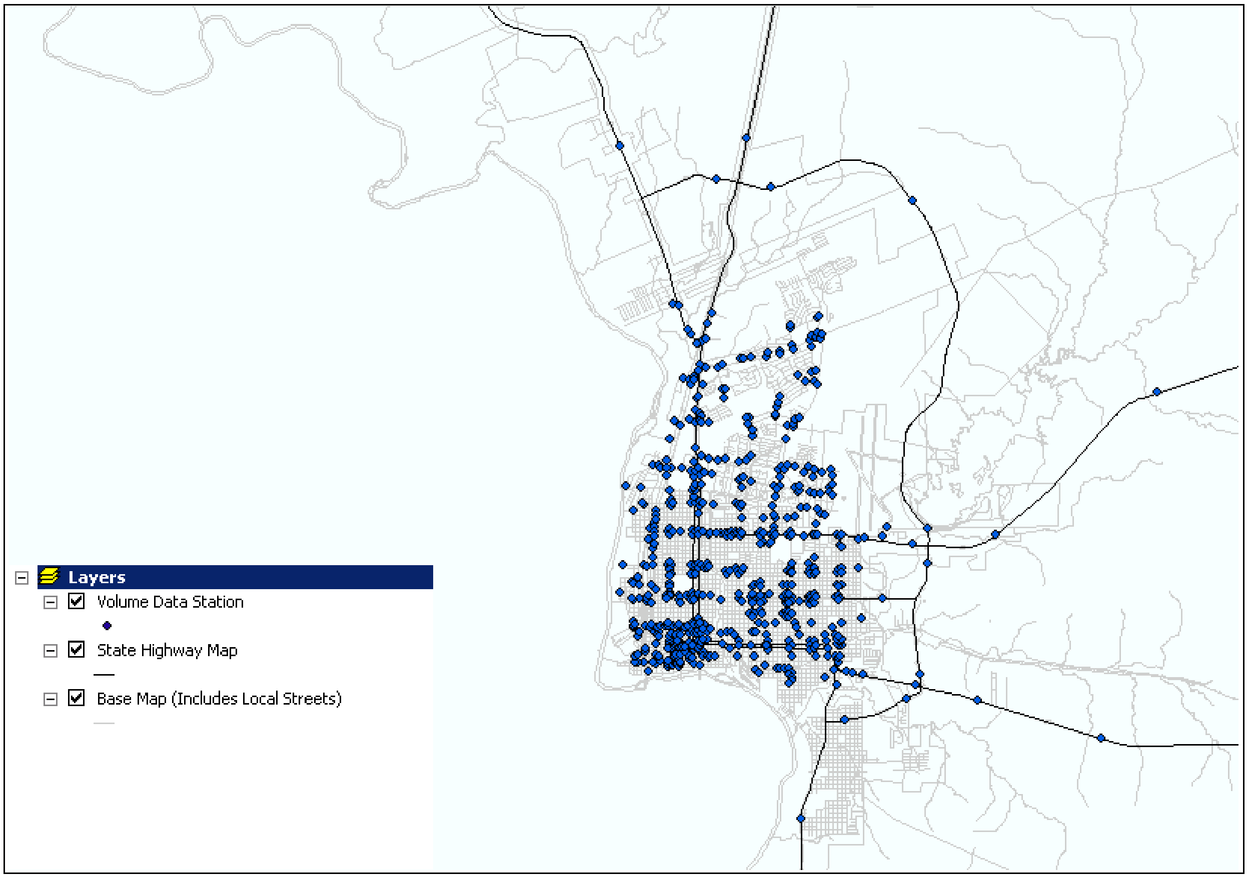


Figure 24. Volume Data Stations in a GIS Format

CHAPTER 5. DATA ON THE MEXICAN SIDE OF THE BORDER

Researchers at the Instituto Mexicano del Transporte (IMT) set up a parallel project to document the availability of data on the Mexican side of the border (Acha and Castrellón, 2003). This chapter summarizes the findings by Acha and Castrellón.

GENERAL INFORMATION ABOUT NUEVO LAREDO

Nuevo Laredo is located on the northwest corner of the Mexican State of Tamaulipas (Figure 1). Like Laredo, Nuevo Laredo has experienced accelerated growth in population and economic development over the last few years, particularly after the introduction of the North American Free Trade Agreement (NAFTA). In 2000, the population in Nuevo Laredo was 311,000 (compared to 193,000 in Laredo, according to the U.S. Census Bureau (2001)).

Manufacturing, commerce, and transportation are the main economic activities in the Tamaulipas and Nuevo Laredo areas. Although not as prevalent as in other urban areas along the Texas-Mexico border, maquiladoras are an important component of the manufacturing economic sector in Nuevo Laredo. International commerce and transportation, particularly in the form of warehousing facilities and cross border traffic between Laredo and Nuevo Laredo, are the main economic activities in Nuevo Laredo. More than 36% of all international commerce in Mexico takes place in Nuevo Laredo.

Figure 25 shows the main routes connecting Nuevo Laredo with other cities in Mexico and the U.S. Table 10 shows year 2001 AADTs on those routes. The most important route is MX 85, which connects Nuevo Laredo with Monterrey and Mexico City. Beginning some 45 mi (75 km) south of Nuevo Laredo, the route is divided into two highways: the old route 85—also known as the 'Free' Mexico – Nuevo Laredo Road—and toll road 85D—also known as the Monterrey-Nuevo Laredo Highway). The border highway (MX 02, also called "Ribereña") connects Nuevo Laredo with Matamoros, Reynosa, Piedras Negras, and Ciudad Acuña (Coahuila). State highway NL 01 connects Nuevo Laredo with Anáhuac (Nuevo León).

Table 10. Year 2001 AADTs on Main Routes Leading to Nuevo Laredo (adapted from Acha and Castrellón, 2003)

Highway	AADT	Percent Buses	Percent Trucks
MX 85D (toll)	15,678	1.6%	12%
MX 85 (no toll)	8,933	3.7%	11%
MX 02 (Nuevo Laredo – Piedras Negras)	3,850	3.3%	9.1%
MX 02 (Nuevo Laredo – Reynosa)	2,986	5.4%	13%
NL 01	4,078	4.5%	14%

DATA IN GIS FORMAT

IMT coordinates the development of SIGET (Sistema de Información Geoestadística para el Transporte—Geostatistical Information System for Transportation). SIGET includes cartographic data such as political divisions, state and federal highways, cities, and toll plazas.

Secretaría de Comunicaciones y Transportes (SCT) Centers nationwide are responsible for maintaining and updating SIGET data.

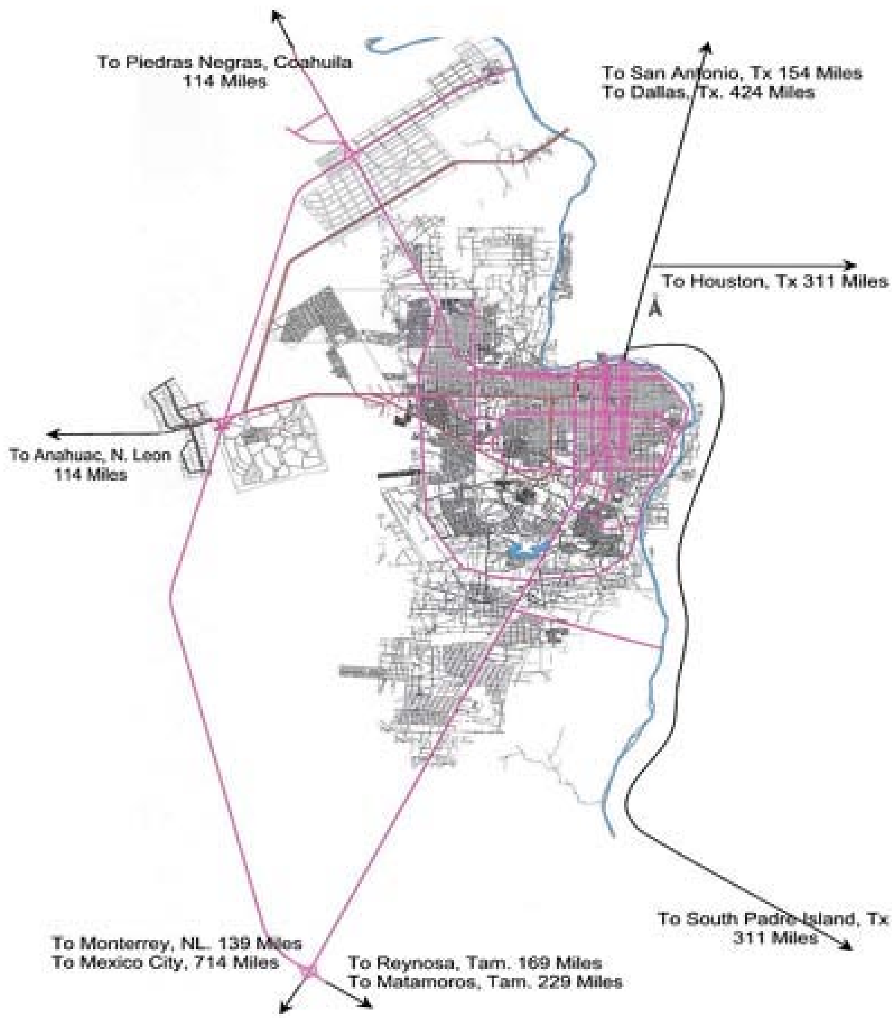


Figure 25. Main Highways in the Nuevo Laredo Area (Nuevo Laredo, 2003)

Figure 26 shows some of the cartographic data in the Nuevo Laredo area imported from SIGET. It may be worth noting that, for dynamic segmentation and linear referencing purposes, the Dirección General de Servicios Técnicos (DGST) divides roads in Mexico into routes, roads, sections, and segments. In general, segments are 500 m (1,640 ft) long.

Traffic Volume Data

DGST publishes an annual summary based on the results of a counting effort at different locations on the federal road network as well as permanent counting stations at toll plazas. The summary is linearly referenced and weighted by distance, i.e., it represents an approximation, by section, of AADTs at different locations on the network. Figure 27 shows AADT ranges at the section level in the Nuevo Laredo area.

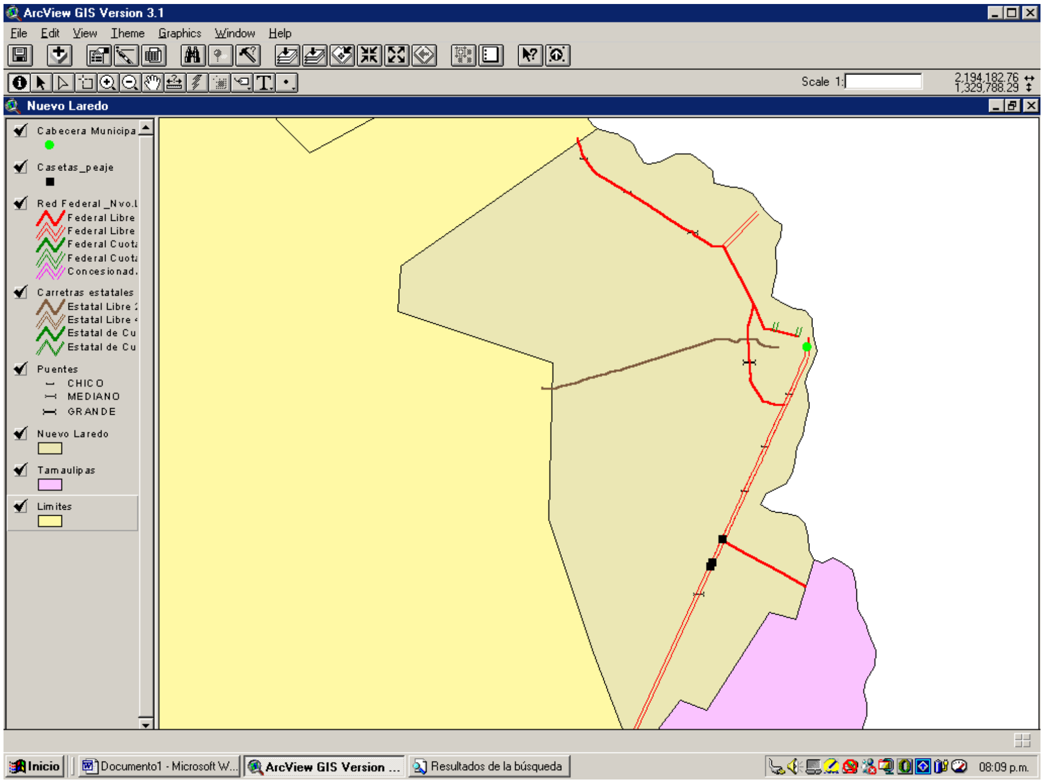


Figure 26. Cartographic Data Imported from SIGET (after Acha and Castrellón, 2003)

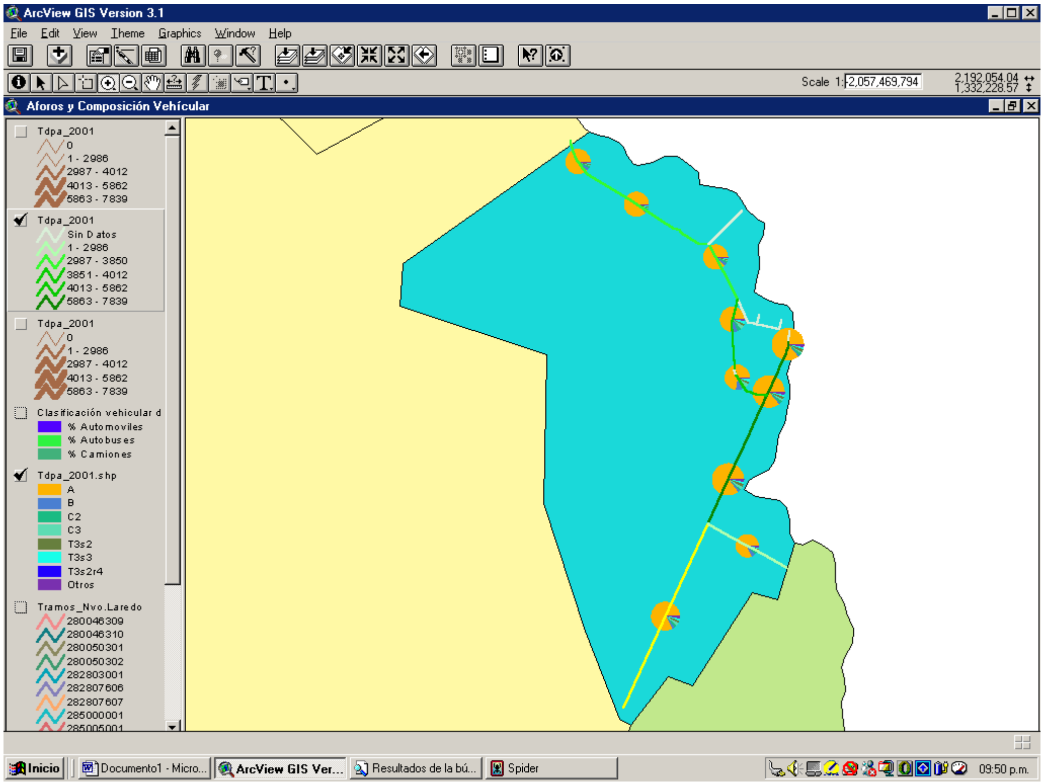


Figure 27. AADTs at the Section Level (after Acha and Castrellón, 2003)

Crash Data

It was only possible to obtain crash data on federal roads. Crashes on federal roads in Mexico, as well as the population and maintenance of the corresponding database, are the responsibility of PFP (Policía Federal Preventiva—Federal Prevention Police). The PFP database provides basic descriptive data about vehicle crashes, including date, time, location, cause, and type. For the project, the researchers conducted a data quality control on the data, eliminated records for which the section information was not correct, and added fields to identify the specific segment associated with a crash. For 2001, the researchers ended up with 48 crashes on federal roads in the Nuevo Laredo area. Figure 28 shows a summarized view of year 2001 crashes on federal roads in the Nuevo Laredo area, where the crashes have been linearly referenced by segment.

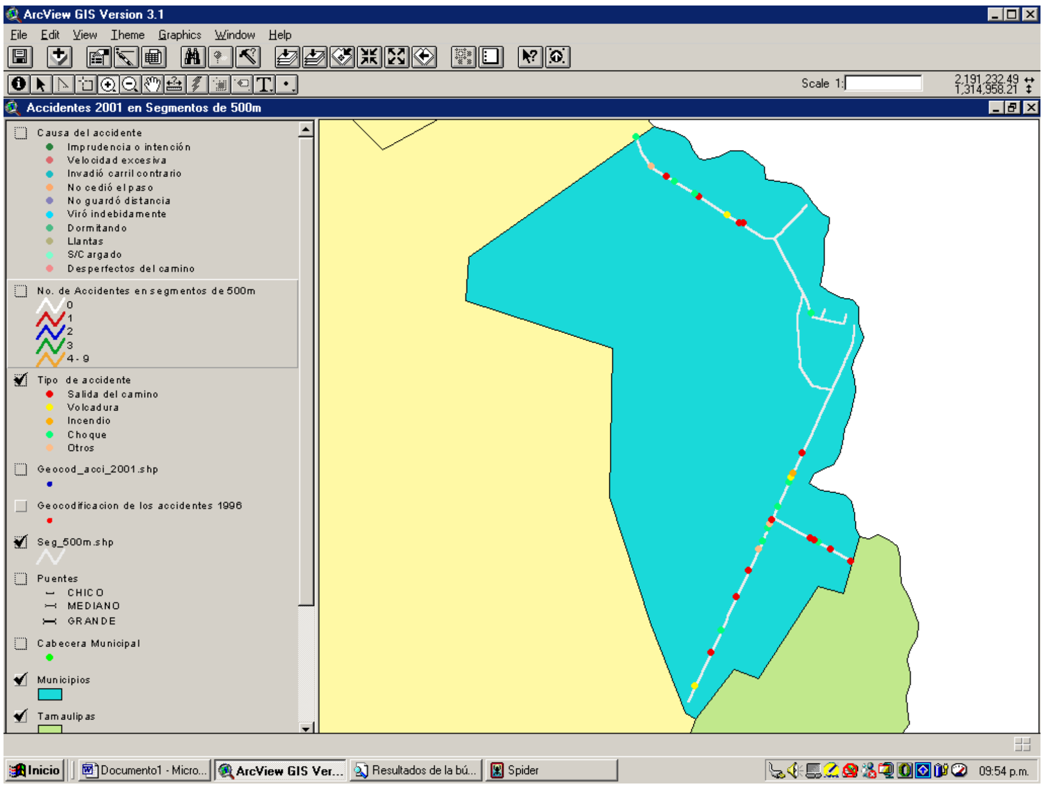


Figure 28. Summarized View of Year 2001 Crashes on Federal Roads in Nuevo Laredo (after Acha and Castrellón, 2003)

Additional Data

In addition to traffic volume and crash data, the researchers also collected aerial photography in digital format and a local street and block map.

CHAPTER 6. DATA INTEGRATION

DATA INTEGRATION ISSUES

After converting each data source to a GIS format, the final step was integrating the data to make sure all layers aligned properly. This process also involved integrating the data into a single geodatabase. Two integration issues were of special interest: coordinate system integration, and base map discrepancies.

Coordinate System Integration

Different data sources had different coordinate systems associated with them (Table 11). ArcGIS has a “projection on the fly” capability that enables data sources that are in different coordinate systems to be automatically projected to a common coordinate system and overlayed on the screen without user intervention. Unfortunately, the “projection on the fly” tool does not always produce consistent or reliable results. To ensure a better display control, the researchers used the projection tool to physically change the coordinate system of each data source to a common coordinate system. As Table 11 shows, the researchers used the coordinate system associated with the aerial photography as the common coordinate system for display purposes.

It may be worth noting that some manual shifting was still necessary in the case of the Lambert Conformal Conic data to properly align the data with other data sources on the screen. In the case of the GIS data from Mexico, it was not necessary to reproject the aerial photography because its associated coordinate system was the same as that associated with the aerial photography on the U.S. side. However, it was necessary to change the coordinate system for the local block and street map. In the case of the federal road network as well as the traffic volume and crash data that had the same coordinate system as the federal road network, the original data source did not include coordinate system data. As a result, it was necessary to manually adjust the data.

After converting all data to Transverse Mercator, the researchers used the import utility tool to import shape files and geodatabase features into a common Access database (Figure 29). As part of the process, the researchers recreated relationships and constraints and confirmed the integrity of the database. The result was a system that displayed all data sources on the same map. Figure 30 provides a view of that map.

Base Map Discrepancies

Different data sources had different road base maps associated with them. The road base maps themselves had different origins and development procedures at their respective agencies, which meant that the road base maps did not necessarily match, even after completing a coordinate system transformation. As an illustration, Figure 31 shows differences in alignment between ESRI’s StreetMap USA and TxDOT’s urban files. Notice the lack of consistency in the differences, which suggests that a simple map translation and/or rotation transformation would not solve the problem. In some cases, there were road segments that appeared to be missing on one of the maps. There were also segments that were longer, shorter, or had a different alignment than their counterparts on the other map.

Table 11. Data Source Coordinate System Data

Data Source	Cordinate System	Parameters
TxDOT urban files TxDOT state highway segments Volume data Crash data on state highways	Lambert Conformal Conic	False Easting: 2999994.000000 False Northing: 2999994.000000 Central Meridian: -100.000000 Standard Parallel 1: 27.416667 Standard Parallel 2: 34.916667 Latitude Of Origin: 31.166667 GCS North American 1927 Datum: North American 1927 Prime Meridian: 0
StreetMap USA Crash data on local jurisdiction streets	USA Contiguous Albers Equal Area Conic	False Easting: 0.000000 False Northing: 0.000000 Central Meridian: -96.000000 Standard Parallel 1: 29.500000 Standard Parallel 2: 45.500000 Latitude Of Origin: 37.500000 GCS North American 1983
Aerial photography	Transverse Mercator Note: Used as the common coordinate system for display purposes.	False Easting: 500000.000000 False Northing: 0.000000 Central Meridian: -99.000000 Scale Factor: 0.999600 Latitude Of Origin: 0.000000 GCS North American 1983 Datum: North American 1983 Prime Meridian: 0
Federal road network, traffic volume data, and crash data from Mexico	Unknown	Unknown
Local block and street map from Mexico	Lambert Conformal Conic	False Easting: 2000000.000000 False Northing: 0.000000 Central Meridian: -102.000000 Standard Parallel 1: 17.500000 Standard Parallel 2: 29.500000 Latitude Of Origin: 15.500000 GCS Clarke 1866

Name	Type
ADT	Personal Geodatabase Table
Agency	Personal Geodatabase Table
Contact	Personal Geodatabase Table
CrashData	Personal Geodatabase Feature Class
Document	Personal Geodatabase Table
DocumentFile	Personal Geodatabase Table
HighwaySegment	Personal Geodatabase Feature Class
Intersection	Personal Geodatabase Table
IntersectionApproach	Personal Geodatabase Table
IntersectionControlType	Personal Geodatabase Table
Route	Personal Geodatabase Table
StationType	Personal Geodatabase Table
StatusType	Personal Geodatabase Table
StreetMapUSA	Personal Geodatabase Feature Class
TMC	Personal Geodatabase Feature Class
VolumeDataStation	Personal Geodatabase Feature Class

Figure 29. Geodatabase Integration Result

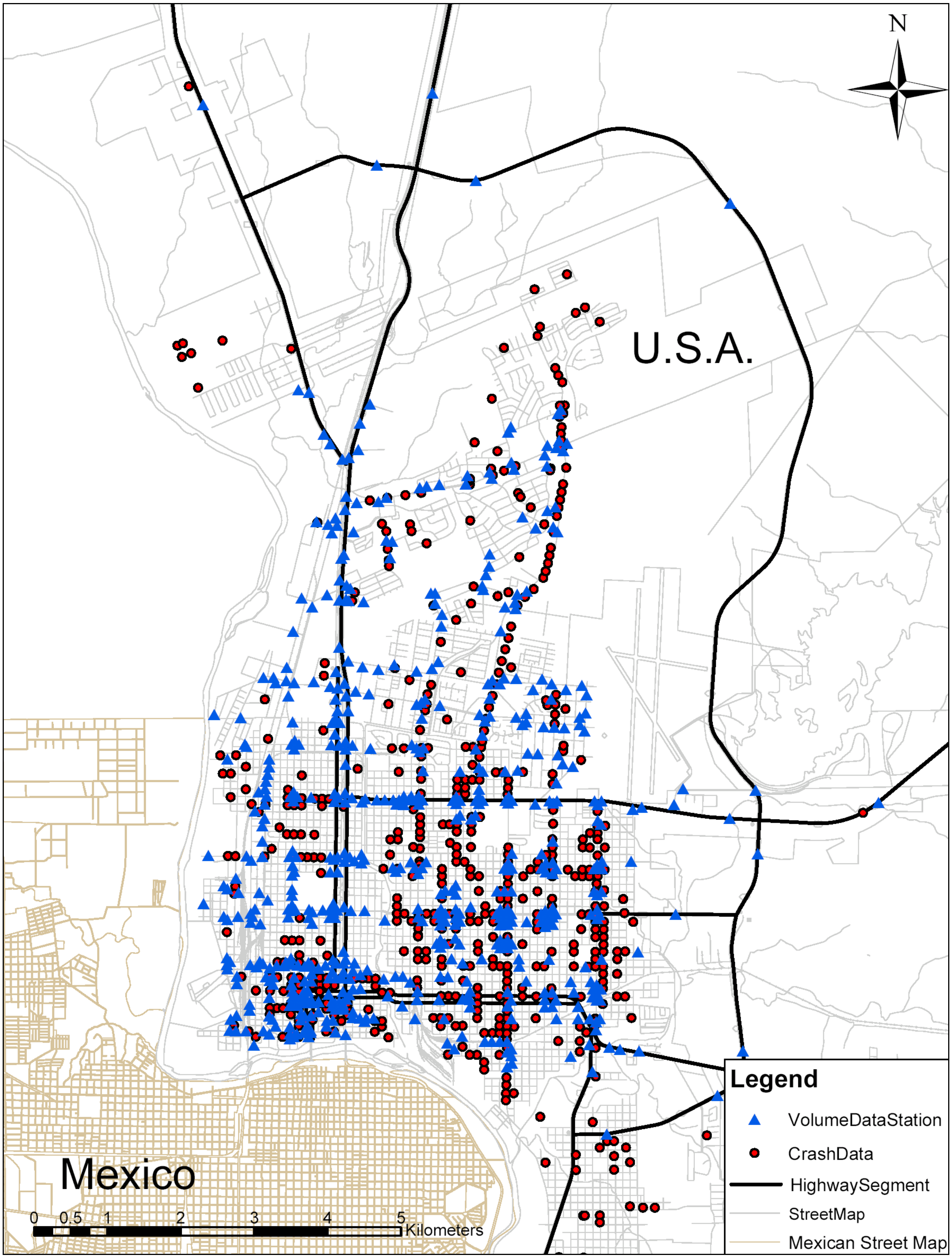


Figure 30. Final GIS Map

One possible solution could be to always use the same road base map for all data sources. Unfortunately, this is not feasible at present because of the dependency between some data sources and their underlying road base maps. More specifically, in the case of crash data on state highways, the positional data associated with each crash data point include a route and a cumulative distance (or measure) along the route. Without this piece of information, it would be practically impossible to locate the crash and, therefore, generate a corresponding feature in the GIS. New crash data collection and data management systems have the capability to remove that limitation by providing the capability to also record latitude-longitude data to identify crash locations in the field. With this capability, having route and measure data to generate GIS features would no longer be necessary, removing therefore the need to depend on the state road base map.

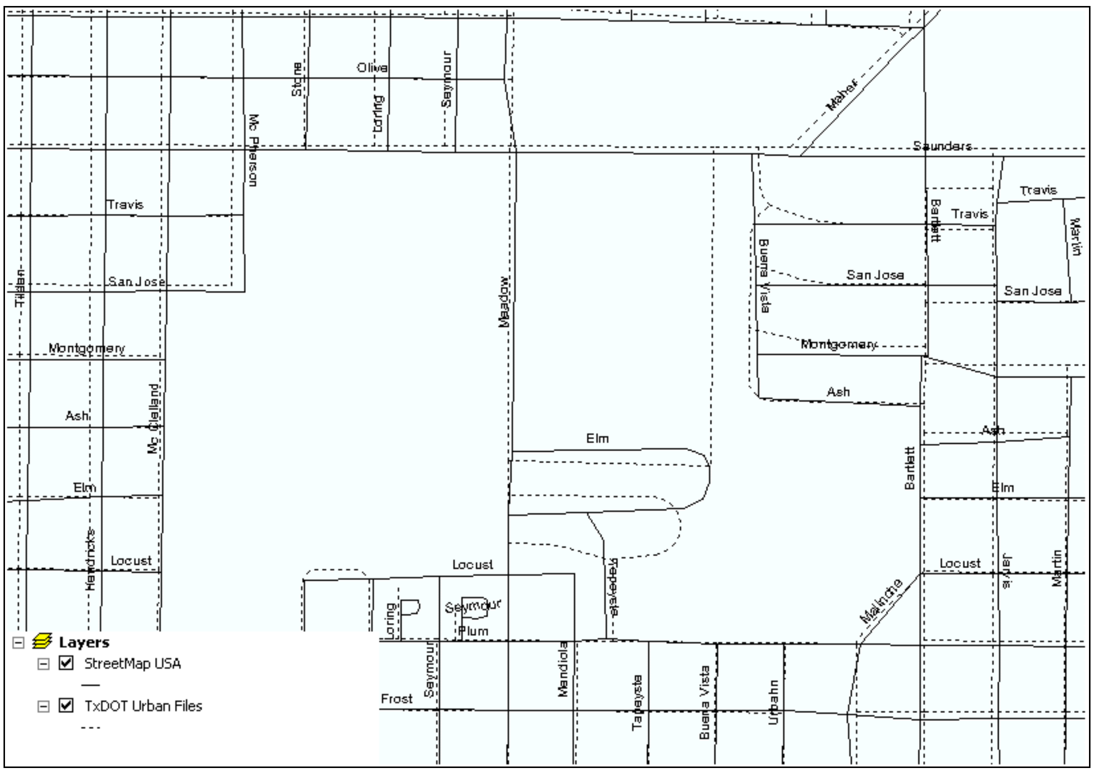
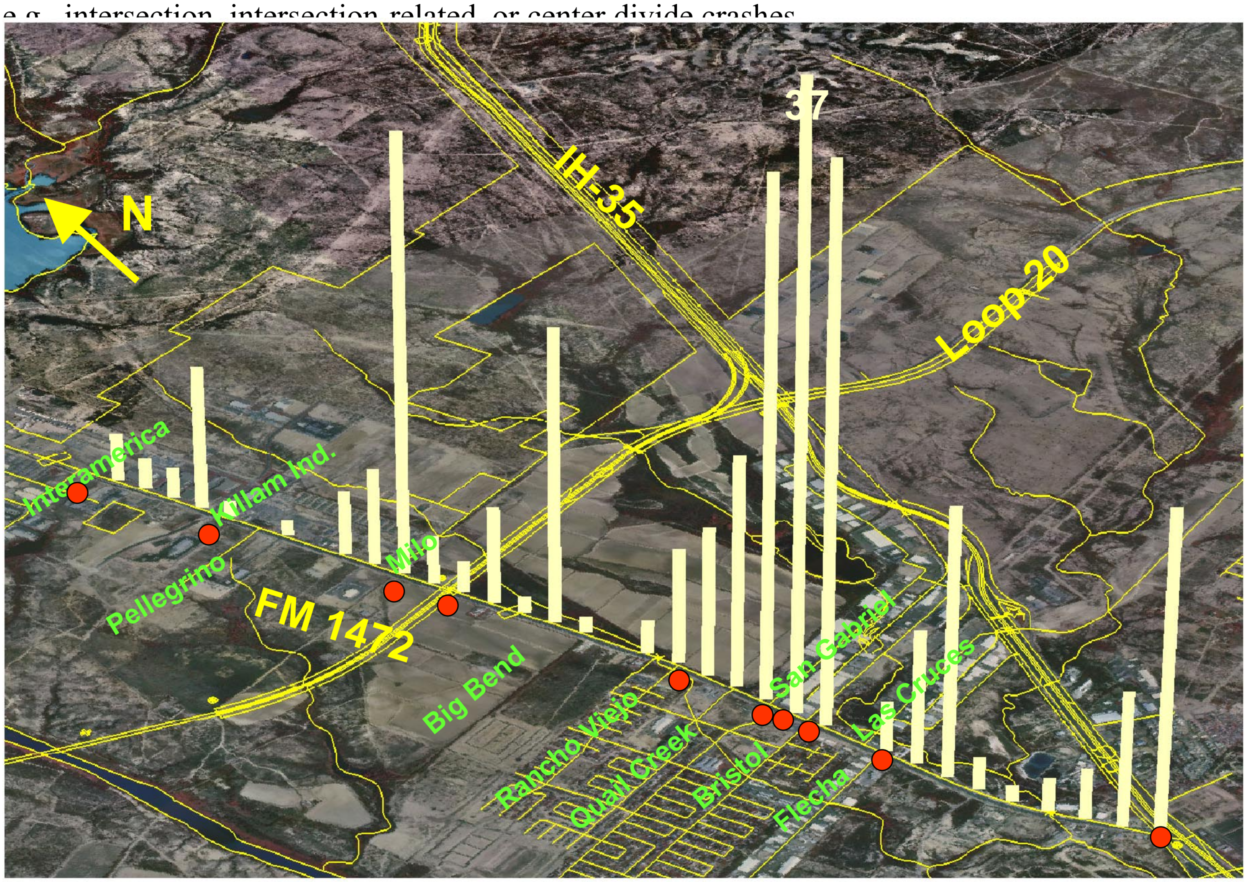


Figure 31. Discrepancies between ESRI's StreetMap USA and TxDOT's Urban Files

APPLICATIONS

Once the data are in a GIS format, a number of applications are possible, including crash rate analyses and evaluation of AADT spatial and temporal trends. The research focused more on system architecture issues and design rather than applications. However, the report does include a couple of application examples to illustrate some of the capabilities of the approach discussed here.

Figure 32 shows the spatial distribution of crashes along FM 1472 from IH-35 to Interamerica Blvd in northwest Laredo from 1997 to 2000. In total, there were 281 crashes during this time period, for an average of about 70 crashes per year. Because the crash data were georeferenced (and in a relational database format), it was straightforward to filter out all crash data that were not located on the specific corridor of interest. Further, it was possible to produce an aggregated view of crash data at regular space intervals (bars in Figure 32 were spaced at 1/10 mile intervals) and overlay the results as a layer in the GIS. Using GIS visualization tools, it was also possible to produce a 3-D view of the crash data on the corridor, which facilitated the interpretation and analysis of the data, as well as its presentation to the public. For example, Figure 32 shows a section between Rancho Viejo and Flecha that exhibited an unusually high density of crashes from 1997 to 2000. When used in combination with available traffic volume data, it was possible to determine that this section of FM 1472 had a crash injury rate (measured as number of people injured or killed in crashes per 100 million of vehicles miles traveled) that was more than twice the average for other urban areas in Texas. Similar maps to Figure 32 could be prepared very easily to show additional or more specific crash-related spatial distributions, e.g., intersection, intersection related, or center divide crashes.

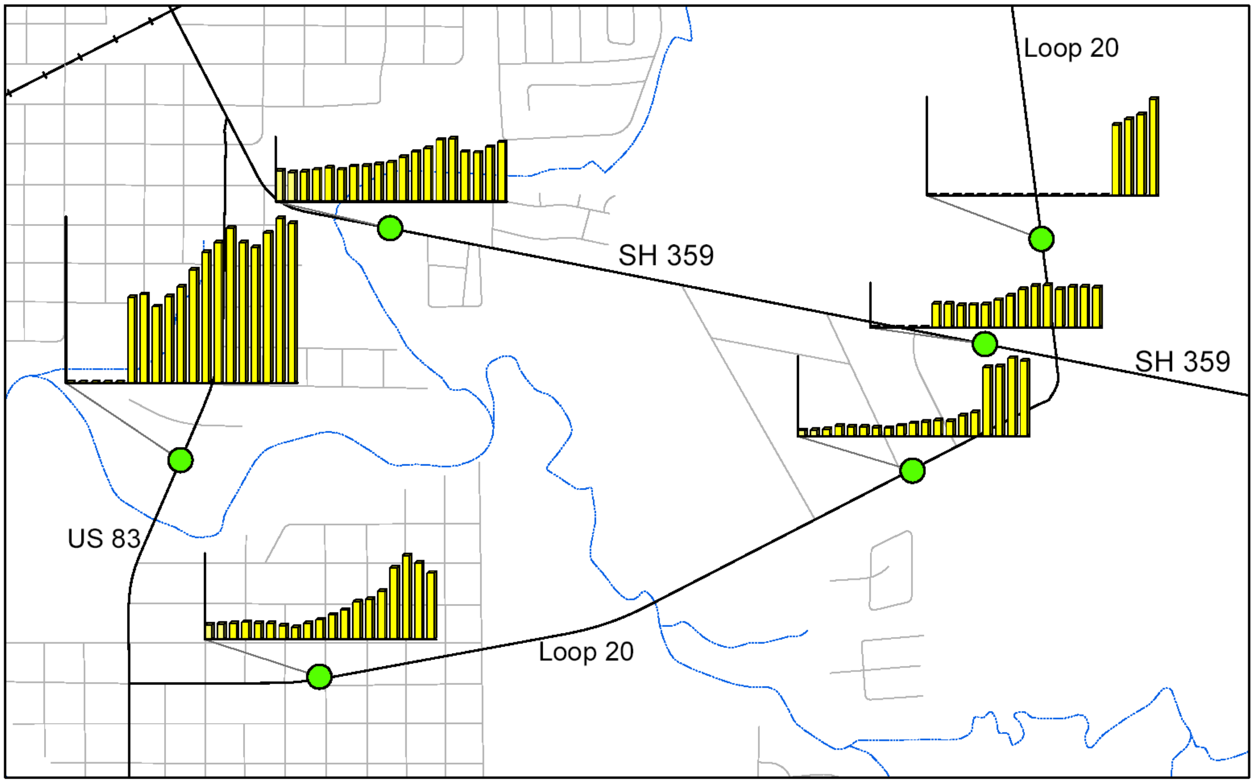


Note: Map prepared by TTI for the TxDOT Laredo District in May 2003. Highway network map is currently available urban file at TNRIS (2003), which does not show some new infrastructure and the World Trade Bridge. Aerial photography is from the 1994 – 1995 time period.

Figure 32. Distribution of 1997-2000 Crashes on FM 1472 from IH-35 to Interamerica Blvd in Northwest Laredo

Figure 33 shows variations in AADT at selected locations on the state highway network in Laredo from 1982 to 2000. In Figure 33, each dot represents the location associated with a TxDOT counter station and each bar chart shows the history of AADT values from 1982 to 2000 at that location. The history of AADT values was stored in a table in the database that was related to the basic table that documents counter station locations. As with the crash data in

Figure 32, a map such as that shown in Figure 33 facilitates both analysis and presentation of the results to interested parties. Similar maps could be produced to show, for instance, variations in AADT along specific corridors, truck percentages, or annual average daily traffic volumes for trucks (AADTTs).



Note: Bar chart shows variations in AADT from 1982 to 2000. Maximum AADT shown is 34,000 (at station on US 83)

Figure 33. Variations in AADT at Selected Locations in Laredo from 1982 to 2000

Another potential application is related to the ability of engineers and planners to have access to the data remotely. With the advent of Internet-based interactive mapping technologies that enable engineers and planners at remote locations to view maps and attribute data using commonly used web browsers, remote access to map and attribute data has become commonplace. To illustrate this approach, the researchers implemented a web site that shows the different GIS data layers developed as part of the research. The web site can be accessed at http://imr.tamu.edu.

CHAPTER 7. CONCLUSIONS

Urban areas along the Texas-Mexico border have experienced phenomenal growth rates in recent years. Transborder trade and transportation volumes are also growing rapidly as well. Despite recent economic slow downs and the implementation of tighter security measures at the border, most estimates point to a continued growth in the region in the foreseeable future. It is therefore critical for planners, engineers, and decision makers to develop a clear understanding of the various processes that drive, and are affected by, such growth.

A number of government agencies are responsible for collecting and processing transportation-related data in the border region. Unfortunately, the available data tend to be scattered among various organizations and rarely include substantial amounts of information from the Mexican side of the border. There is a need to provide an integrated approach to the issue of transportation data collection and analysis for the Texas-Mexico Border Region. This research developed a prototype GIS-based framework for transportation data with a goal to better understand the characteristics of the transportation system along border areas.

The research examined existing transportation data sources at various jurisdictional levels and evaluated the degree to which these data sources could be integrated in a GIS environment. While some data were already available in GIS format, e.g., road networks, land cover, and land use, most of the data available were aggregated and not in a format suitable for inclusion in a GIS. Some other data, mainly volume data and crash data, which are critical data elements for most transportation analyses, contained some georeferencing elements; however, they were not in a GIS format. Some of the agencies that develop and maintain those data elements have started programs to develop GIS-based inventories of volume and crash data. However, it will be several years before those programs are in place. In general, the researchers detected a lack of documentation in the transportation community with respect to procedures for documenting and integrating existing traffic volume and crash data into a GIS format. The decision was therefore made to focus on the development of those procedures.

The research did not address border crossings. A number of organizations compile and publish border crossing trends and statistics. Unfortunately, most of the data available are aggregated by port (i.e., data are not divided by individual border crossings). Some descriptive information about individual border crossings is generally available, but not individual border crossing traffic volume data. If these data become available, the database schema could be very easily expanded to accommodate the new data source.

Geocoding crash data varied depending on whether the crash occurred on a state highway or a local (jurisdiction) street. In the case of state highways, geocoding was possible by using control section and milepoint data. In the TxDPS database, field CONTSEC1 represents a control section, and field MILE1 represents milepoint. Because TxDOT's GIS-based road inventory uses a reference marker-distance approach for linearly referencing objects or events along the state highway network, it was necessary to also use an equivalence table to convert data from the control section-mile point system to the reference marker-distance system. The researchers obtained the equivalence table from TxDOT. In the case of local streets, geocoding was possible by using street name and block number data. In the TxDPS database, two fields, PRIM and

BLOCK, represent street name and block number, respectively. In reality, the PRIM field contained a street name code, which means that, for geocoding purposes, it was necessary to also have a lookup table containing the equivalence between street name codes and street names. The researchers obtained a copy of the lookup table from the Laredo Police Department.

The researchers geocoded traffic volume data that TxDOT provided. Three types of traffic volume data were available: (a) ATR report, which summarized data for all ATR stations around the state, including average hourly volumes, AADT data, and vehicle classification data; (b) TxDOT Laredo District traffic map, which contained AADT data at specific count stations on state highways in the Laredo District; and (c) Webb County Urban Study map, which contained ADT data for local (jurisdiction) streets in Webb County. In general, traffic volume data were not georeferenced, which meant that the researchers had to follow a manual approach for converting the data to a GIS format. The procedure essentially consisted of locating volume data stations using the approximate location information available on the maps and cross-referencing that information with road base map data in the GIS. After finding an approximate location for the volume data stations, the researchers snapped the volume data station point features to linear features in the GIS.

The report also summarizes an assessment of transportation data on the Mexican side of the border that resulted from a collaborative effort with IMT. IMT researchers evaluated existing data at the federal, state, and local levels in Mexico and produced a document that summarized their findings. They also focused on existing highway network, traffic volume, and crash data sources. IMT researchers faced a number of challenges derived from the lack of common data frameworks and standards for minimum data content and positional accuracy. They concluded, however, that the prototype system they developed for the Mexican side laid the foundation for more integrated approaches to transportation data management for the border region and more effective interagency collaboration.

After converting each data source to a GIS format, the final step was integrating the data to make sure all layers aligned properly. This process also involved integrating the data into a single geodatabase, which included dealing with coordinate system integration issues and base map discrepancies. In some cases, the researchers observed noticeable offsets when projecting data from one coordinate system to another and had to manually correct the offsets to ensure proper data overlaying. Also noticeable were the differences between ESRI's StreetMap USA map and TxDOT's urban files. The differences were not uniform, clearly indicating that a simple map translation and/or rotation transformation would not solve the problem. It would be ideal if the same road base map could be used for all data sources. Unfortunately, this is not feasible at present because of the dependency between some data sources and their underlying road base maps. As new data collection and data management systems with the capability to record latitude-longitude data in the field become available, the need to depend on road base maps to generate GIS features will decrease. In the mean time, analysts need to be aware of the differences between various maps and account for such differences when conducting analyses that involve spatial overlaying and querying.

The research focused more on system architecture issues and design rather than applications, although the report did include a couple of application examples to illustrate some of the capabilities of the approach discussed here. In general, the research addressed two fundamental issues affecting the availability and use of traffic volume and crash data: georeferencing and integration. The research addressed these issues by developing a GIS-based prototype and by laying the foundation for the integration of additional data sources as these data sources become available. As a prototype, the system only handles a few data types. However, it is sufficiently generic and expansions appear feasible because of the GIS-based relational database structure that was chosen for its development.

Although the prototype system incorporates data elements from both Texas and Mexico, clearly a lot more needs to be accomplished to develop a truly binational transportation data management system. Transportation agencies on both sides of the border frequently meet and interact to deal with traffic operation issues affecting crossborder traffic. If developed and properly implemented, a system based on the prototype discussed here could help make the crossborder interaction more effective, therefore increasing the chances that strategies that are intended to meet the needs on both sides of the border are successful. With the advent of Internet-based interactive mapping technologies that enable engineers and planners at remote locations to view maps and attribute data using commonly used web browsers, crossborder interaction could become even more effective. To illustrate this approach, the researchers implemented a web site that shows the different GIS data layers developed as part of the research. The web site can be accessed at http://imr.tamu.edu.

REFERENCES

Acha, J., and Castrellón, R.A. (2003). Evaluación Integral de la Información de Transporte de la Región de Nuevo Laredo (Documento de Trabajo). Coordinación de Integración del Transporte, Instituto Mexicano del Transporte, 60 p.

Binational Border Transportation Planning and Programming Study (1998). Retrieved August 1, 2003, from http://www.fhwa.dot.gov/binational.

Czerniak, R.J. (2003). Binational Transportation Geographic Information System Project. New Mexico State University, New Mexico Department of Transportation and Federal Highway Administration, Las Cruces, NM. (email communication).

Environmental Systems Research Institute (2002). Geocoding Matching Rules Used for StreetMap. File stmap.mat (ArcGIS\Geocode\ installation folder).

Environmental Systems Research Institute (2003a). ArcGIS Streetmap USA. Retrieved June 20, 2003, from http://www.esri.com.

Environmental Systems Research Institute (2003b). Summary of Controlling Match Weights and Computing Scores in ArcView Geocoding. Retrieved July 2, 2003, from http://www.esri.com.

Elmasri, R., and Navathe, S.B. (1989). Fundamentals of Database Systems. Redwood City, CA: Benjamin/Cummings Publishing Company.

Federal Geographic Data Committee (2003). Content Standard for Digital Geospatial Metadata (CSDGM). Retrieved June 20, 2003, from http://www.fgdc.gov/metadata/contstan.html.

H.R. 2230, 2001 Leg., 77th Reg. Sess. (Tex. 2001) (enacted). Retrieved October 18, 2002, from http://tlo2.tlc.state.tx.us/cgi-bin/tlo/textframe.cmd?LEG=77&SESS=R&CHAMBER=H&BILLTYPE=B&BILLSUFFIX=02230&VERSION=5&TYPE=B.

Laredo Development Foundation (2003). Retrieved August 23, 2003, from http://www.laredo-ldf.com.

Navigation Technologies (2003). Retrieved June 20, 2003, from http://www.navtech.com

Nuevo Laredo (2003). Sistema Carretero. Retrieved August 1, 2003, from http://www.nuevolaredo.gob.mx/tumunicipio/economia/sistema_carretero.htm.

Tele Atlas (2003). MultiNet Street Data. Retrieved June 20, 2003, from http://www.etak.com.

Texas Center for Border Economic and Enterprise Development (2003). Retrieved August 1, 2003, from http://texascenter.tamiu.edu.

Texas Department of Public Safety (1996). Motor vehicle traffic accidents. Austin, TX: Author.

Texas Department of Public Safety (2003). Accident Records Forms. Retrieved August 23, 2003, from http://www.txdps.state.tx.us/forms.

Texas Department of Transportation (2003). Retrieved August 1, 2003, from http://www.dot.state.tx.us.

Texas Department of Transportation (1998). District Highway Traffic Map of Laredo District. Austin, TX: Author.

Texas Department of Transportation (2000a). 2000 ATR Vehicle Classification & Border Trend Report. Austin, TX: Author.

Texas Department of Transportation (2000b). Traffic Map of Laredo District. Austin, TX: Author.

Texas Department of Transportation (2002). Texas-Mexico International Bridges and Border Crossings. Existing and Proposed. Austin, TX: Author. Retrieved August 28, 2003, from http://www.dot.state.tx.us/iro/.

Texas Natural Resources Information System (2003). Retrieved August 1, 2003, from http://www.tnris.state.tx.us.

Texas State Data Center (2003). Retrieved August 1, 2003, from http://txsdc.tamu.edu.

U.S. Census Bureau (2001). Metropolitan Areas Ranked by Percent Population Change: 1990 to 2000. Census 2000 PHC-T-3. Ranking Tables for Metropolitan Areas. Retrieved August 26, 2003, from http://www.census.gov/population/cen2000/phc-t3/tab01.pdf, 2001.

U.S. Census Bureau (2003). Retrieved August 1, 2003, from http://www.census.gov.