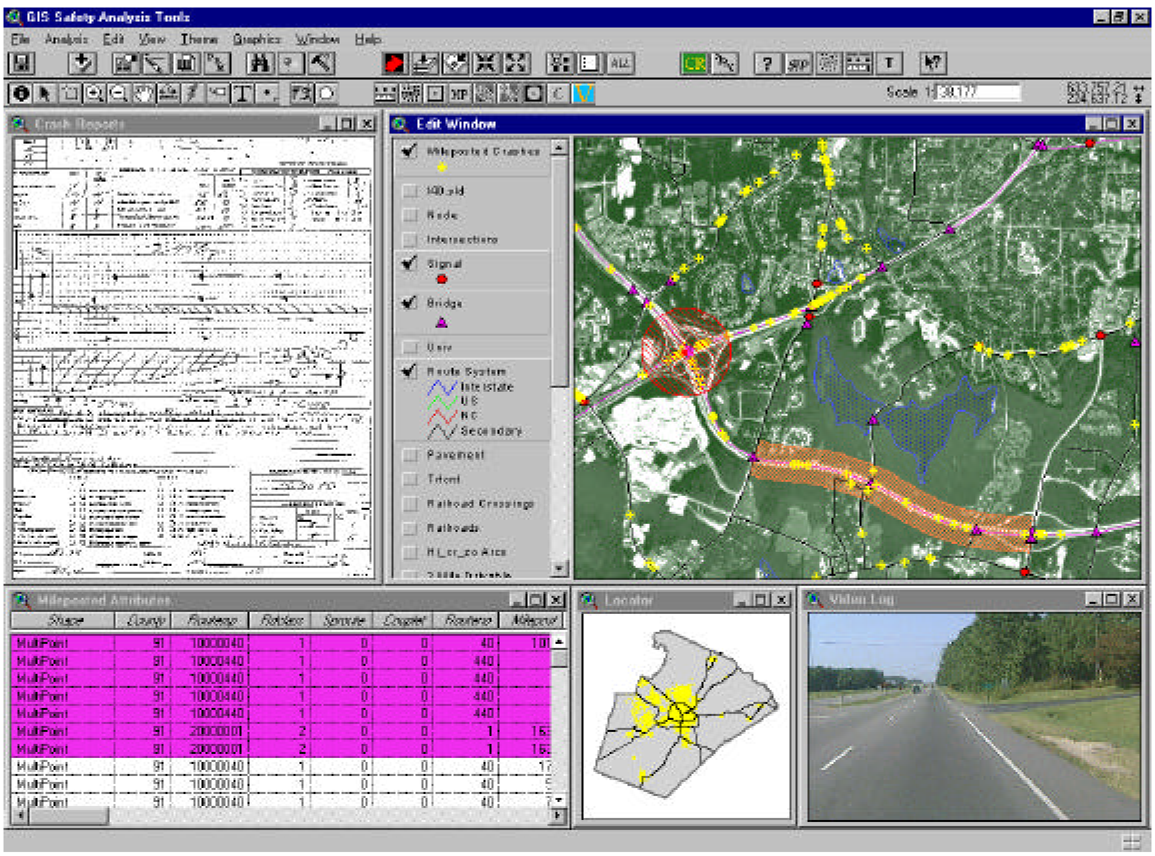
Implementation of GIS-Based Highway Safety Analyses: Bridging the Gap

PUBLICATION NO. FHWA-RD-01-039 January 2001



U.S. Department of Transportation
Federal Highway Administration
Research and Development
Turner-Fairbank Highway Research Center
6300 Georgetown Pike
McLean, VA 22101-2296

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA-RD-01-039		

4. Title and Subtitle	5. Report Date / 6. Performing Organization Code
IMPLEMENTATION OF GIS-BASED HIGHWAY SAFETY ANALYSES: BRIDGING THE GAP	
7. Author(s)	**8. Performing Organization Report No.**
Richard C. Smith, David L. Harkey, and Bobby Harris	
9. Performing Organization Name and Address	**10. Work Unit No. (TRAIS)**
GIS/Trans, Ltd. 8555 16th Street Suite 700 Silver Spring, MD 20910 University of North Carolina Highway Safety Research Center 730 Airport Road, CB #3430 Chapel Hill, NC 27599	**11. Contract or Grant No.** DTFH61-96-C-00063
12. Sponsoring Agency Name and Address	**13. Type of Report and Period Covered**
Office of Safety Research and Development Federal Highway Administration 6300 Georgetown Pike McLean, VA 22101-2296	Task Report April 1999 – Sept. 2000 **14. Sponsoring Agency Code**

15. Supplementary Notes

Contracting Officer's Technical Representative (COTR): Michael Griffith, HRDS

16. Abstract

In recent years, efforts have been made to expand the analytical features of the Highway Safety Information System (HSIS) by integrating Geographic Information Systems (GIS) capabilities. The original version of the GIS Safety Analysis Tools was released in 1999 and provided practitioners with programs to perform spot/intersection analysis, cluster analysis, strip analysis, sliding-scale evaluations, and corridor analysis. The updated version of this product has just been released and includes additional pedestrian and bicycle safety tools to select safe routes to schools, assess the bicycle compatibility of roadways, and define high pedestrian crash zones.[1-2]

One of the continuing goals of distributing the GIS Safety Analysis Tools is to encourage the safety engineers and others within State and municipal departments of transportation and metropolitan planning organizations to explore the capabilities of the GIS-based highway safety analysis tools and to adapt those ideas and applications to fit their particular needs. However, due to the variety of implementations of GIS that exist within these organizations, developing capabilities in highway safety analysis requires an understanding of the requirements of GIS, linear referencing systems (LRS), and GIS-based highway safety analysis applications.

The primary goal of this current effort was to discuss GIS/Safety integration in terms that can be understood by both safety engineers and GIS specialists, and to describe issues and solutions involved in the integration of GIS into safety-related analysis efforts. This task report is intended to serve as an educational document for both safety engineers and GIS professionals and to initiate a common dialogue. Hopefully, this report will begin to bridge the gap between the desire to implement GIS highway safety analysis within an organization and the development of a Geographic Information System – Transportation (GIS-T) infrastructure to support that effort.

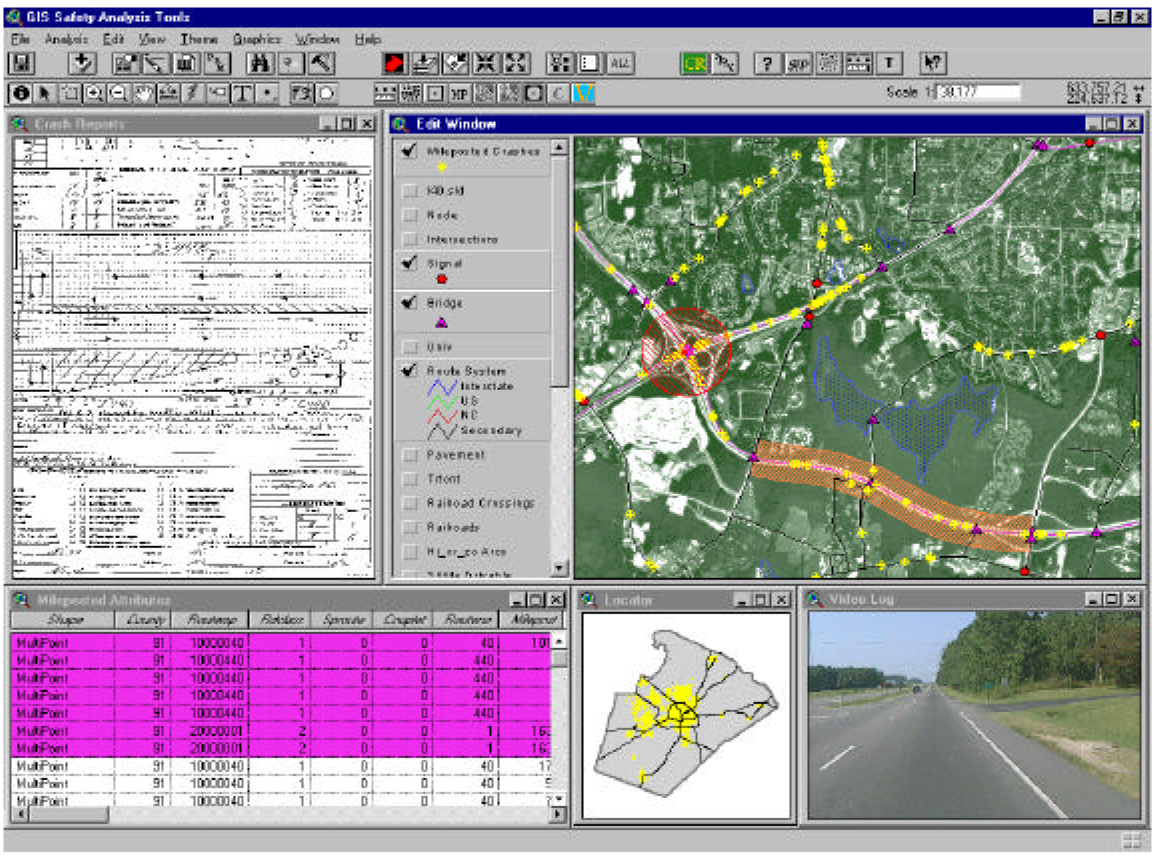
17. Key Words	18. Distribution Statement
Geographic information system, GIS, highway safety analysis, linear referencing system.	No restrictions. This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	44	

Form DOT F 1700.7 (8-72) Reproduction of form and completed page is authorized

Implementation of GIS-Based Highway Safety Analyses: Bridging the Gap

PUBLICATION NO. FHWA-RD-01-039 January 2001



U.S. Department of Transportation
Federal Highway Administration
Research and Development
Turner-Fairbank Highway Research Center
6300 Georgetown Pike
McLean, VA 22101-2296

FOREWORD

The primary goal of this current effort was to discuss GIS/Safety integration in terms that can be understood by both safety engineers and GIS specialists, and to describe issues and solutions involved in the integration of GIS into safety-related analysis efforts. This task report is intended to serve as an educational document for both safety engineers and GIS professionals and to initiate a common dialogue. Hopefully, this report will begin to bridge the gap between the desire to implement GIS highway safety analysis within an organization and the development of a Geographic Information System – Transportation (GIS-T) infrastructure to support that effort.

Michael F. Trentacoste
Director, Office of Safety
Research and Development

NOTICE

Technical Report Documentation Page

1. Report No. FHWA-RD-01-039	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle IMPLEMENTATION OF GIS-BASED HIGHWAY SAFETY ANALYSES: BRIDGING THE GAP		5. Report Date 6. Performing Organization Code
7. Author(s) Richard C. Smith, David L. Harkey, and Bobby Harris		8. Performing Organization Report No.
9. Performing Organization Name and Address GIS/Trans, Ltd. 8555 16th Street Suite 700 Silver Spring, MD 20910 University of North Carolina Highway Safety Research Center 730 Airport Road, CB #3430 Chapel Hill, NC 27599		10. Work Unit No. (TRAIS) 11. Contract or Grant No. DTFH61-96-C-00063
12. Sponsoring Agency Name and Address Office of Safety Research and Development Federal Highway Administration 6300 Georgetown Pike McLean, VA 22101-2296		13. Type of Report and Period Covered Task Report April 1999 – Sept. 2000 14. Sponsoring Agency Code
15. Supplementary Notes Contracting Officer's Technical Representative (COTR): Michael Griffith, HRDS		

16. Abstract In recent years, efforts have been made to expand the analytical features of the Highway Safety Information System (HSIS) by integrating Geographic Information Systems (GIS) capabilities. The original version of the GIS Safety Analysis Tools was released in 1999 and provided practitioners with programs to perform spot/intersection analysis, cluster analysis, strip analysis, sliding-scale evaluations, and corridor analysis. The updated version of this product has just been released and includes additional pedestrian and bicycle safety tools to select safe routes to schools, assess the bicycle compatibility of roadways, and define high pedestrian crash zones.[1-2]
One of the continuing goals of distributing the GIS Safety Analysis Tools is to encourage the safety engineers and others within State and municipal departments of transportation and metropolitan planning organizations to explore the capabilities of the GIS-based highway safety analysis tools and to adapt those ideas and applications to fit their particular needs. However, due to the variety of implementations of GIS that exist within these organizations, developing capabilities in highway safety analysis requires an understanding of the requirements of GIS, linear referencing systems (LRS), and GIS-based highway safety analysis applications.

The primary goal of this current effort was to discuss GIS/Safety integration in terms that can be understood by both safety engineers and GIS specialists, and to describe issues and solutions involved in the integration of GIS into safety-related analysis efforts. This task report is intended to serve as an educational document for both safety engineers and GIS professionals and to initiate a common dialogue. Hopefully, this report will begin to bridge the gap between the desire to implement GIS highway safety analysis within an organization and the development of a Geographic Information System – Transportation (GIS-T) infrastructure to support that effort.

17. Key Words Geographic information system, GIS, highway safety analysis, linear referencing system.		18. Distribution Statement No restrictions. This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 44	22. Price

Form DOT F 1700.7 (8-72) Reproduction of form and completed page is authorized

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yards	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	millimeters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
NOTE: Volumes greater than 1000 l shall be shown in m^3.				
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact)		
EF	Fahrenheit temperature	5(F-32)/9 or (F-32)/1.8	Celsius temperature	EC
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.71	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE (exact)		
EC	Celsius temperature	1.8C + 32	Fahrenheit temperature	EF
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
fl	candela/m^2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

* SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380.

Table of Contents

List of Figures

List of Tables

Table **Page**

Chapter 1. Introduction

The Federal Highway Administration (FHWA) operates and maintains the Highway Safety Information System (HSIS) database. [1] The HSIS integrates police-reported crash data and roadway inventory and operations data already collected by eight States for the management of the highway system and it uses these data to study roadway and roadside safety issues. Recently, efforts have been made to expand the analytical features of HSIS by integrating Geographic Information Systems (GIS) capabilities. The GIS Safety Analysis Tools represent a recent example of the work in this arena to promote the use of GIS for highway safety analyses. [2] The original version of the tools was released in 1998 and provided practitioners with programs to perform spot/intersection analysis, cluster analysis, strip analysis, sliding-scale evaluations, and corridor analysis. [3] Version 2.0 was released in July 2000 and includes additional pedestrian and bicycle safety tools to select safe routes to schools, assess the bicycle compatibility of roadways, and define high pedestrian crash zones. [4]

One of the continuing goals of distributing the GIS Safety Analysis Tools is to encourage the safety engineers and others within State and municipal departments of transportation (DOTs) and metropolitan planning organizations (MPOs) to explore the capabilities of the GIS-based highway safety analysis tools and adapt those ideas and applications to fit their particular needs. However, due to the variety of implementations of GIS that exist within these organizations, developing capabilities in highway safety analysis requires an understanding of the requirements of GIS, Linear Referencing Systems (LRS), and GIS-based highway safety analysis applications.

The primary goal of this current effort was to discuss the integration of GIS and traditional safety data in terms that can be understood by both safety engineers and GIS specialists, and to describe issues and solutions involved in developing a GIS-based highway safety analysis system. To accomplish this goal, a survey of all eight HSIS States was conducted to assess their current GIS capabilities and to determine their methods for integrating GIS and their safety data. Subsequently, two States (Maine and Washington) were selected as case studies to more fully understand the intricacies associated with this type of integration.

This final report is intended to serve as an educational document for both safety engineers and GIS professionals and to initiate a common dialogue. Hopefully, this report will begin to bridge the gap between the desire to implement highway safety analysis within an organization and the development of a Geographic Information System – Transportation (GIS-T) infrastructure to support that effort. The report does so by providing the following:

- The benefits that GIS technology offers in general analyses, including display, spatial, and network evaluations, as well as cell-based modeling. The applications from the already-developed GIS Safety Analysis Tools are discussed as examples.

- A description of how historical safety data (crashes and roadway inventory) are acquired, why such data are collected as linear referenced data, and how linear referenced data are different from spatial data. Definitions of common route systems are provided along with illustrations to show how each is different.

- General background information on Linear Location Referencing Systems (LLRS or LRS), which includes an explanation of routes and their measures, common types of LRS, how linear referencing methods (LRMs) are used to locate crashes and roadway inventory, and how GIS uses LRS to locate linear features.

- A general understanding of how GIS manages road network data and how in GIS route data are different from road network data. The impact of resolution, scale, and route calibration is discussed as they relate to data accuracy.

- A detailed discussion of the process of integrating GIS and safety data, including the need to plan for the integration and development of the GIS road network and route system, and the processing of the LRS data within GIS.

Chapter 2. What GIS Has to Offer Safety Analysis

In recent years, many transportation departments, metropolitan planning organizations, and other related agencies have begun to use GIS for a variety of data management, systems management, and planning efforts, including:

- Pavement and bridge maintenance management.
- Modeling disaster response plans.
- Quantifying the potential impacts of transportation alternatives.
- Routing of overweight and oversized vehicles.
- Flood prediction.
- Risk assessment and risk management.
- Seismic slope-performance analysis and mapping of landslide hazard zones.
- Study of air emissions on health.
- Truck traffic analysis for the management of rural highway networks.

However, one area where GIS has not been extensively used is highway safety analysis. In part, this may be due to a lack of understanding of the potential benefits of such an application. Thus, prior to developing a GIS highway safety analysis system, there is a need to have a better understanding of what GIS is and how it can benefit traditional analyses. Provided in this chapter is information that will hopefully answer the following question:

> *What does GIS offer, in terms of capabilities and features, that improves upon traditional analytical techniques and should make one consider integrating GIS and safety data?*

The present-day benefits of GIS are well established in a number of disciplines. GIS provides the capability of storing and maintaining large data sets of spatial and tabular information. GIS has its strength in providing display and analytical capabilities that model the physical proximity of spatial features. One powerful aspect of GIS is the flexibility in modeling spatial objects to suit the particular needs of the user or application. These capabilities have been developed as the technology has matured. In its infancy, GIS provided rudimentary analysis capabilities for areas that were represented as discrete points distributed throughout a uniform grid. This type of analysis is referred to as “grid” or “cell-based” analysis.

GIS has since matured to include systems based on cartographic representation of points, lines, and area feature types. These systems provide a topological data model that allows for more robust analysis capabilities, referred to as “vector-based” analysis (e.g., point-in-polygon analysis or buffer analysis). Other common GIS capabilities include database integration, image overlay capabilities, and network analyses (e.g., shortest path routing). Over the past 10 years, GIS has adapted to accommodate linear referenced data. Crash and roadway inventory data are examples of this type of linear data and can now be brought into GIS for display and analysis. This capability offers the safety engineer specific analytical methods for understanding the spatial relationship of data that are not found in other information systems.

In addition, GIS offers a programming or scripting environment that allows the user to develop specific analysis programs or customize existing programs. All functions for display and analysis can be employed in a single-system design for Rapid Application Development (RAD) using common programming languages, such as Visual Basic, C++, and Java. This capability is evident in the GIS Safety Analysis Tools, which were developed in ArcView GIS using the Avenue scripting language. [(2)] More importantly, with recent developments in interoperability, GIS can be integrated into more mainstream enterprise applications, as well as web-based thin-client applications. Spatially enabling a website to include maps of high crash areas would be one example of the latter applications.

GIS provides the ability to display and view crash and roadway inventory location, and offers great rewards not available in a linear referencing system (LRS) alone. This capability is broader than simply mapping data and includes several types of analytical capabilities that can be broadly categorized into four groups:

- Display/Query analysis.
- Spatial analysis.
- Network analysis.
- Cell-based modeling.

The remainder of this chapter discusses each of these capabilities in more detail. Where appropriate, examples of existing applications (available on the GIS Safety Analysis Tools CD) are provided.

Display/Query Analysis

The primary appeal of GIS to many is the graphical capabilities. As it has been stated that “a picture is worth a thousand words.” Maps are the pictures GIS uses to communicate complex spatial relationships that the human eyes and mind are capable of understanding. The computer makes this possible, but still, it is the GIS user that determines what data and spatial relationships will be analyzed and portrayed, or how the data will be thematically presented to its intended audience.

Using the database capabilities of GIS, the safety engineer can query the database and have the results graphically displayed. This query analysis, when spoken in everyday conversation, takes on the form of a “show me” question, such as “Can you show me all head-on collisions that resulted in a fatality?” However, query analysis capabilities in GIS can also be exploited for other purposes, such as database automation, which might be used for error checking and quality control of coded data. As an example, the GIS roadway database could be queried automatically during the crash data entry process to verify the accuracy of speed limit and other crash report variables coded by an officer.

For linear referenced data to be displayed in GIS, it first must be integrated with spatial data. GIS can integrate spatial data of various scales, resolution, and projection, although use of spatial data integration warrants caution on inappropriate use. One example of poor use of GIS data integration capabilities would be statewide roadway feature data developed from 1:500,000-scale source maps. These data will not have the same line delineation and will not fit well or be appropriate for integration with data from large-scale sources (e.g., 1:24,000).

The use of imagery in GIS in conjunction with terrain modeling can provide a virtual reality display for highway safety analysis, giving the safety engineer a realistic view of the landscape (for instance, an aerial view of an intersection or a view of trees along the roadside). Satellite imagery and digital aerial photographs are two sources that can be used for this application. Both can be rectified, which involves image processing, such as rotating, scaling, and re-sampling. The imagery data can then be fit to overlay with the GIS spatial data (or linked to features), which involves determining the image map extent coordinates. Then, the imagery can be used for feature data collection or used as a backdrop image reference.

Data integration provides a microscopic level of analysis through the ability to spatially integrate and merge the data into a single view. Data not ordinarily used by the safety engineer, data that would otherwise be external to the LRS or not have a linear reference, such as demographic data, meteorological data, environmental data, economic data, and terrain data, to name a few, can be integrated using GIS. LRS data that is not ordinarily integrated, such as work-zone data, can also be integrated within GIS, thus expanding the data sources available to the safety engineer.

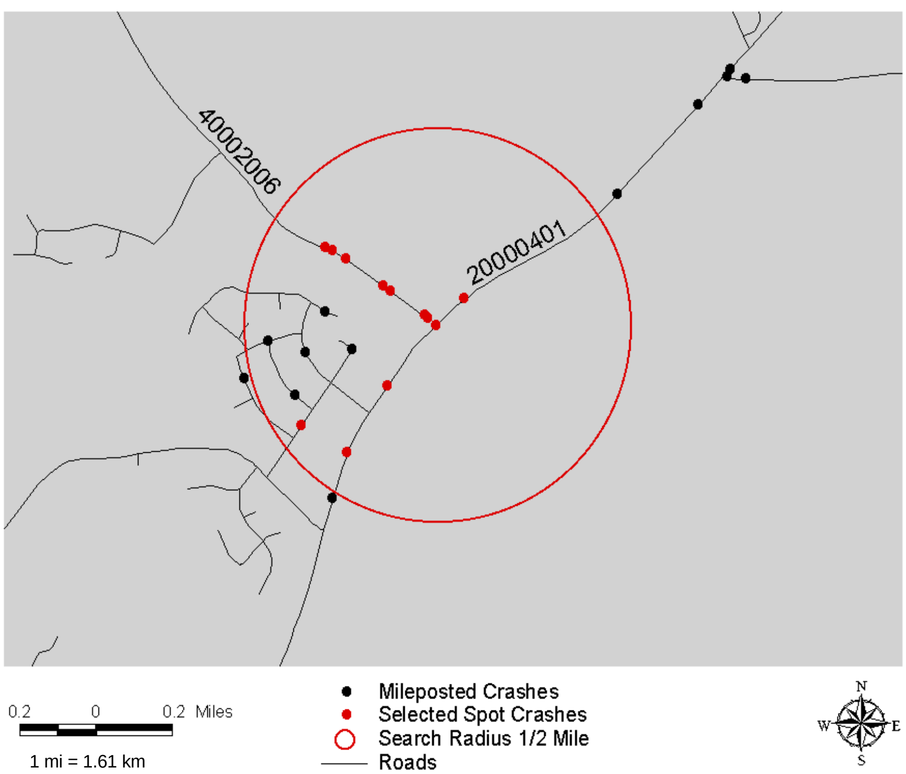
Thematic mapping of highway safety data provides a macroscopic level of analysis. Linear and spatial data integrated into GIS can be selected, differentiated by type or class, and displayed thematically. The safety engineer will be able to symbolize crashes for thematic mapping to distinguish between crashes, such as the severity of a crash resulting in fatalities and non-fatalities. These simple capabilities are the most commonly used to quickly digest large amounts of information, such as showing high crash locations or showing crash histories of road segments through the use of graduated line weight symbolization.

Spatial Analysis

Several analytical techniques, grouped under the general heading "overlay analysis," are available in GIS for spatial analysis and data integration. GIS provides tools to combine data, identify overlaps across data, and join the attributes of data sets together using feature location and feature extent as the selection criteria. Overlay techniques will combine spatial data in other ways, such as features that can be combined to simply add one spatial data set to another, or to update or replace portions of one data set with another data set. Overlay analysis can be used to merge spatial data by combining two or more spatial data sets to produce a new spatial data set where the feature attributes are a union of the input data sets. As an example, the safety engineer can use these spatial techniques to combine demographic data, such as the number of households, showing the average number of school age children, with road segments having crash data showing pedestrian-related crashes, in order to derive risk factors for the total number of pedestrian-related crashes relative to the total number of school age children per road segment, for pedestrian-to-school safety analysis.

Proximity analysis is a type of GIS query capability and a category of spatial analysis that represents the fundamental difference of GIS from all other information systems. Buffering is a means of performing this practical spatial query to determine the proximity of neighboring features. In GIS, buffering will locate all features within a prescribed distance from a point, line, or area, such as determining the number of crashes that occurred within 800 m (0.5 mi) of an interchange, or locating secondary crashes that occurred within a certain distance and time (e.g., 400 m (0.25 mi) and 30 min) of other crash events, although reliability of these variables may not always support this example. Examples of proximity analysis applications on the GIS Safety Analysis Tools CD include Spot/Intersection Analysis, Strip Analysis, and Cluster Analysis.

The *Spot/Intersection Analysis* routine is used to evaluate crashes at a user-designated point or intersection for a given search radius. The spot or intersection of interest can be selected by clicking on the map using the mouse or by entering the intersecting route/street names. The end result of this analysis is a report that lists the number of crashes, fatalities, injuries, costs, etc. (as defined by the user) and a graphic that can be output as a hardcopy map (see figure 1) depicting the spot, search radius, and selected crashes.



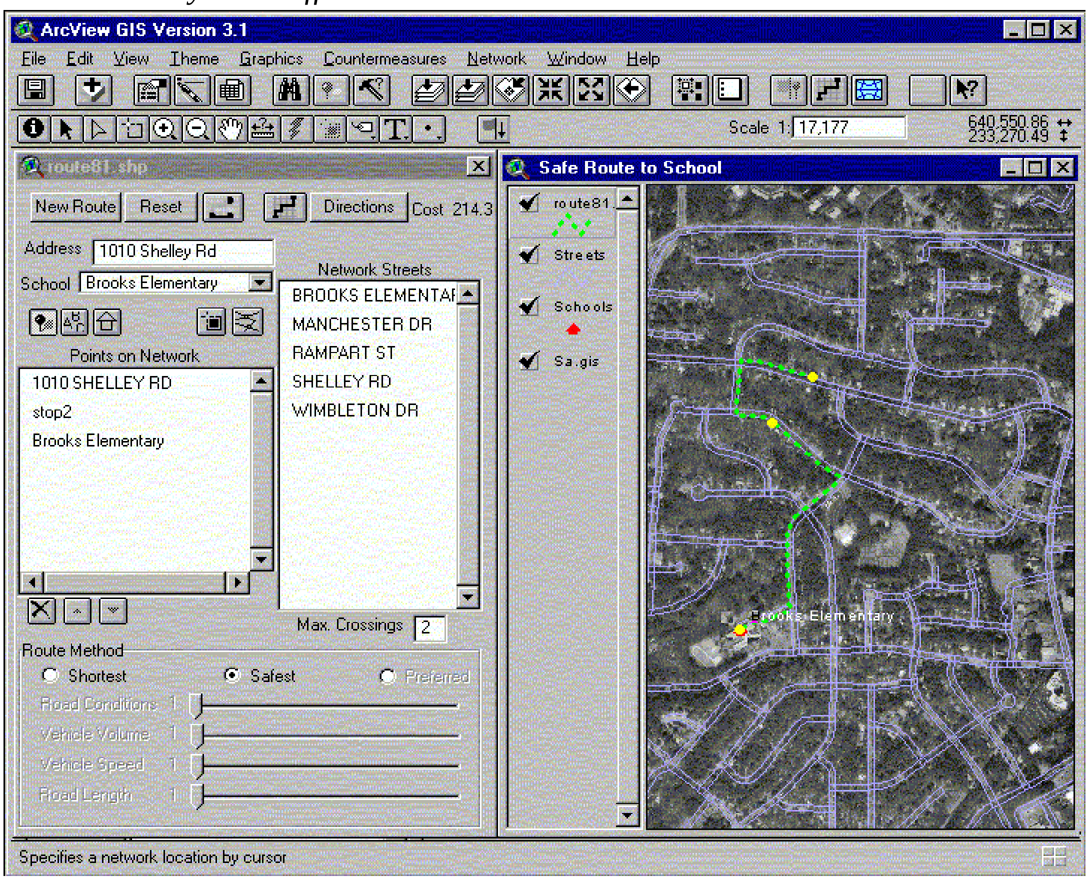
The *Strip Analysis* routine is used to study crashes along a length of roadway rather than a finite location, spot, or intersection. The user must provide the section length to be used for the analysis as the program traverses the route (e.g., every 1.0 km) and the name/number of the route. The end result of this analysis is a report that lists the number of crashes and other user-defined attributes, and a graphic that can be output as a hardcopy map depicting the buffer that makes up the strip, selected crashes, and roadway identifiers.

The *Cluster Analysis* routine is used to study crashes clustered around a specific roadway feature, such as a bridge or railroad crossing. Crashes are identified that fall within a given distance on all selected routes. Again, the output is a report that lists various summary statistics selected by the user and a map depicting the high crash locations.

Network Analysis

Unlike proximity analysis that searches in all directions from a point, line, or area, network analysis is restricted to searching along a line, such as a route, or throughout a network of linear features, such as the road network. Network analysis can be used to define or identify route corridors and determine travel paths, travel distances, and response times. For example, network analysis may be used to assess

Figure 2. The Safe Route to School application selects the best route between an origin and a school based on roadway and traffic conditions.



the traffic volume impact of a road closure on adjacent roadways.

GIS networking capabilities can also be used for the selection of optimal paths or routes. The Safe Route to School application (see figure 2) on the GIS Safety Analysis Tools CD is an example of this type of application. The user inputs the origin and destination, and the program produces a map and walking directions for the preferred route, which is based on the level of hazard associated with the various roadway and traffic elements.

To improve the network model and provide the capability of automated route selection, the road network can be developed to include turning points, avoid improper turns onto one-way streets, represent posted traffic control restrictions, and include impedance factors to travel (such as mean travel speeds, number of travel lanes, and traffic volumes) to enhance the network analysis. *Note: Network routing capability is not available with all GIS, some GIS vendors offer network capabilities as an extension or additional modules to their software products at an additional cost.*

Other examples of network analysis tools that have been developed and are available on the GIS Safety Analysis Tools CD include the Sliding-Scale Analysis and Corridor Analysis programs. The *Sliding-Scale Analysis* routine is used to identify roadway segments with a high crash occurrence. This program differs from the Strip Analysis program in that the analysis segment is not fixed, but rather slides along the route in an incremental fashion. The user defines the segment length and the increment length for analysis. The end result of the analysis includes a table showing the high crash locations that exceeded a calculated or user-defined threshold, along with a variety of summary statistics and a map showing these locations.

The *Corridor Analysis* routine provides a visual means to locate high crash concentrations within a corridor. Using traditional methods, segments along a specific route could be examined (e.g., by using the sliding-scale analysis), but multiple routes within a corridor could not be easily linked and analyzed as a group. This program allows routes to be linked together in a manner that allows the analyst to assess the overall safety performance within a transportation corridor. In a recent evaluation, the program was used to examine truck crashes along designated truck corridors in a county in North Carolina. [5] In this case, State laws permit trucks to drive on any designated truck route and along any intersecting routes for a distance of up to 3 mi (4.8 km). The Corridor Analysis program was subsequently developed to identify truck crashes on roadways within the 3-mi (4.8-km) driveable zones. The output of the analysis included crash statistics and a variety of roadway characteristics for each high crash zone in the corridor. In addition, several plots depicting high crash segments and zones were also produced. The plot shown in figure 3 shows the high truck crash segments, including three such segments that were not on designated truck routes and were outside the 3-mi (4.8-km) driveable buffer.

Cell-Based Modeling

Cell-based modeling, also referred to as "grid-based" analysis, uses a grid or cells to aggregate spatial data for discrete distribution. In cell-based modeling, the spatial data are developed as tiles of a given

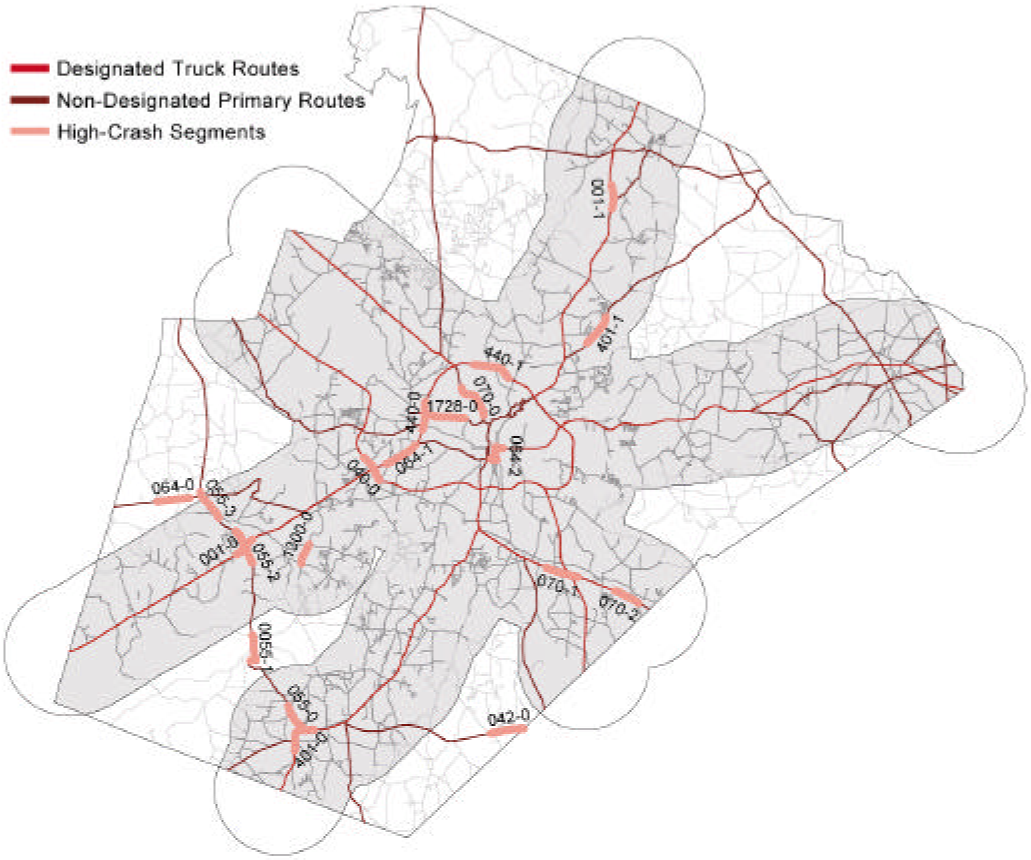


Figure 3. High truck crash segments that were identified using the Corridor Analysis program, including three segments outside of the designated truck zone.

dimension, or points of a uniform distribution, as defined by the user, for display and analysis. Cell-based modeling is effective in displaying patterns over larger areas, such as representing the sum total of crashes that are located within a cell. This capability provides a quick means to view spatial clustering of crash data. This technique is favored among DOTs and MPOs that assign crash data to street midpoints and street intersections, a method that in and of itself forms data clusters. Since cell-based modeling aggregates data at a specified grid resolution, it *would not* be appropriate for site-specific spatial analysis.

In cell-based modeling, special tools are available to merge grid data for overlay analysis. Cell-based overlay analysis is similar to the GIS overlay analysis previously discussed; however, the techniques and functions available in cell-based modeling are somewhat different. When the cells of different data sets have been developed using the same spatial dimensions, they can be merged on a cell-by-cell basis to produce a resulting data set. The functions and processes used in cell-based modeling to merge grid data are referred to as “map algebra,” because the grid data sets in cell-based modeling are merged using arithmetic and Boolean operators called “spatial operators.”

The High Pedestrian Crash Zone application on the GIS Safety Analysis Tools CD makes use of this technique. The program uses a discrete point file to calculate the density of selected crashes and generates a contour map identifying areas of high crash occurrence (see figure 4). Summary statistics of the various zones can also be produced in tabular or graphical formats.

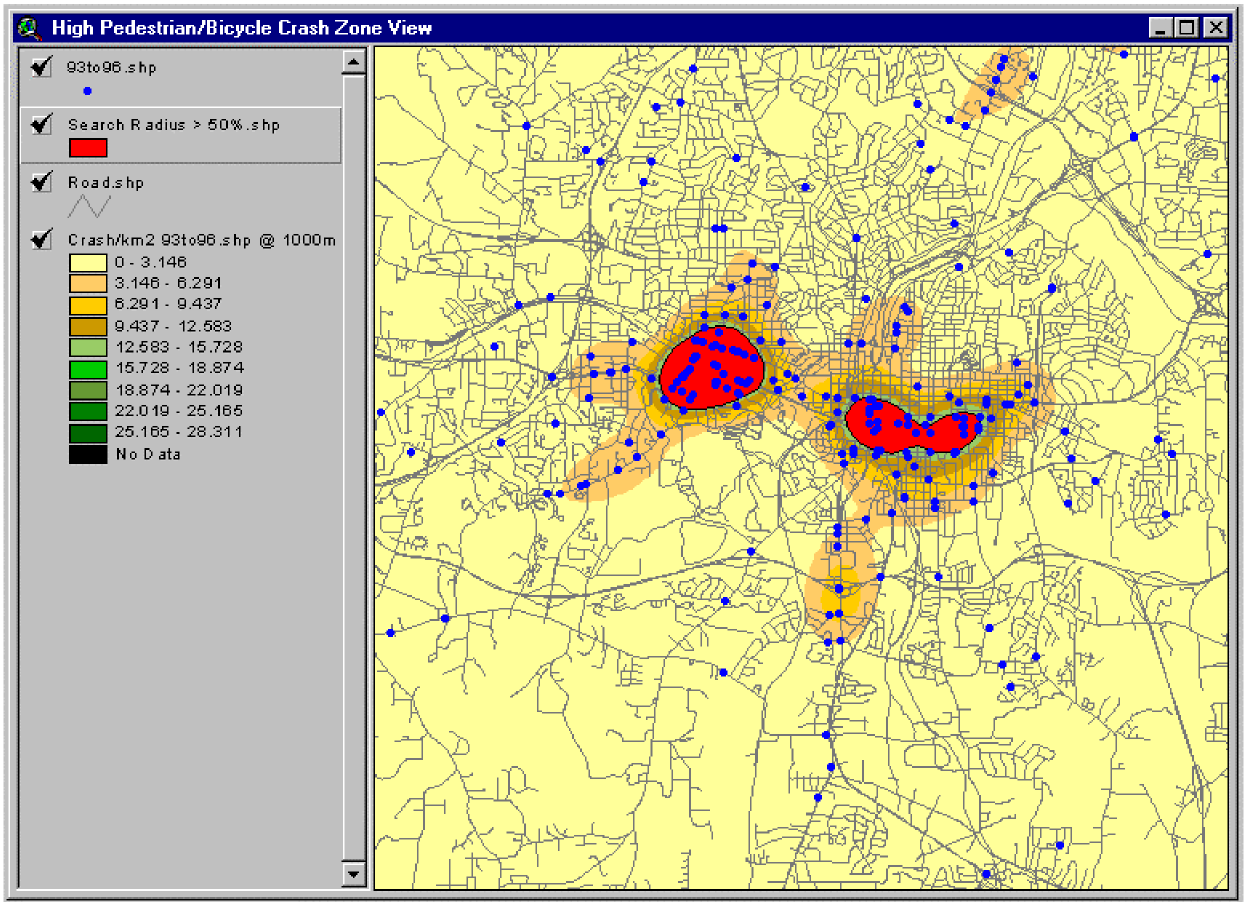


Figure 4. A view of crashes/km^2 grid with high crash zone of 50 percent and greater created in the High Pedestrian/Bicycle Crash Zone application.

Chapter 3. Understanding Traditional Data Collection Methods

Prior to integrating the GIS and safety analysis efforts, it is important to understand how data used in traditional safety analyses are collected and how GIS interprets and makes use of these data. This chapter provides an understanding of the former, while subsequent chapters explore GIS and data interpretation.

Locating crashes and roadway features is a process that traditionally has been accomplished using either references to the roadway or references to monuments along the roadway. This method is known as "linear referencing." Many different variations of linear location referencing systems (LLRS or LRS) have been defined and implemented by States and municipalities, each using various linear referencing methods (LRMs), and various designations and naming conventions. For clarification, the distinction between an LRS and an LRM is as follows:

> *Linear Location Referencing System (LLRS or LRS)* is the total set of procedures for determining and retaining a record of specific points along a [highway]. The system includes the location referencing method(s), together with the procedures for storing, maintaining, and retrieving location information about points and segments on the highways. (6)
>
> *Linear* Referencing *Method (LRM)* is the technique used to identify a specific point (location) or segment of highway, either in the field or in the office.

At times, the reference to the type of LRM or LRS is used interchangeably. However, it is important to recognize the difference when discussing route systems and to understand that the LRS is developed from the LRM.

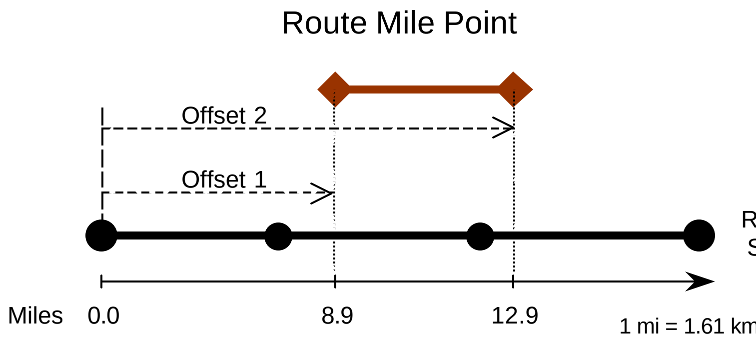
The most common location methods generally fall into one of five categories, with the last one being relatively new with the increasing use of global positioning system (GPS) technologies:

- Route-Milepost (RMP).
- Route-Reference Post (RRP).
- Link-Node (LN).
- Route-Street Reference (RSR).
- Geographic Coordinates.

> *Note: LRMs are supported in a variety of ways by the different GIS vendors. Not all of the GIS software products support all route systems, and the necessary functionality to support a particular route system may require development on the part of the user. This may be particularly true for the RRP or LN systems.*

Route-Milepost System

The Route-Milepost (RMP) system is, perhaps, the most common method used, particularly at the State DOT level. It is sometimes referred to as the “Route Mileage” system because mileage is typically the unit of measurement. In the RMP system, distance is measured from a given or known point, such as the route beginning or a jurisdictional boundary (e.g., a county line), to the referenced location. The distance is usually specified to the nearest hundredth of a mile, although some States may only specify crashes to the nearest tenth of a mile. The point of interest (i.e., crash or roadway feature) is always offset in a positive direction from the zero milepoint, and is not referenced to other intermediate points along the route. This point is illustrated in figure 5 using roadway surface condition as the roadway feature of interest.



Surface Condition Database

Route	Begin MP	End MP	Condition
SR1	8.9	12.9	Good
...			

Figure 5. A roadway surface condition specified as "good," located along an RMP between mile point (MP) 8.9 and 12.9 as an offset from the route beginning, not referencing intermediate points in the LRM.

Route-Reference Post System

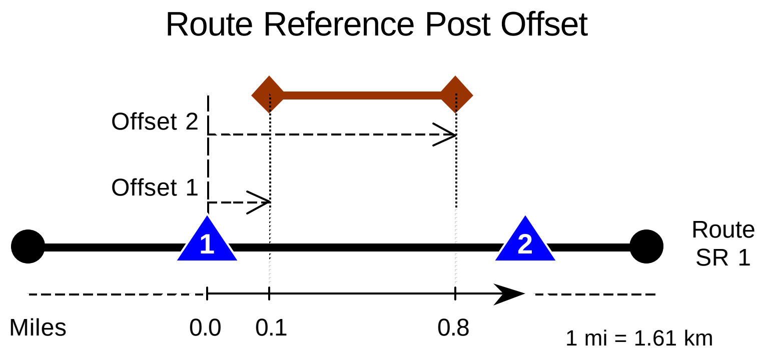
The Route-Reference Post (RRP) system is a method that uses signs posted in the field to indicate known locations. These signs, known as “reference posts,” may or may not reflect mileposts. All crash and roadway feature data collected in the field are referenced to these markers in terms of distance and direction. These field-recorded events can later be converted to corresponding mileposts using cross-referencing tables and maps. The advantage of this system over an RMP system is the elimination of the problems caused by changes in route length that may be the result of realignment. Figure 6 illustrates the RRP LRM and uses roadway median type as an example.

Link Node System

In a Link-Node (LN) system, specific physical features, such as intersections, are identified as nodes. Each node is considered unique and is assigned a unique identifier or node number. Links are the logical connection between nodes and may vary in length. Links also have unique identifiers that are often derived from the associated pair of Node identification (ID) numbers. All crashes or roadway features are measured as an offset distance from the nearest or lowest node number along a link. Figure 7 illustrates the LN system and shows a schema for assigning Link IDs.

Route-Street Reference System

The Route-Street Reference (RSR) system is more commonly used in many municipalities and relies on the local system of streets to locate crashes and roadway features. In this system, an event is typically recorded as occurring on one street at a specified distance and direction from another street that is used as a reference. A variation of this system is the use of two reference streets and no distance measurement. For example, a crash may be coded as occurring on Street A between Streets 22 and 23. This option results in a loss of detail with regard to precise location, but still provides enough information to determine sections of roadway that may have a high number of crashes.



Median Database

Route	Reference Post	Offset 1	Offset 2	Median Type
SR1	001	0.1	0.8	Barrier
...				

Figure 6. A roadway median type specified as a "barrier" is located along an RRP at 0.1 and 0.8 mi (0.16 and 1.29 km) offset from reference post 1, while other reference points in the LRM are not considered.

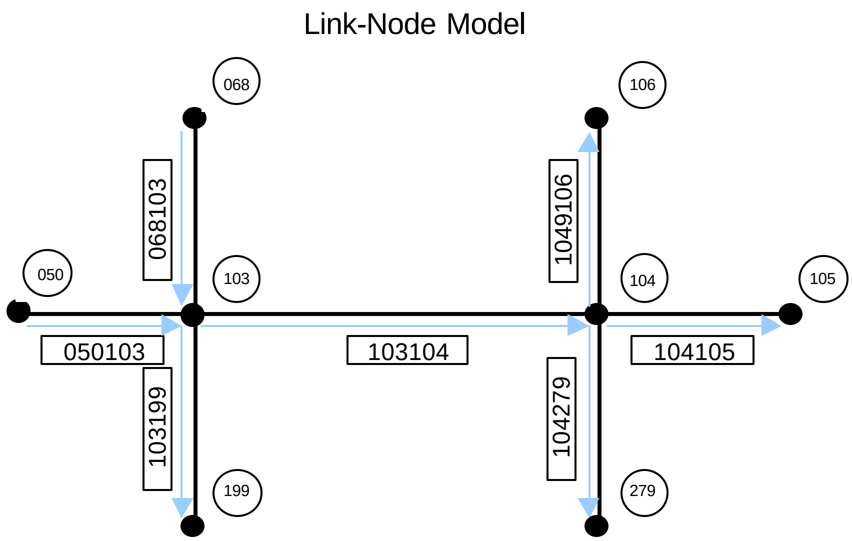
Geographic Coordinate System

Newer methods of reporting crash location information using GPS and other technologies are now available or are being developed. Unlike linear systems, coordinate systems use two or more spatial references that have equal significance. Cartesian coordinates use x and y (x-y) to measure distance along perpendicular axes of a coordinate plane. Geographic coordinates use latitude and longitude to measure distance in degrees along the axes of the sphere of the earth.

Crashes (and beginnings and endings of route segments) can also be located using GPS technology to reference, by latitude and longitude, a location on the earth’s surface. Local transportation authorities may use State plane coordinates to measure (in meters or feet) the distance east and west or north and south along a State origin or datum.

Developing the LRS for Crash Data and Roadway Inventory Data

Regardless of which LRM is used, the procedures used by State and local DOTs to collect and process the crash and roadway inventory data are generally the same. A brief overview of these generic procedures and the resulting data formats that are available for safety analysis efforts are provided in this section. The example provided refers to an RMP system, but would be applicable to any LRM.



Route Section Database Table		
Link ID	Begin Node ID	End Node ID
050103	050	103
068103	068	103
103104	103	104
103199	103	199
104105	104	105
104106	104	106
104279	104	279

Figure 7. LN LRM showing links composed of node pairs, with each Link ID being unique and composed of unique Node IDs.

Collecting roadway inventory data, such as number of lanes, shoulder type, and pavement surface, is often accomplished in the field by driving along the roadway. As the inventory item is located, its attributes are recorded, along with the road name (or Route ID) and the mileage driven (or milepost). Mileage attributes for the various elements are generally recorded in one of two ways. For *point features*, such as a signpost or a culvert, a *single* mileage attribute is recorded. For an item located along a stretch of roadway, such as the number of lanes, shoulder type, or pavement surface, a *beginning* mileage and *ending* mileage is recorded. In GIS, the data and attributes associated with the LRS are known as “events,” i.e., point events or linear events. The result of this type of roadway inventory data collection can be represented graphically as shown in figure 8, where each inventory item along the route is associated with specific beginning and ending milepoints.

Most States collect and maintain attribute data on roadway characteristics as a single table containing records representing homogeneous sections of highway, such as represented in table 1, depicting pavement and shoulder type. This information may also be entered into a relational database

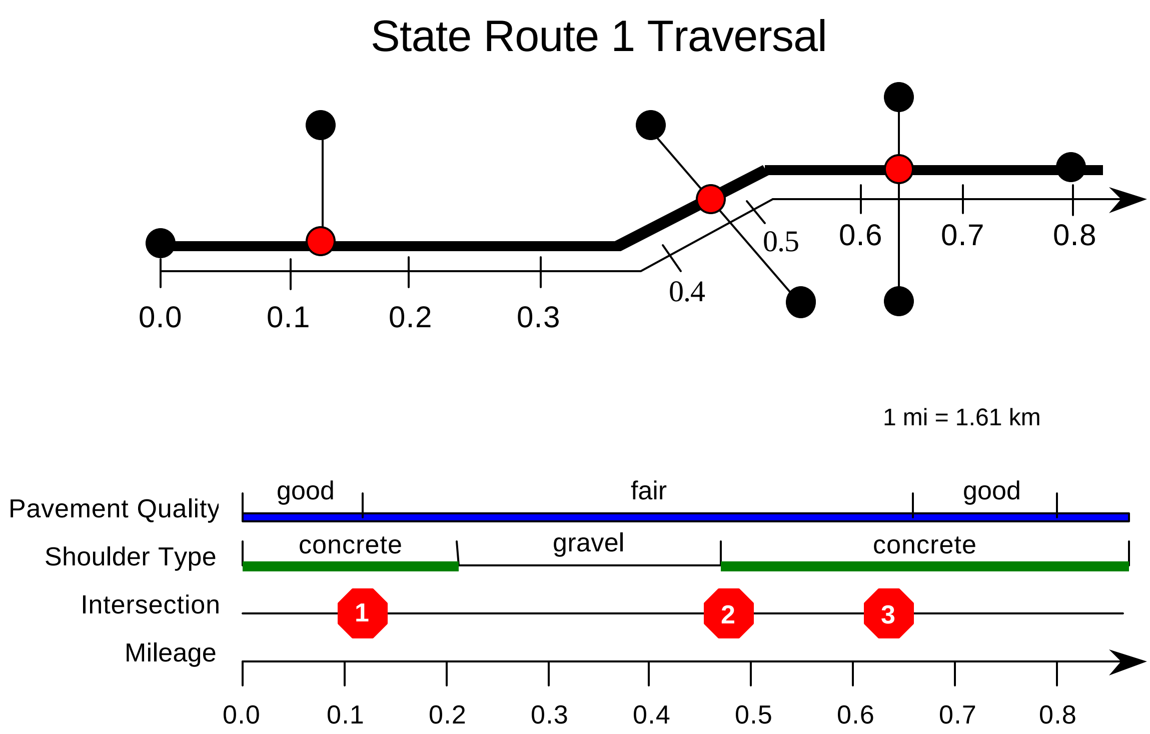


Figure 8. The traversal of State Route 1 graphically shows collection of pavement, shoulder type, and intersection items as the roadway is measured.

management system (RDBMS). Each record in the database would be entered for each observed and recorded occurrence. The attributes for roadway inventory would include Route ID, Mileage, and Inventory Type. Each data type could be entered into database tables, such as a Pavement File, Shoulder File, and Intersection File, as illustrated in figure 9.

Collecting crash location information is somewhat different in that no planning usually takes place to measure a crash location from the route beginning. Instead, crash location is usually measured from the nearest reference (e.g., milepost or intersection). However, crash locations are brought into the same LRS as roadway inventory through the coding process.

Table 1. Depiction of roadway characteristics showing pavement quality and shoulder type (each record represents a homogeneous section of highway).

Section File					
Route ID	**Beginning Mileage**	**Ending Mileage**	**Section ID**	**Pavement Quality**	**Shoulder Type**
SR1	**0.00**	**0.13**	**1**	**good**	**concrete**
SR1	**0.13**	**0.21**	**2**	**fair**	**concrete**
SR1	**0.21**	**0.46**	**3**	**fair**	**gravel**
SR1	**0.46**	**0.65**	**4**	**fair**	**concrete**
n					

The officer at the scene of a crash usually cites observable features and states crash location as route, direction, and

LLRS as Database Tables

Pavement File

Route ID	Begin Mileage	End Mileage	Condition
SR1	0.00	0.13	Good
SR1	0.13	0.65	Fair
SR1	0.65	0.80	Good
...			

Shoulder File

Route ID	Begin Mileage	End Mileage	Type
SR1	0.00	0.21	Concrete
SR1	0.21	0.46	Gravel
SR1	0.46	0.90	Concrete
...			

Intersection File

Route ID	Mileage	Cross-Street Name
SR1	0.13	Main Street
SR1	0.47	Maple Street
SR1	0.64	State Route 5
...		

Figure 9. Partial illustration of a route system roadway inventory data model.

offset. Then a DOT “Coder” interprets the officer’s location description and assigns a route code and mileage attributes to the crash location. An exception may be an MPO or urban area authority that might use street intersection coordinates or a street mid-point designation instead of a standard LRS. GPS use for crash location is also being used in some instances, but for the most part, it is not in widespread use by the enforcement community at the time of this writing.

The Coder puts the crash into the LRS by interpreting the location information from the crash report and determining or interpolating a precise linear location. Coders rely on additional information sources, such as roadlog reports that provide a listing of route mileage for cross-streets, roadside features, etc., and assist with correctly locating a crash. Crash attributes would include Crash Case Number, Route ID, and Mileage, and would be entered as crash records within the LRS. For example, a crash report may describe the location for Crash Case No. 2000-0954 as “Interstate 65, 50 ft north of intersection with U.S. 10.” The Coder may translate this information into a linear location crash event as “Case No. 2000-0954, Route ID I-65, Milepoint 2.71.”

Chapter 4. Understanding the Roadway Within GIS

In GIS, the roadway is represented as a collection of lines with endpoints defined in coordinate space. A combined collection of graphical links form a roadway network, but this representation alone is considered as having no “intelligence.” That is, connectivity and designated route topology are not present. “Routes” are special feature types constructed from the roadway line features (e.g., route number) and can be designated in GIS using relational database tables to identify those lines that make up each route. Routes also have a location method associated with them that allows event locations, such as a crash location milepoint, to be positioned on the route. To implement this capability in GIS, route “measures” are assigned as attributes to the route at the starting, ending, and intermediate points along the route. The intermediate route measures are used to control location placement accuracy for events along the route (see route calibration discussion later in this chapter).

> *The development of routes in GIS varies by vendor and available GIS software. In general though, routes can be: (1) wholly or partially coincident with other routes, (2) disjointed or disconnected, and (3) defined with sections containing route measure attributes.*

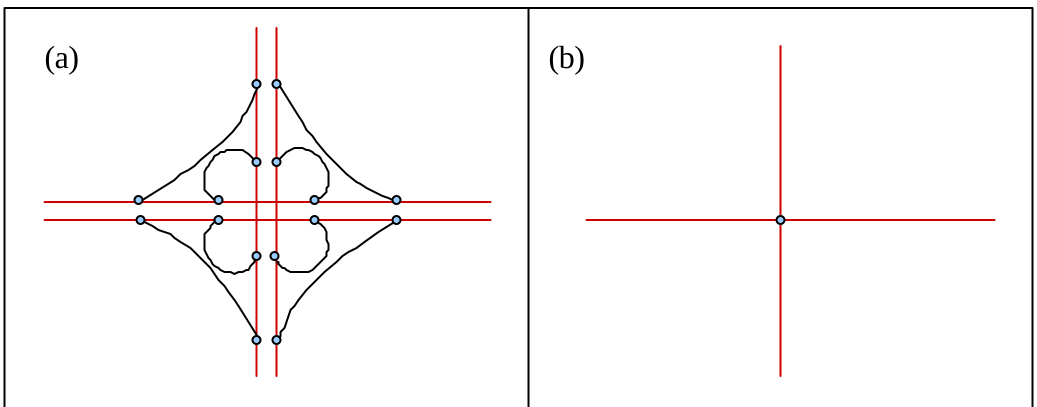
Resolution and Generalization

GIS has historically relied on a cartographic data model, similar to Computer-Aided Design (CAD) systems, to represent roadway and other feature elements. Like CAD, GIS uses a coordinate system to store and display primitive feature elements of points, lines, and areas. In the cartographic data model, the roadway is represented as a “line” feature.

Unlike purely graphical software applications however, GIS builds and manages topology in the cartographic model. A “road network” is the connection of a series of roadway “line” features having the same defined attributes. This interconnectivity of line features is important for using routing applications or in network data modeling. The GIS network applications may be used for optimal routing analysis to find efficient travel routes, closest facility analysis to determine which roadway or other facility is closest, or service area analysis to learn what is near a particular site. Appropriate connectivity and related information can be designed and built into the GIS data to develop the network data model for the support of network analysis.

Modeling the road network as a spatial or graphical layer in GIS is a planning exercise that needs to be compatible with and reflect the needs and requirements of the DOT, as it might support the daily operations of the organization. Mainline, secondary routes, collectors, and interchange features can be represented in GIS at various levels of detail. In the discussion that follows, it is important to understand the difference between small-scale and large-scale mapping. The smaller the scale, which is represented by a larger number in the ratio (e.g., 1:500,000), the less detail that can be represented. Small-scale mapping of State-maintained roads could be represented as a simple roadway or right-of-way centerline. This depiction would show intersections of mainline and secondary roads as a point, thereby showing no interchange features, such as ramps. On a moderately large mapping scale, such as 1:24,000, roadway features could be resolved at a spatial accuracy of ∀40 ft (12.2 m) to fully depict the road network for all directions of travel, showing ramps and collectors. On this mapping scale, lanes of travel, i.e., lanes for the same direction of travel, would be generalized to a single pavement centerline and would be sufficient for most LRS application needs. On much larger scales, such as 1:600 (1 in = 50 ft), features such as pavement markings, actual lane designations, and specific design elements could be graphically depicted. The latter is the level of resolution often used in roadway design work.

Consider, as an example, how an interstate interchange, having roadway mainline and collector features, would be depicted on a map. On a larger scale mapping (e.g., 1:24,000), each feature of the interchange (e.g., the mainline roadways, ramps, and intersecting collectors) would be depicted in GIS as a separate line feature, as illustrated in figure 10(a). With small-scale mapping, typically 1:500,000 or smaller, the GIS cartographic data model *would not* support the depiction of ramps and collectors, and the mainline roadway features would be generalized to a centerline representation. At this level of generalization, the interchange would be represented as a single-point feature, as illustrated in figure 10(b).



Depiction of interchange generalized as large-scale representation of mainline travel lanes and collectors, representing greater detail.	Depiction of interchange generalized as small-scale representation of roadway centerline intersection, having a reduced level of detail.

Figure 10. Same roadway interchange represented in GIS at two levels of generalization and detail.

Scale and Accuracy

In GIS, scale and accuracy are important considerations, but often these aspects of data collection are overrated when dealing with routes and highway safety analysis. Most site location analyses can be performed with nearly any scale mapping. This should not be misunderstood to mean that knowledge of the scale and accuracy of the base map or linear referenced data is not important. Generally, the source material and the standards of data development determine both the scale and precision of geospatial data sets. As an example, the U.S. Geological Survey (USGS) publishes accuracy standards for Digital Line Graph (DLG) data, such as follows:

As applied to the U.S. Geological Survey 7.5-minute quadrangle topographic map, the horizontal accuracy standard requires that the positions of 90 percent of all points tested must be accurate within one-fiftieth of an inch (0.05 cm) on the map. At 1:24,000 scale, one-fiftieth of an inch is 40 ft (12.2 m). [7]

For highway safety analysis, GIS brings together data from various sources the GIS roadway network, the LRS crash database, and the LRS roadway inventory data. When merging data from different sources, the least accurate data source (i.e., least common denominator) is used for determining overall data accuracy. In many cases, the location of the crash within the LRS crash database will be that least common denominator. Generally, a crash is recorded by the police officer and subsequently coded by an analyst to within 0.01 mi (0.016 km) or approximately 50 ft (15.2 m). Thus, for highway safety analysis, the USGS standard of ∀40 ft (12.2 m) would be considered acceptable.

The accuracy of linear referenced data is relative to, and thus mostly dependent on, the calibration of route measures along the road network, and less dependent on the accuracy of the road network data in GIS. However, the accuracy of spatial data will come into play in two ways: (1) when the road network data is overlaid with other spatial data, and (2) as the LRS is linked with the GIS route system through dynamic segmentation (the latter being the degree to which calibration needs to be performed to improve placement of crashes given the spatial resolution of the road network data set).

Another important consideration in positional accuracy is the distinction between locations referenced in different dimensions, i.e., a location referenced relative to the linear distance versus x-y coordinate space versus x-y-z spherical space. Consider the LRM that uses the RMP system as an example. Measurements are taken from the beginning of the route (or perhaps from the beginning of the route in each county) and are used to specify the offset of a feature or event along that route. Since only the length of the roadway geometry is taken into consideration, these offsets are accurate in only one dimension (i.e., linear accuracy). The linear distance for any given point along the route has accounted for the roadway curvature and grade since it essentially represents the driving distance on the roadway. In the GIS software, roads and other features are referenced using a minimum of two dimensions. In two dimensions, roads appear as if they were in plan view (i.e., being seen from above). Curves, turns, tangent sections, and intersections appear as they would on an aerial photograph or map, i.e., an orthogonal view. Curves and turns in a road obviously impact the measured distance between two nodes, as illustrated in figure 11.

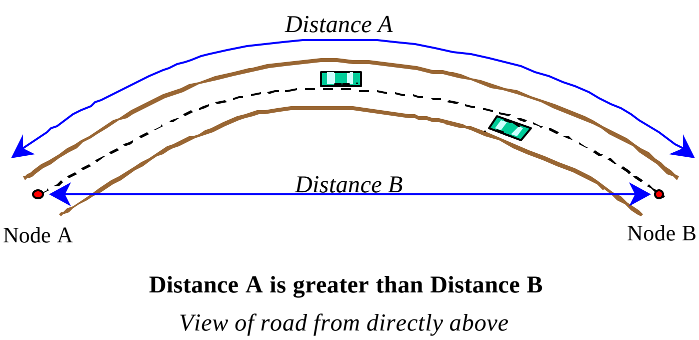
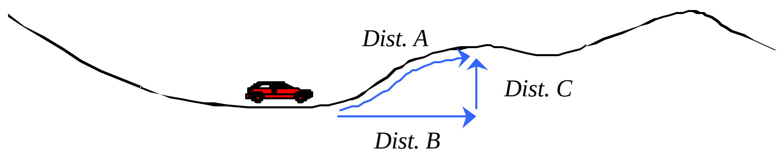


Figure 11. Location references in two dimensions.

The real world, however, is three-dimensional. In three dimensions, topography also affects the distance along a roadway the same way horizontal curves do in two dimensions (see figure 12). The measurements taken by field personnel are obviously made in the real world and, therefore, accurately record distances as they are measured along both horizontal and vertical curves. These measurements reflect the geographic accuracy of a roadway and are the distances used to reference features in an LRS. It is important that GIS properly reflect this level of geographic accuracy.



Distance A is greater than the two dimensional Distance B because of vertical component C

Cross Sectional View of Roadway

Figure 12. Location references in three dimensions.

While some of the GIS systems accurately capture data in three dimensions, most do not. This creates a problem when comparing distances calculated in a two-dimensional GIS with distances measured in the three-dimensional real world. However, this problem is fairly insignificant in most transportation applications, since the slope of most roadways has a minimal effect on distance. For example, on a 10 percent slope, the difference between horizontal and surface distance is just 0.5 percent. [8] This problem will, however, be compounded further down the length of a route, especially on roadways in mountainous terrain.

Another source of error that should be noted with regard to the length accuracy of the GIS links appears when digitizing road features from a paper map or aerial photograph. Most base maps are created from two-dimensional maps, and precision in the road network database is determined by two factors: (1) the scale of the source data, and (2) the skill and abilities of the person digitizing the road network. The use of scanning, character recognition, and raster-to-vector conversion technologies has aided in the task of converting hardcopy to digital data and has mitigated operator-introduced errors.

Route Calibration Using Control Points

Regardless of the source of the error, differences between the distances measured in the field and those calculated by the GIS software will make it difficult to precisely locate attributes and events referenced by an offset from a node or from the route's origin on a two-dimensional map. When the GIS lengths differ from the actual distances as measured in the field, events can "float" away from their actual linear location. The process of adjusting the two-dimensional GIS link lengths based on three-dimensional field measurements taken at control points (points at known distances along the route) is known as "route calibration."

This concept of "float" is illustrated by the following example. A crash occurs at milepoint 6.3 along a route, which is measured as being 6.5 mi (10.5 km) long based on accurate measurements in the field.

The road has a number of vertical and horizontal curves and was digitized using a 1:24,000-scale paper map or aerial photograph. Because of the difference between two-dimensional and three-dimensional distances (as described in figures 11 and 12), GIS only calculates a distance of 6.1 mi (9.8 km) for the route. When the GIS system tries to place the crash that occurred at milepoint 6.3 on a route that it believes is 6.1 mi (9.8 km) long, the crash "floats" off the end of the route and cannot be located (illustrated in figure 13).

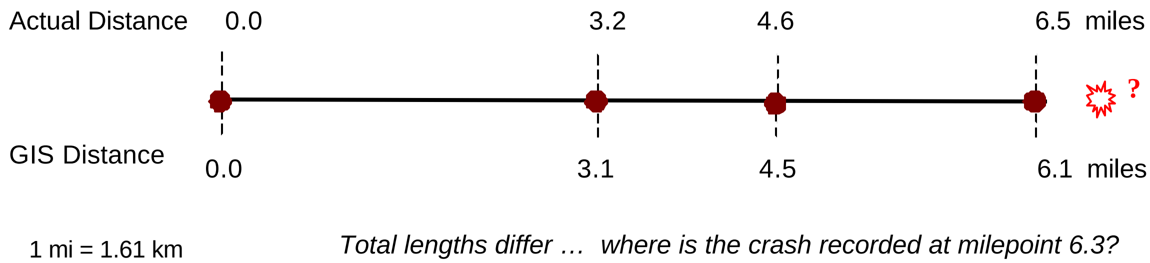


Figure 13. Non-calibrated roadway causes events to shift from their actual locations.

The process of calibration effectively shifts points referenced by an offset along a specified route closer to their actual location. Associating accurate cumulative distance measurements of known, observable features to points on the graphic representation in GIS causes GIS to "know" the three-dimensional mileage rather than simply its calculated two-dimensional measurements. These points in GIS of known three-dimensional cumulative measurements are known as "control points." Thus, a control point is one with a known set of coordinates and a known real-world distance from another control point (e.g., the beginning of a route or an intersection). The GIS software automatically shifts points in between control points, or intermediate points, proportionately to the shift of the node to its control point. This result of

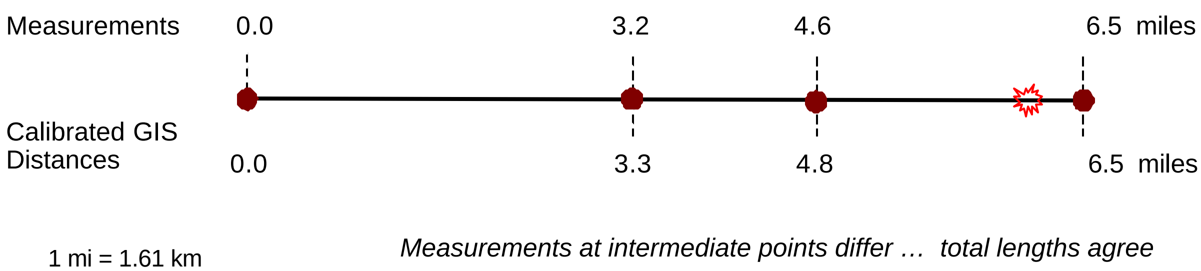


Figure 14. Calibrated GIS distances along section of roadway correct "float."

the calibration for the above example is illustrated in figure 14, assuming milepoints 0.0 and 6.5 were used as control points.

The more control points used in the calibration, the more accurate the GIS link lengths become and the more precisely event data can be located. GPS, combined with measured distances from vehicles, is making it possible to calibrate roadways in GIS at very short intervals, thus further removing the two-

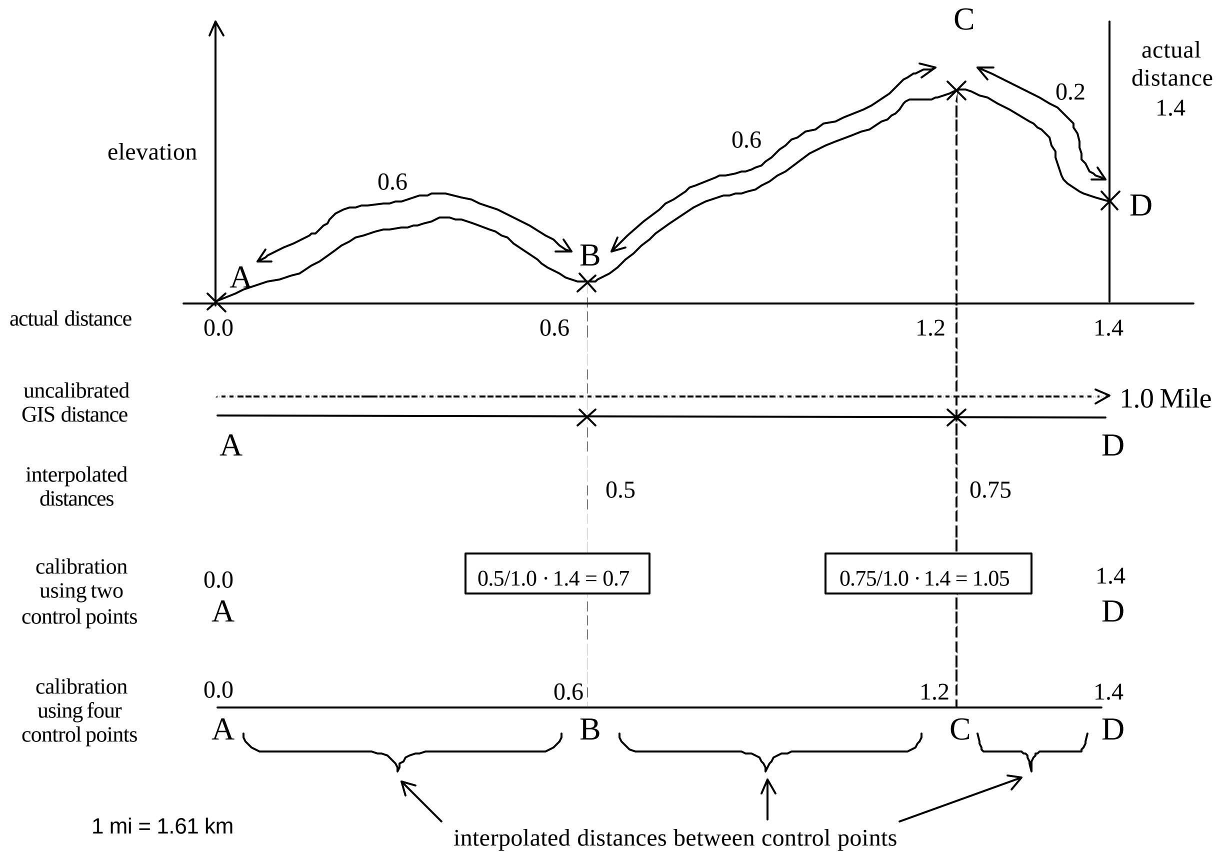


Figure 15. An increase in the number of control points increases relative accuracy.

dimensional versus three-dimensional distance problem. The following example (illustrated in figure 15) shows how adding control points between two existing control points on a particularly hilly segment of the roadway can dramatically improve the accuracy of the references along this stretch of road. The actual linear distance between Point A and Point D is 1.4 mi (2.3 km) as measured in the field. The distance as measured in GIS is only 1.0 mi (1.6 km). The difference of 0.4 mi (0.64 km) is caused by the inaccuracies of digitizing the map and by the accumulation of distance traveled going up and down the hills, which is lost in the two-dimensional representation of the road in GIS.

Without calibration, the GIS software interpolates between points A and D using the computed 1.0-mi (1.6-km) two-dimensional length of the section, placing Point B at milepoint 0.5 (actual distance 0.6) and Point C at milepoint 0.75 (actual distance 1.2). References along this roadway would be highly inaccurate, and the amount of error increases at points further down the road. Calibration using measured distances at points A and D would improve the accuracy of the intermediate references by adjusting the interpolated lengths based on the actual length of the segment by using the following formula:

$$d_i = D_i \cdot (d_{cp}/D_{cp})$$

where:

	d_i	= Calibrated distance for Point i.
	D_i	= Calibrated distance between control points.
	d_{cp}	= Measured GIS distance for Point i.
	D_{cp}	= Measured GIS distance between control points.

Using this equation and the calibration points A and D in the above example, the calibrated distance for points B and C would be calculated as follows:

$$d_B = 1.4 \cdot (0.5/1.0) = 0.7$$
$$d_C = 1.4 \cdot (0.75/1.0) = 1.05$$

In this case, the accuracy of Point B is not improved, but simply changed. It moved from an uncalibrated distance of 0.5 to 0.7, when the real-world value was 0.6. However, the accuracy of Point C was greatly improved. It changed from an uncalibrated distance of 0.75 to 1.05, with a real-world distance of 1.2. Taking accurately measured distances at points B and C (which have known coordinates) would enable these locations to be used as additional control points. Attributes or events specifically at these points would be located precisely (i.e., at milepoint 0.6 and milepoint 1.2, respectively). References between these new sets of control points would then be interpolated using the above formula and would be much more accurate when compared to using fewer or no control points.

Chapter 5. Integrating GIS and Safety Data

With a basic understanding of how roadway inventory data and crash data are traditionally referenced and how GIS interprets these data, one can turn to the integration of GIS and safety data for analysis. Linking highway safety data to GIS will provide challenges for State DOTs, MPOs, and other agencies. To make this integration a reality, three steps must be taken:

1. The LRS for the crash and roadway inventory data must be developed and made available for integration. In most cases, this development step has been completed by State DOTs in setting up their traditional systems, as previously described in chapter 3. Therefore, a good understanding of how the LRM has been implemented is necessary to plan for the development of an appropriate GIS that avoids linkage-related issues.

2. The spatial data model and a GIS route system must be developed. The GIS road network and the GIS route system are the foundation of GIS and are critical to the long-term success of any GIS for transportation applications.

3. The GIS is then used to process LRS data for display and spatial analysis.

This chapter discusses the technical issues associated with these steps as components of a GIS-based safety analysis system. Also provided is a discussion of other issues that must be considered when implementing GIS for safety analysis, including potential problems related to linkage of GIS and the LRS.

Planning for Integration

Perhaps the most critical step in developing a GIS safety analysis system is understanding the existing database design and planning for the development of the model that will integrate the newer technology into the older, well-established computing environments. The linkage between the existing LRS and GIS is dependent on several factors, including adhering to the naming convention and data type in use by the LRS, in particular for key attributes and data standards that may be in place for specific systems. Both of these issues are discussed below.

Attribute Coding Issues

Key field names are required to establish a database linkage for the crash and roadway inventory database. Small mistakes, such as improperly defining field names, data widths, or data type, could add unnecessary hurdles and delays in GIS development and linkage to linear referenced data. The key fields and items for linking the LRS to the GIS route system are Route ID and Route Measures for both point and linear event data (including crash and roadway inventory data). Other LRS items may also be key, depending on the scale and level of generalization of the LRS, such as with the implementation of multiple-route systems (refer to section at the end of this chapter).

For some agencies, a fully functional LRS linkage may also depend on additional key fields. A County-Route-Milepoint (CRM) is one example of an LRM common to some State DOTs, where an additional LRS attribute would have a "functional dependency" in the GIS LRS data model. In the CRM LRS, County (or jurisdiction) is a key attribute, because the route beginning mileage measurement is reinitiated for each county the route passes through. In this case, were GIS modeled using *only* Route ID and Milepoint, an incorrect linkage would probably occur. This is because without the use of the County attribute (where milepoint is measured independently for each county), the CRM data model would function like an RMP data model (where the route measurement runs continuously across the entire State). In such a case, several crashes occurring in *different* counties, but having the same route and mileage attributes, would probably be improperly mapped to the same point location, thus placing them in the wrong county.

Standardization Issues

Data standardization is a fundamental consideration in developing GIS for integration with existing databases. All working groups depend on standards being established within and outside of organizations and agencies to allow for cooperative efforts. Standards should be established for the LRS and GIS, and should address simple integration and processing of data within GIS. The DOTs have established the LRS based on standards that should include linear referenced data modeling, the data file naming convention, attribute coding, and the design of relational database tables. For placing crashes on the map, spatial data standards are less of an issue. However, the GIS route system standards should include spatial data modeling and considerations for scale, accuracy, resolution, and generalization. Standards for datum and projection mapping should also be considered.

The standards for hardware platforms, operating systems, network environments, database systems, and applications software are generally *not* an issue. The interoperability evidenced by the success of the Internet has proven this point. The standards for data definitions are much more important in order to provide reliability and portability in developing and maintaining systems and applications. The GIS software standards can also add to the complexity, since not all GIS share a common route system that is easily transferred from one vendor-specific application to another.

Developing the GIS Road Network Data Set

A GIS route system, based on a GIS road network data set, is required to display linear referenced data such as crashes or roadway inventory. Each route in the LRS coded in the GIS route system will be used as a reference for the display and analysis of the LRS data associated with that route. A GIS road network data set is produced from a transportation base map, which is developed through one of three means of digital data acquisition:

- Digitization of Source Materials C This method is a common, widely used, and well-tested means of data acquisition in which the road network line features, such as roadway centerline, are digitized from aerial photographs (considered to be primary source material) or from hardcopy maps (considered to be secondary source material). The process involves collecting the x-y coordinate values of the line features by tracing over each one using a digitizing tablet with a cursor or puck as the input device to locate and input map features into the computer. This type of manual production requires planning, source material preparation, and production setup, in addition to digital data post-processing. The costs for this type of data acquisition are significant and can represent the majority of system startup costs. Semi-automated methods using map scanning and line tracing technologies are being used to lower the cost and improve the accuracy of the digitization process.

- Acquiring Existing Digital Data From Other Sources C A cost-effective alternative to digitizing is to acquire digital data from a third-party source, such as USGS Digital Line Graphs (DLG). *Note: Large-scale 7.5 transportation overlay data from USGS may not be widely available for a given State and possibly may require updates to meet the completeness or accuracy standards for DOT use.*

- Directly Collecting Road Centerline Data Using GPS Technology C While this method is gaining popularity with DOTs, it is not widely used at this time due to some limitations in technology and an overall high cost for statewide coverage.

Correcting Data Topology

Road network data are often developed from a hardcopy map source or acquired from other sources of digital data and must undergo quality assurance and quality control procedures. To complete the development of the road network data, the spatial data must be cleaned up and edited to eliminate line overshoots, line undershoots, and to close all open polygons. Figure 16 depicts lines that overshoot or undershoot, and polygons that require node editing to close the links. The features represented in figure 16(a) may look like those of figure 16(b) when viewed on smaller scales. But in performing analyses, GIS can distinguish between differences of less than an inch in measured ground distance. Thus, these unedited links can cause significant GIS-T problems.

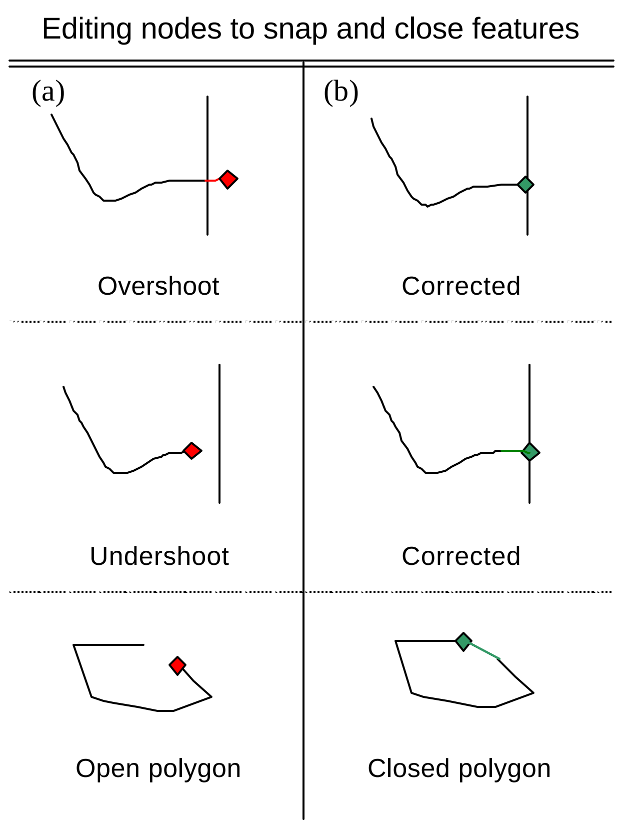


Figure 16. Line features require editing to correct for nodes that overshoot or undershoot, and polygons that do not close.

Special GIS tools are available to correct overshoots, undershoots, open polygons, and other types of topological problems. The connectivity problems illustrated above are cleaned up or corrected using system capabilities to properly connect the overshoot node to the neighboring line feature. When precision in placement along the neighboring line feature is required, the neighboring feature is split at the appropriate location and the dangle node is snapped to the newly created node feature.

Edge-Matching Map Sheets

Spatial data are often developed independently as map sheets. These map sheets must be post-processed or edge-matched to ensure that features on adjoining map sheets spatially match. Figure 17 illustrates how the edge-matching process adjusts features so that they are coincident with adjacent features that do not align on the adjoining map sheets.

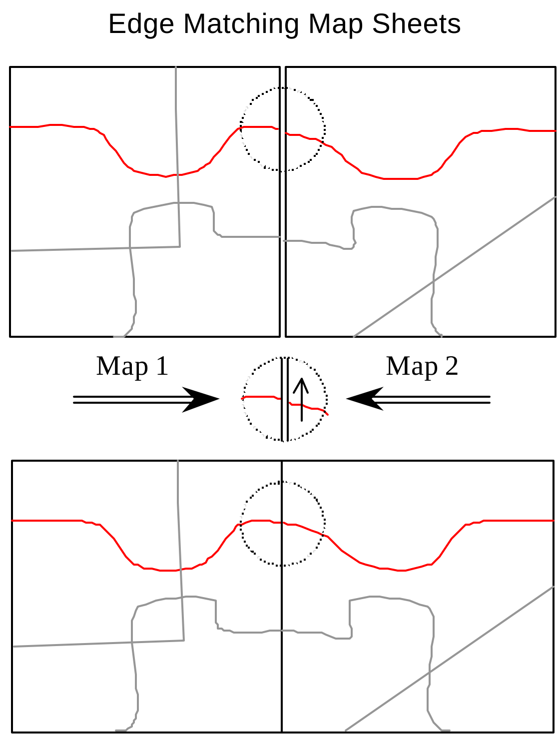


Figure 17. Features along the map sheet edge are aligned to match the location of an adjoining feature.

Creating Route Feature Data Types

Using available GIS tools, route features are created using the developed road network, following a two-stage process. First, the road network line feature elements are identified and coded as route feature elements. Then, route measures are added to the route feature at the beginning and end of the route and at additional locations along the route that will serve as control points. A simplified example of the process used by a safety engineer to create a GIS route system is as follows:

> *The safety engineer interested in linking an RMP LRS with the GIS road network database for safety analysis would require that the road network database have a route system representing the LRS. Given that the GIS route system had not been created, the safety engineer would go about that task using the GIS tools to create route features. Each route in the LRS that is of interest would first be identified. This could represent all routes for a DOT or only those represented in a particular study area. Next, using the tools available in GIS, the road network line features that represent each unique route are first selected to define the route features. Then, the appropriate Route ID is assigned to that route feature. This process would be repeated for each route.*
>
> *Next, the safety engineer might want to map crash locations using LRS data and the routes defined in GIS. Route measures in GIS will ensure good positional accuracy. However, route features will not contain measures until they are coded to the route. To*

complete the route development, the safety engineer would add route measures in the GIS route system. In the typical example, routes start at zero mileage and the length of each route is known and specified in the LRS. The route mileage of other locations, such as where routes intersect, may also be specified in the LRS. The safety engineer would first select a route, next select one of the known point locations, and then assign a correct route mileage to that location using the GIS tools for that purpose. Again, this process would be repeated for each route.

Figure 18 shows a route defined as a single feature composed of four sections defining the measures of the route. Routes do not have to coincide with the start and end of existing lines; they can be disjointed, but should not branch.

Use of Control Points

The process of adding route measures is often automated in GIS using control point data having route measure attributes. The control point data are determined from the LRS as the critical points on the road network, where linear control is required, then used to develop and maintain the GIS route system. For a route model that uses sections as a measurement control, each section's beginning location and ending location are coded with route measure values. For a route model that uses milepost markers, all milepost markers could be added to the GIS route system (as a route measure) to provide a highly accurate linear calibration of the GIS route system (see chapter 4 for a discussion of route calibration), although that additional effort is not necessarily required for GIS functionality.

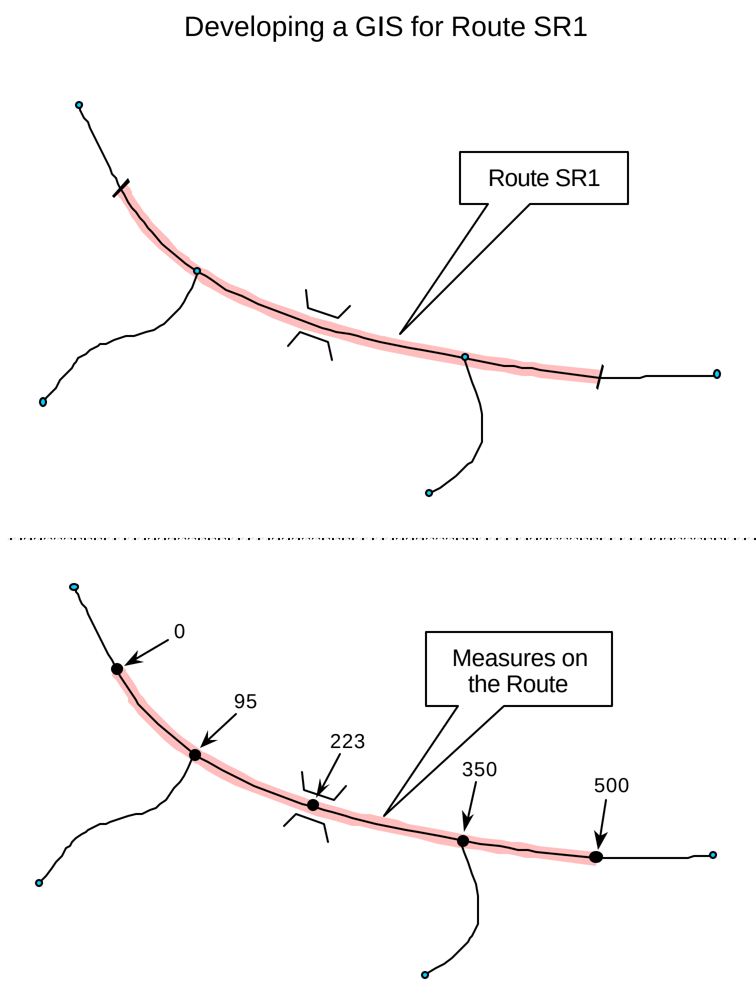


Figure 18. Routes are first defined in GIS and then a measuring system is defined along the route.

A GIS route system can be developed using only the beginning and ending route measures and can still support the capabilities that GIS has to offer. This is because GIS maps linear referenced data to a single route feature (the linear reference), relative to its linear measurements. This becomes significant for a point location of cross-streets (or other significant features). To accurately map the linear location of the linear referenced data of one route relative to a cross-street, the intersection of the two routes would have to be developed as a control-point location. This would ensure that events such as intersection-related crashes are mapped at that point location in GIS.

Processing the LRS Data Using the GIS Route System

Once the GIS route system has been developed with a linear location referencing data model, the LRS data (crashes and roadway inventory) can be displayed in GIS using the GIS capabilities and functionality. The spatial accuracy will depend on the spatial accuracy of the road network database and the linear accuracy will depend on the use of control points. This section provides a brief overview of how the GIS capabilities manage the LRS data.

With the GIS route system developed, crashes and roadway inventory can be displayed in GIS without having to perform further data conversion or data development as long as the route measures that have been applied to the GIS route system are inclusive of the measurements in the LRS data. This is accomplished by establishing a database connection and relating the linear referenced route attributes, found in the crashes and roadway inventory data sets, to the route and measure attributes in GIS. This linkage between the LRS and the GIS route system is established during dynamic segmentation (also known as Dyn Seg), which is a set of GIS tools and processes that permit linear referenced data to be placed along a measured line or route system, and spatial attributes to be derived from that location placement. Figure 19 illustrates segmentation by the placement of linear referenced data along a measured route. The LRS is what allows dynamic segmentation to take place in GIS.

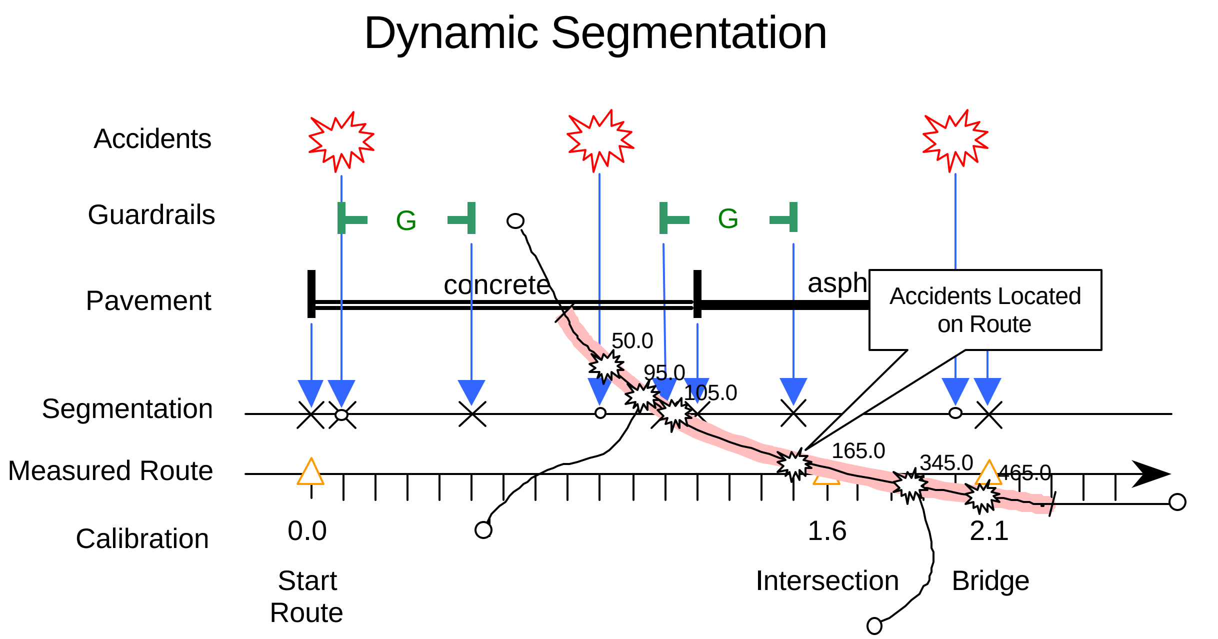


Figure 19. Using Dynamic Segmentation, point events and linear events are located along a measured line that has been calibrated at the measurements for an intersection and a bridge.

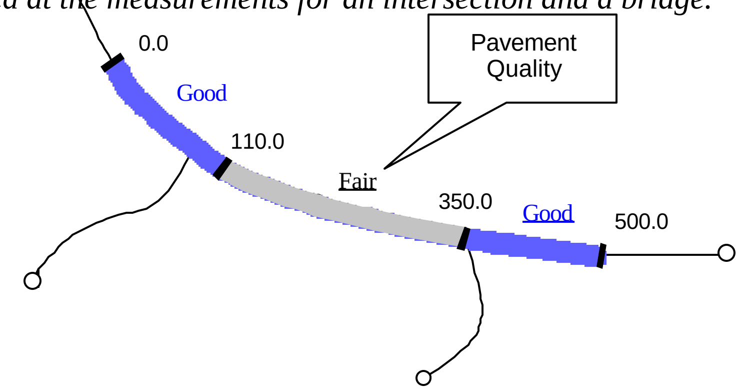


Figure 20. Both crash and pavement data are located on a route using Dynamic Segmentation.

Dynamic segmentation was implemented as a means of modeling linear features and point events independently of the route feature type. No longer was it necessary to statically store route information as line feature elements representing homogeneous sections. Rather, the LRS could be organized as database tables (similar to figure 9), and the routes and measures could be used to dynamically display the LRS as linear events along the route feature element. If the linear referenced data attributes happen to change in the LRS, GIS could redisplay the linear referenced data using the same route system without having to redevelop the GIS routes.

Although implementation of dynamic segmentation will vary by GIS vendor, GIS uses dynamic segmentation to locate and display linear features along a route and/or to segment the route itself. This definition of "dynamic segmentation" has taken on a generic meaning of locating linear event data along a measured route. In either case, dynamic segmentation is used in GIS to produce linear referenced data that can then either be displayed on a monitor or produced as hardcopy. These data may also be converted to a spatial data file. In other words, the linear referenced data can be mapped to geographic coordinates, and the coordinates and linear attributes can then be stored as a spatial data set. This process is done by interpolating the distance along the measured line of the GIS route from the beginning measure to the ending measure of the line. Figure 20 illustrates both crashes and roadway surface conditions located along a route using dynamic segmentation.

When each discrete event is located and displayed in real time, the process is said to be “dynamic.” That is, the GIS road network is used and segmented with the selected linear referenced data set. However, dynamic segmentation is often performed once in batch processing for reasons of system performance or to fix a spatial coordinate to an event for historical reference (see the section below on Historical Linear Reference). This is especially true for data warehouse applications where all events along the route system are used to segment the route for a transactional database.

The safety engineer can use the same road network to analyze crashes by any tabular attribute or data column (e.g., year, crash severity, etc.). Linear referenced data can be joined to create new linear referenced data sets and can be used for dynamic segmentation. In fact, it may be useful to join two or more linear referenced data sets, such as crashes and roadway inventory, for statistical analysis for thematic map display (color-coded by data values). However, there may be cases where the safety engineer would require more than one LRM or LRS. In such a case, additional consideration should be made for using GIS to bring together linear referenced data, which is also discussed below.

Other Considerations

Discussed below are several additional issues that will need to be addressed in order to properly develop and maintain a GIS-based safety analysis system.

Multiple LRMs

Some agencies may have more than one LRM. For example, an MPO may share its LN linear referenced data with a DOT that uses an RMP LRM for all State routes. Of the eight HSIS States, three currently use multiple LRMs. As a result, the DOT would have to support both LRMs to use the MPO data. Whether it is the case that a DOT has to support disparate LRS's, or that the DOT wants to utilize an externally supported LRS, the DOT will have two issues to consider when integrating the LRS into GIS. First, the two LRS's will have been developed from two differing road network databases. This means that the LRS data will have to be attached to a common road network to make it useable. One means of accomplishing this feat is to use geometrical and rotational transformation techniques to match and merge the LRS attributes developed in one GIS route system into the other GIS route system. This process is called “conflation.” Second, not all vendors have fully implemented all LRS types. This means that each LRM must be supported and implemented in GIS, or the DOT will be required to perform data conversion or develop custom GIS programming.

LRS Versus Coordinates

As noted above, dynamic segmentation is used to develop spatial coordinates of crashes and other linear referenced data. GPS technology is beginning to be used to assist in the crash data collection task by providing x-y coordinates of the crash site location, and will be an improvement far superior to most current collection methods for crash locations. However, it would be expected that when overlaid with linear referenced crash data, the LRS data would not align well with the GPS data due to the difference in the datum, or set of parameters and control points used to accurately define horizontal or vertical measurements. Data derived from different sources can be resolved for accurate display and meaningful analyses if the datum is known. It is suggested that metadata be available for all coordinate data and include projection, datum, and unit of measure information.

As GPS data are more widely used, the precision of the road network layer (developed from digitization or another non-GPS method) will be questioned relative to the precision of the GPS crash data. The solution for the road network data to spatially "fit" other data having a higher spatial precision, such as GPS data, is to conflate one data set to the other more precise data set. Conflation is used to rectify spatial accuracy between two data sets by adjusting all coordinates of the data points in the less precise data set to allow for a better match between selected data points and their more accurate locations. This process is also referred to as "rubber sheeting."

Another technique is to adjust the GPS data positional accuracy to the linear datum or snap the GPS data to the linear features in the GIS route system. Coordinate-based crash data derived from GPS or other sources, such as a different road network, will require adjustment to snap to the roadway as depicted in GIS. This will be expected for site location analysis mapped against the road network data. For States having GIS-located crash data, the buffering distance along routes, available in the GIS Safety Analysis Tools, will have to be considered to allow for the spatial margin of error in crash x-y placement relative to the GIS-defined roadway feature.

Address Geocoding

Crashes located by street address require a special set of GIS tools and a different GIS road network data. The process of linear location referencing by street address is called "Address Geocoding" or "Address Matching." GIS does this in a manner similar to dynamic segmentation (except not dynamically). First, all streets in the GIS road network database are attributed by street name. Then the GIS street network is further developed to include a beginning and ending address for each street block.

After the street network has been processed to contain beginning and ending address ranges for each street section, crash locations can be displayed using the GIS street network and the crash street address designation using GIS tools for address geocoding. The GIS tools do this by first parsing the address into its parts: number, street name, street type, etc., with each address part stored in the crash address data fields. GIS can then locate the crash by street name and interpolate the location of the street number as a distance along the street block using the street network beginning address and ending address. The spatial location along the road network is assigned to the street address and the coordinate values are used for display.

The numeric value of the address need not be an actual postal address, but merely a legitimate address value within a range of beginning and ending addresses assigned to each block face. Thus, this is not really a true address location in that each address does not have an accurate location associated with any cadastral survey or postal assignment.

Historical Linear Reference

Over time, roads change. New highways are built, roads are realigned, roads are abandoned, routes are renamed, and roadway inventory continually changes. Route identifiers and road measurements may change in the process, and the system that maintains this linear information would be updated accordingly. As changes in the LRS occur, changes in the spatial representation of linear features in the road network layer need to be updated also. Often this synchronization of databases requires an interdepartmental cooperative effort.

As early as 1985, the HSIS States have provided crash and other related data to HSIS. Each year's data set represents an annual snapshot of the linear representation and events for the State's roadways. The annual data sets are adjusted to correct for changes in linear measurements for that year.

There will always be uncertainty in spatial accuracy in locating linear events using a method that relies on a current GIS data set to map historical linear referenced data. The best way to initially locate historical data is to have a separate view of the LRS for each year of data, both in the linear referenced database and in the GIS roadway network. This approach would provide a snapshot of the LRS and would ensure complete and accurate LRS linkage of the linear referenced database with the GIS route system for that time period. This method assumes that the data model and all roadway realignments and other similar changes are fixed in the LRS for that period of time.

It becomes a challenge for agencies to develop procedures and methodology for GIS to adopt for the accurate representation of the road network over the life of the system. It may be that all historical data cannot be confidently located. However, the key is to plan for the future and use old data as well as one can. For those States that implement a data warehouse approach to their LRS or linear referenced data, spatially enabling the data warehouse will provide a solution to historical data reference by generating coordinate locations for linear referenced data within the data warehouse.

Chapter 6. Conclusions

This report was written to discuss GIS/safety integration in terms that can be understood by both safety engineers and GIS specialists, and to describe the issues and solutions involved in this type of systems integration. This report is intended to serve as an educational document for both audiences to initiate a common dialogue. Hopefully, the content of the report will begin to bridge the gap between the desire to implement GIS highway safety analysis within an organization and the development of a GIS-T infrastructure to support that effort. The specific topics discussed included:

- The benefits that GIS technology offers in general analyses, including display, spatial, and network evaluations, as well as cell-based modeling. The applications from the already-developed GIS Safety Analysis Tools are discussed as examples (see chapter 2).

- A description of how historical safety data (crashes and roadway inventory) are acquired, why such data are collected as linear referenced data, and how linear referenced data are different from spatial data. Definitions of common route systems are provided with illustrations to show how each is different (see chapter 3).

- General background information on Linear Location Referencing Systems (LLRS or LRS), which includes an explanation of routes and measures, common types of LRS's, how linear referencing methods (LRMs) are used to locate crashes and roadway inventory, and how GIS uses LRS's to locate linear features (see chapter 3).

- A general understanding of how GIS manages road network data, and how route features are developed using the road network feature data. The impact of resolution, scale, and route calibration are discussed as related to data accuracy (see chapter 4).

- A detailed discussion of the process of integrating GIS and safety data, including the need to plan for the integration and development of the GIS road network and route system, and processing the LRS data within GIS (see chapter 5).

References

1. Federal Highway Administration, *HSIS C Highway Safety Information System: The Essential Information System for Making Informed Decisions About Highway Safety*, Publication No. FHWA-RD-00-033, Washington, DC, 2000.

2. Federal Highway Administration, GIS Safety Analysis Tools, Version 2.0 (CD format), Washington, DC, July 2000.

3. Federal Highway Administration, *GIS-Based Crash Referencing and Analysis System*, HSIS Summary Report, Publication No. FHWA-RD-99-081, Washington, DC, February 1999.

4. Federal Highway Administration, *GIS Tools for Improving Pedestrian and Bike Safety*, TechBrief, Publication No. FHWA-RD-00-153, Washington, DC, July 2000.

5. Federal Highway Administration, *Using GIS in the Analysis of Truck Crashes*, HSIS Summary Report, Publication No. FHWA-RD-99-119, Washington, DC, June 1999.

6. Transportation Research Board, *Highway Location Reference Methods*, NCHRP Synthesis 21, Washington, DC, 1974.

7. National Map Accuracy Standards, http://www.usgs.gov/fact-sheet/map-accuracy/map-accuracy.html.

8. GIS/Trans, Ltd., *Linear Reference Practitioner's Handbook* (Draft), November 1997.

9. *TIDE: A GIS-Linked Data Warehousing Approach for Building an Integrated Transportation Information Environment*, American Association of State and Highway Transportation Offices, GIS-T '98, June 1998.

Appendix A. Case Studies

In an attempt to better understand some of the issues associated with integrating GIS and safety analysis, two case studies were conducted using data from the HSIS States of Maine and Washington. Spatial data sets were acquired from both States and were integrated with HSIS data currently maintained in the system. These two studies are described below and provide examples of the successes and problems of developing GIS linkage to an LRS.

Maine Case Study

In HSIS, crash, roadway, traffic volume, and interchange data files are maintained for Maine back to 1985, representing 22,000 roadway miles (35,405 roadway kilometers), and an average of 38,000 crashes per year. The Maine Department of Transportation (MeDOT) relies on the Transportation Integrated Network Information System (TINIS) to bring together data for crashes, roadway inventory, bridges, railroads, and project history/maintenance, and to support their LRS. Recently, MeDOT, with the assistance of GIS/Trans, Ltd., implemented the Transportation Information for Decision Enhancement (TIDE) system as a data warehouse to integrate their legacy systems with GIS and to augment the LRS to provide new system-wide access and capabilities. [9]

One of the many benefits of TIDE is in the area of historical data referencing. Using a process referred to by MeDOT as "static segmentation," the GIS coordinates for all data linked to the LRS are managed on a periodic weekly basis, such that any changes occurring in the LRS during that time period are reflected in GIS. For historical data referencing, this process addresses the issue of linear referencing and fulfills the department's goal of providing historical analysis capabilities for crashes. Thus, when road realignment takes place, the crash *will not* be imprecisely placed in an improper location along the new alignment, but will be located more accurately to a coordinate position that matches the location of the roadway at that point in time.

Maine uses a link-node (LN) system for their LRS, which means that the Link ID is a key variable for routes and is defined as a composite field made up of beginning Node ID plus ending Node ID. The system has been fairly stable in Maine, but over the years, new links were created that required additional Node IDs to be added. These additions also resulted in changing the four-digit Node ID number to five digits, which, in turn, increased the link number by two digits. All of these changes were implemented in TINIS. The TINIS data was then migrated to TIDE the source of the MeDOT GIS route system made available to HSIS. This seemingly small change to the Node ID number had a great impact on the ability to integrate HSIS data and the GIS data. These additional digits in the link numbers had not been changed within HSIS. In order to link the two systems, the Link IDs were changed in all 12 years of Maine data residing in HSIS. This problem clearly illustrated a key point that even with a well-managed GIS, such as the one Maine has, integration solutions will need to be found for existing incompatibilities.

In summary, the following conditions and situations, both advantageous and problematic, were encountered:

- Maine's data warehouse approach implemented in the TIDE system provided a very reliable and desirable approach to historical data referencing and mapping crash locations.

- The Maine road network and LN route system implemented in GIS provided a solid basis for mapping crashes and other linear referenced data.

- Changes for key linear feature data formats that had migrated to GIS had not been applied to existing Maine data in HSIS, which resulted in linkage problems until the HSIS data were brought up to the new Maine data standard.

- A business decision long established by Maine to round up the crash data linear location reference found a different solution in TIDE than the solution implemented years earlier in HSIS. HSIS had to be reconciled with the Maine DOT source data to permit complete mapping of available data.

- The Maine LN LRS spatial data were found not to adapt well with the GIS Safety Analysis Tools, which anticipates an RMP LRS route system.

Washington State Case Study

The Washington State DOT (WSDOT) relies on the Transportation Information and Planning Support (TRIPS) system to bring together data for crashes, roadway inventory, bridges, curve/grade/features data, roadway crossings, roadside facilities, special-use lane information, railroad grade crossing index, and traffic data to support their LRS. In HSIS, crash, roadway, traffic volume, curve/grade, and interchange data files are maintained for WSDOT for all years since 1993, representing 8,400 roadway miles (13,518 roadway kilometers), and an average of 35,000 crashes per year.

Washington State has a great investment in developing their GIS road network and route system data. They have developed GIS route systems on two scales of resolution 1:500,000 (good for small-scale mapping) and a higher resolution GIS road network based on 1:24,000-scale maps. The WSDOT GIS route systems contain route measures based on the TRIPS system's State Routes and Accumulated Route Mileage (ARM), a type of RMP LRS. Although both route systems contain the same Route ID and similar ARM values, they must be treated differently in the linkage with the LRS.

In developing GIS capabilities for use with Washington State HSIS data, the WSDOT GIS route system was used for linkage to HSIS. Working with the two GIS road networks available, it was easily discerned that road features are depicted differently, as would be expected. For example, at the 1:500,000 scale (small-scale mapping), a highway interchange containing ramps and collectors is generalized as a simple intersection of mainline routes. This generalization and reduction of detail is adequate for the mapping of mainline features and crashes, but the lack of ramp features degrades the accuracy of the GIS linkage to the LRS. At the 1:24,000 scale, the roadway has been modeled differently using additional details. Ramp features and divided roadways are present. This spatial data model represents a truer depiction of the roadway, where each lane of travel is represented as a route with an increasing or decreasing direction and ramps are represented as other routes.

The effort to map Washington State crashes and roadway inventory data (referred to as “WSDOT Roadlog” data) represented the first attempt, by anyone, to use the WSDOT GIS route data for that purpose. After a clear understanding of the linear referenced data model deployed by WSDOT, the linkage with the HSIS data was established for four years (1993-1996). As shown in table 2, the linkage with the roadway inventory data across all routes to each corresponding year of WSDOT 1:500,000-scale route data was achieved with an average success rate of 86 percent for all route types. When broken down by road type, a 98.9 percent success rate was achieved for linkage of mainline roadway inventory data to the 1:500,000-scale route data. This linkage would be equivalent to mapping to the road centerline. The difference between mapping all data and mapping mainline data is thought to be the result of ramps not having a representation in GIS at that scale.

Table 2. Summary of success in mapping Washington State Roadlog data to 1:500,000-scale route system.

All Data				Mainline Data			
Data Set	Roadlog	Miles	% Mapped	Data Set	Mainline	Miles	% Mapped
1993	8,583	7,314	85.2%	1993	7,265	7,207	99.2%
1994	8,659	7,321	84.5%	1994	7,265	7,209	99.2%
1995	8,352	7,317	87.6%	1995	7,265	7,204	99.2%
1996	8,397	7,240	86.2%	1996	7,265	7,122	98.0%
Average	8,498	7,298	85.9%	Average	7,265	7,186	98.9%

1 mi = 1.61 km

Washington State data mapped at the 1:24,000 scale presented several challenges in terms of geographic division of data and functional dependency. First, the WSDOT route systems were developed independently for the 39 counties in the State, which provided a technical challenge to working with HSIS data that are maintained on a statewide basis by year. Scripts had to be developed to handle the multiple-route systems and the geographic division of the data by county jurisdiction. Secondly, the large-scale mapping permitted greater feature resolution and less generalization, and depicted ramps and divided highways not shown in the smaller scale mapping.

As previously noted, the WSDOT TRIPS system uses ARM values for locating crashes and features. These values, computed from the State Route Milepost (SRMP) equations, contain measurements for increasing and decreasing directions on the roadway. For undivided highways, the ARM values would be the same, regardless of direction. But for divided highways, the increasing and decreasing side of the same route section can have different ARM values. To manage the differences in measurements for increasing and decreasing directions, WSDOT represented the TRIPS LRS data model in the GIS spatial data model by developing separate route systems within the same GIS one for increasing routes, a second for decreasing routes, and a third for ramps (each as separate route systems). This solution preserved the functional dependency inherent in the TRIPS data for direction of route measurement.

Unfortunately, the HSIS Washington State roadway inventory files do not contain a key variable for direction of mileposting on the State route, which separated features by increasing and decreasing the direction of travel. As a result, the GIS linkage of the Roadlog data with the 1:24,000-scale route system data model could not be achieved. Subsequently, comparisons between the two mapping scales for the Roadlog data could not be made.

For crash data, however, both scales could be linked. Taking advantage of a little-used crash data variable for direction of crash impact, crash locations were mapped with great success. The larger scale mapping allowed better accuracy in mapping events for divided roadways and interchanges by mapping crash data to the proper side of the roadway or to a specific ramp. The results for the 1:24,000-scale route data (see table 3) show that linkage with HSIS crash data was achieved for 97.1 percent of all available crashes and 98.9 percent of all mainline crashes. The 1.8 percent difference in mapping all crash data and mainline data is attributed to being able to accurately map crashes occurring on interchange ramps and couplets.

Table 3. Summary of success in mapping Washington State crash data to 1:24,000-scale route system.

All Data				Mainline Data			
Data Set	Records	Crashes	% Mapped	Data Set	Records	Crashes	% Mapped
1993	33,837	32,972	97.4%	1993	30,315	30,017	99.0%
1994	36,784	35,806	97.3%	1994	32,933	32,525	98.8%
1995	38,935	37,660	96.7%	1995	34,711	34,284	98.8%
1996	42,141	40,801	96.8%	1996	37,737	37,365	99.0%
Average	37,924	36,810	97.1%	Average	33,924	33,548	98.9%

For comparison, the use of the 1:500,000-scale model resulted in 89 percent of all crashes being properly linked (see table 4), which is 8 percent lower than achieved with the larger scale model. This lower value is due to the generalized representation of the roadway within GIS at this scale, where only mainline roadway features are represented, and ramp and collector features are not shown. Thus, one cannot map crashes to a roadway feature not depicted. This phenomenon was previously illustrated in figure 10 and described in chapter 4. Note that there was a very small increase (0.6 percent) in the mapping of mainline crash data for the 1:500,000-scale model over the 1:24,000-scale model. This increase is believed to be caused by the complexities and possible inaccuracies in the spatial data sets of the larger scale mapping.

Table 4. Summary of success in mapping Washington State crash data using the WSDOT 1:500,000-scale route system.

All Data				Mainline Data			
Data Set	Records	Crashes	% Mapped	Data Set	Records	Crashes	% Mapped
1993	33,837	30,214	89.3%	1993	30,315	30,214	99.7%
1994	36,784	32,741	89.0%	1994	32,933	32,741	99.4%
1995	38,935	34,525	88.7%	1995	34,711	34,525	99.5%
1996	42,141	37,588	89.2%	1996	37,737	37,588	99.6%
Average	37,924	33,767	89.0%	Average	33,924	33,767	99.5%

In summary, this case study highlighted the need for a complete understanding of the LRMs in use in order to develop GIS linkage to safety data for highway safety analysis. Below is a summary of the conditions and situations, both advantageous and problematic, that were encountered in the Washington State case study:

- The Washington State road network and ARM route system implemented in GIS provided a solid basis for mapping crashes and other linear referenced data.

- The quality of development and the completeness of the Washington State GIS provided a high degree of success (99 percent) in mapping crash data from mainline roads.

- To fully exploit the complexity of the Washington State GIS route system, a thorough understanding of the LRMs and the LRS was required.

- Newly discovered methods for using HSIS data for GIS integration exploiting key linear and spatial data field attributes provided opportunities and challenges in GIS development.

- GIS linkage of the HSIS Roadlog data with the WSDOT 1:24,000-scale spatial data could not be achieved due to a critical variable not being available in HSIS. This variable supported the functional dependency for direction of travel in Washington State roadway inventory data.

- The Washington State spatial data, developed using geographical coordinates and mileage as a route measurement, were found not to adapt well with the GIS Safety Analysis Tools, which anticipates State plane coordinates having units of measure in meters.

Glossary of Terms

Cartesian Coordinates:

A two-dimensional x-y location of a point on a plane (planar) in relation to two intersecting straight lines (axes). If the axes are perpendicular to each other, the coordinates are rectangular; if not, they are oblique. The x-axis measures the horizontal distance and the y-axis measures the vertical distance from the origin. An x-y coordinate defines every point on the plane. Relative measurement of distance, area, and direction are constant throughout the Cartesian coordinate plane.

Conflation:

A process by which two digital maps, usually of the same area at different points in time, or two different thematic maps of the same area, may be matched and merged into one through geometrical and rotational transformations. (Association for Geographic Information (AGI), the AGI GIS dictionary, http://www.agi.org.uk/pag-es/dict-ion/dict-agi.htm).

Coordinate:

Pairs of numbers expressing horizontal distances along orthogonal axes; alternatively, triplets of numbers measuring horizontal and vertical distances. Any of a set of numbers used in specifying the location of a point or position.

Coordinate System:

A framework used to define the position of a point, line, curve, or plane, and derivative map features within a two- or three-dimensional space. A reference system for defining points in space or on a particular surface by means of distances or angles, or both, with relation to designated map projection, datum, one or more standard parallels, and a central meridian.

Datum:

A set of parameters and control points used to accurately define the three-dimensional shape of the Earth (e.g., as an ellipsoid). The corresponding datum is the basis for a planar coordinate system. A reference surface for horizontal or vertical measurements.

A base reference level for the third dimension of elevation for the earth's surface. A datum can depend on the ellipsoid, the earth model, and the definition of sea level.

Divided Highways:

A divided highway is a roadway where the opposing directions are separated by a median that restricts movement between the two directional roadbeds. Note that some GIS installations consider highways to be divided only if the scale of the map and the size of the median are such that the two roadbeds can be mapped separately.

Dynamic Segmentation:

Dynamic segmentation of lineal spatial objects provides a means by which new point or line objects can be created by relating the distance-referenced attributes with a manageable set of distance-referenced linear objects. Dynamic segmentation removes the need for a set of spatial objects for each attribute. Spatial objects and distance referencing of routes are used to create attribute-based spatial objects as needed. A method of referencing attribute data on demand, based on variable segmentation of a single route or network structure.

Generalization:

A reduction of detail and a transformation of cartographic data into a representation at a reduced scale. The process of moving from one map scale to a smaller (less detailed) scale, changing the form of the features by simplification, etc.

Global Positioning System:

A satellite-based navigational system allowing the determination of any point on the earth's surface with a high degree of accuracy given a suitable GPS receiver. In the past the U.S. Department of Defense has intentionally degraded the accuracy of the satellite signal for non-U.S. military users. The error introduced into the signal is known as "selective availability." Error in the accuracy of GPS-derived positions can also be introduced through the nature of local conditions, for example, multipath. These errors can be greatly reduced using a technique known as "differential GPS." (Modified from the Association for Geographic Information (AGI), http://www.agi.org.uk/).

Linear Feature:

A geographic feature that can be represented by a line or set of lines. For example, rivers, roads, and electric and telecommunications networks can all be represented as linear features.

Linear Location Referencing Method:

A mechanism for finding and stating the location of an unknown point along a network by referencing it to a known point. All linear referencing methods consist of traversals and associated traversal reference points that together provide a set of known points, a metric, and a direction for referencing the locations of unknown points. No attributes are assigned to linear referencing methods.

Linear Location Referencing Systems:

The total set of procedures for determining and retaining a record of specific points along a linear feature. The system includes the location reference method(s), together with the procedures for storing, maintaining, and retrieving location information about points and segments on the highways.

Linear Referencing:

Process of identifying a location(s) on a network or specific link in a network by specifying a start position, direction, and distance.

Mileage (mileage measurement):

A given distance expressed in miles.

Milepoint:

The name given to the numerical value of the mileage displacement from a base point to any location.

Milepost (mileage marker):

One of a series of posts or markers set along a highway or other thoroughfare to indicate distance in miles. A physical entity, ordinarily a sign, placed beside a highway that contains a number that indicating the mileage to that point from some zero point on the highway.

Reference Markers:

Physical objects along roads that may or may not have a simple relationship to the length of roads and that form control points with a route and milepost measurement.

Reference Point:

A fixed identifiable feature, such as an intersection, railroad crossing, or bridge, from which a location can be measured or referenced.

Reference Post:

A physical entity, ordinarily a sign, placed beside a highway that contains a number that does not reflect a mile point (MP), but is an identification number for the location of the post. The identification number is associated with the actual MP of the location in office records.

Scale:

The proportion between two sets of dimensions.

In relation to maps, the best scale for your map depends on the resolution of the original data, as well as the level of detail you want your map to include. For example, 0.25 in^2 on a 1:250,000-scale map represents approximately 1.0 mi^2 (640 acres) on the ground. But 0.25 in^2 on a 1:63,360-scale map represents 0.25 mi^2 (160 acres).

State Plane Coordinate System (SPCS):

The plane-rectangular coordinate systems developed by the U.S. Coast and Geodetic Survey (now known as the National Geodetic Survey or NGS), one for each State in the United States, for use in defining positions of geodetic stations. Each State is covered by one or more zones, over each of which is placed a grid imposed upon a conformal map projection. Zones having limited north-south dimension and indefinite east-west extent have the Lambert conformal conic map projection with two standard parallels as the based for the State plane coordinate system. Zones in which this sequence is reversed (i.e., limited east-west dimension and indefinite north-south extent) have the transverse Mercator projection as the basis.

Traverse:

A method of surveying in which lengths and directions of lines between points on the earth are obtained by or from field measurements, and are used in determining the positions of the points.

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yards	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	millimeters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
NOTE: Volumes greater than 1000 l shall be shown in m^3.				
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact)		
EF	Fahrenheit temperature	5(F-32)/9 or (F-32)/1.8	Celsius temperature	EC
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.71	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE (exact)		
EC	Celsius temperature	1.8C + 32	Fahrenheit temperature	EF
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
fl	candela/m^2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

* SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380.

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

The Federal Highway Administration (FHWA) operates and maintains the Highway Safety Information System (HSIS) database. (1) The HSIS integrates police-reported crash data and roadway inventory and operations data already collected by eight States for the management of the highway system and it uses these data to study roadway and roadside safety issues. Recently, efforts have been made to expand the analytical features of HSIS by integrating Geographic Information Systems (GIS) capabilities. The GIS Safety Analysis Tools represent a recent example of the work in this arena to promote the use of GIS for highway safety analyses. (2) The original version of the tools was released in 1998 and provided practitioners with programs to perform spot/intersection analysis, cluster analysis, strip analysis, sliding-scale evaluations, and corridor analysis. (3) Version 2.0 was released in July 2000 and includes additional pedestrian and bicycle safety tools to select safe routes to schools, assess the bicycle compatibility of roadways, and define high pedestrian crash zones. (4)

One of the continuing goals of distributing the GIS Safety Analysis Tools is to encourage the safety engineers and others within State and municipal departments of transportation (DOTs) and metropolitan planning organizations (MPOs) to explore the capabilities of the GIS-based highway safety analysis tools and adapt those ideas and applications to fit their particular needs. However, due to the variety of implementations of GIS that exist within these organizations, developing capabilities in highway safety analysis requires an understanding of the requirements of GIS, Linear Referencing Systems (LRS), and GIS-based highway safety analysis applications.

The primary goal of this current effort was to discuss the integration of GIS and traditional safety data in terms that can be understood by both safety engineers and GIS specialists, and to describe issues and solutions involved in developing a GIS-based highway safety analysis system. To accomplish this goal, a survey of all eight HSIS States was conducted to assess their current GIS capabilities and to determine their methods for integrating GIS and their safety data. Subsequently, two States (Maine and Washington) were selected as case studies to more fully understand the intricacies associated with this type of integration.

This final report is intended to serve as an educational document for both safety engineers and GIS professionals and to initiate a common dialogue. Hopefully, this report will begin to bridge the gap between the desire to implement highway safety analysis within an organization and the development of a Geographic Information System – Transportation (GIS-T) infrastructure to support that effort. The report does so by providing the following:

- The benefits that GIS technology offers in general analyses, including display, spatial, and network evaluations, as well as cell-based modeling. The applications from the already-developed GIS Safety Analysis Tools are discussed as examples.

- A description of how historical safety data (crashes and roadway inventory) are acquired, why such data are collected as linear referenced data, and how linear referenced data are different from spatial data. Definitions of common route systems are provided along with illustrations to show how each is different.

- General background information on Linear Location Referencing Systems (LLRS or LRS), which includes an explanation of routes and their measures, common types of LRS, how linear referencing methods (LRMs) are used to locate crashes and roadway inventory, and how GIS uses LRS to locate linear features.

- A general understanding of how GIS manages road network data and how in GIS route data are different from road network data. The impact of resolution, scale, and route calibration is discussed as they relate to data accuracy.

- A detailed discussion of the process of integrating GIS and safety data, including the need to plan for the integration and development of the GIS road network and route system, and the processing of the LRS data within GIS.

Chapter 2. What GIS Has to Offer Safety Analysis

In recent years, many transportation departments, metropolitan planning organizations, and other related agencies have begun to use GIS for a variety of data management, systems management, and planning efforts, including:

- Pavement and bridge maintenance management.
- Modeling disaster response plans.
- Quantifying the potential impacts of transportation alternatives.
- Routing of overweight and oversized vehicles.
- Flood prediction.
- Risk assessment and risk management.
- Seismic slope-performance analysis and mapping of landslide hazard zones.
- Study of air emissions on health.
- Truck traffic analysis for the management of rural highway networks.

However, one area where GIS has not been extensively used is highway safety analysis. In part, this may be due to a lack of understanding of the potential benefits of such an application. Thus, prior to developing a GIS highway safety analysis system, there is a need to have a better understanding of what GIS is and how it can benefit traditional analyses. Provided in this chapter is information that will hopefully answer the following question:

> *What does GIS offer, in terms of capabilities and features, that improves upon traditional analytical techniques and should make one consider integrating GIS and safety data?*

The present-day benefits of GIS are well established in a number of disciplines. GIS provides the capability of storing and maintaining large data sets of spatial and tabular information. GIS has its strength in providing display and analytical capabilities that model the physical proximity of spatial features. One powerful aspect of GIS is the flexibility in modeling spatial objects to suit the particular needs of the user or application. These capabilities have been developed as the technology has matured. In its infancy, GIS provided rudimentary analysis capabilities for areas that were represented as discrete points distributed throughout a uniform grid. This type of analysis is referred to as “grid” or “cell-based” analysis.

GIS has since matured to include systems based on cartographic representation of points, lines, and area feature types. These systems provide a topological data model that allows for more robust analysis capabilities, referred to as “vector-based” analysis (e.g., point-in-polygon analysis or buffer analysis). Other common GIS capabilities include database integration, image overlay capabilities, and network analyses (e.g., shortest path routing). Over the past 10 years, GIS has adapted to accommodate linear referenced data. Crash and roadway inventory data are examples of this type of linear data and can now be brought into GIS for display and analysis. This capability offers the safety engineer specific analytical methods for understanding the spatial relationship of data that are not found in other information systems.

In addition, GIS offers a programming or scripting environment that allows the user to develop specific analysis programs or customize existing programs. All functions for display and analysis can be employed in a single-system design for Rapid Application Development (RAD) using common programming languages, such as Visual Basic, C++, and Java. This capability is evident in the GIS Safety Analysis Tools, which were developed in ArcView GIS using the Avenue scripting language. [(2)] More importantly, with recent developments in interoperability, GIS can be integrated into more mainstream enterprise applications, as well as web-based thin-client applications. Spatially enabling a website to include maps of high crash areas would be one example of the latter applications.

GIS provides the ability to display and view crash and roadway inventory location, and offers great rewards not available in a linear referencing system (LRS) alone. This capability is broader than simply mapping data and includes several types of analytical capabilities that can be broadly categorized into four groups:

- Display/Query analysis.
- Spatial analysis.
- Network analysis.
- Cell-based modeling.

The remainder of this chapter discusses each of these capabilities in more detail. Where appropriate, examples of existing applications (available on the GIS Safety Analysis Tools CD) are provided.

Display/Query Analysis

The primary appeal of GIS to many is the graphical capabilities. As it has been stated that “a picture is worth a thousand words.” Maps are the pictures GIS uses to communicate complex spatial relationships that the human eyes and mind are capable of understanding. The computer makes this possible, but still, it is the GIS user that determines what data and spatial relationships will be analyzed and portrayed, or how the data will be thematically presented to its intended audience.

Using the database capabilities of GIS, the safety engineer can query the database and have the results graphically displayed. This query analysis, when spoken in everyday conversation, takes on the form of a “show me” question, such as “Can you show me all head-on collisions that resulted in a fatality?” However, query analysis capabilities in GIS can also be exploited for other purposes, such as database automation, which might be used for error checking and quality control of coded data. As an example, the GIS roadway database could be queried automatically during the crash data entry process to verify the accuracy of speed limit and other crash report variables coded by an officer.

For linear referenced data to be displayed in GIS, it first must be integrated with spatial data. GIS can integrate spatial data of various scales, resolution, and projection, although use of spatial data integration warrants caution on inappropriate use. One example of poor use of GIS data integration capabilities would be statewide roadway feature data developed from 1:500,000-scale source maps. These data will not have the same line delineation and will not fit well or be appropriate for integration with data from large-scale sources (e.g., 1:24,000).

The use of imagery in GIS in conjunction with terrain modeling can provide a virtual reality display for highway safety analysis, giving the safety engineer a realistic view of the landscape (for instance, an aerial view of an intersection or a view of trees along the roadside). Satellite imagery and digital aerial photographs are two sources that can be used for this application. Both can be rectified, which involves image processing, such as rotating, scaling, and re-sampling. The imagery data can then be fit to overlay with the GIS spatial data (or linked to features), which involves determining the image map extent coordinates. Then, the imagery can be used for feature data collection or used as a backdrop image reference.

Data integration provides a microscopic level of analysis through the ability to spatially integrate and merge the data into a single view. Data not ordinarily used by the safety engineer, data that would otherwise be external to the LRS or not have a linear reference, such as demographic data, meteorological data, environmental data, economic data, and terrain data, to name a few, can be integrated using GIS. LRS data that is not ordinarily integrated, such as work-zone data, can also be integrated within GIS, thus expanding the data sources available to the safety engineer.

Thematic mapping of highway safety data provides a macroscopic level of analysis. Linear and spatial data integrated into GIS can be selected, differentiated by type or class, and displayed thematically. The safety engineer will be able to symbolize crashes for thematic mapping to distinguish between crashes, such as the severity of a crash resulting in fatalities and non-fatalities. These simple capabilities are the most commonly used to quickly digest large amounts of information, such as showing high crash locations or showing crash histories of road segments through the use of graduated line weight symbolization.

Spatial Analysis

Several analytical techniques, grouped under the general heading “overlay analysis,” are available in GIS for spatial analysis and data integration. GIS provides tools to combine data, identify overlaps across data, and join the attributes of data sets together using feature location and feature extent as the selection criteria. Overlay techniques will combine spatial data in other ways, such as features that can be combined to simply add one spatial data set to another, or to update or replace portions of one data set with another data set. Overlay analysis can be used to merge spatial data by combining two or more spatial data sets to produce a new spatial data set where the feature attributes are a union of the input data sets. As an example, the safety engineer can use these spatial techniques to combine demographic data, such as the number of households, showing the average number of school age children, with road segments having crash data showing pedestrian-related crashes, in order to derive risk factors for the total number of pedestrian-related crashes relative to the total number of school age children per road segment, for pedestrian-to-school safety analysis.

Proximity analysis is a type of GIS query capability and a category of spatial analysis that represents the fundamental difference of GIS from all other information systems. Buffering is a means of performing this practical spatial query to determine the proximity of neighboring features. In GIS, buffering will locate all features within a prescribed distance from a point, line, or area, such as determining the number of crashes that occurred within 800 m (0.5 mi) of an interchange, or locating secondary crashes that occurred within a certain distance and time (e.g., 400 m (0.25 mi) and 30 min) of other crash events, although reliability of these variables may not always support this example. Examples of proximity analysis applications on the GIS Safety Analysis Tools CD include Spot/Intersection Analysis, Strip Analysis, and Cluster Analysis.

The *Spot/Intersection Analysis* routine is used to evaluate crashes at a user-designated point or intersection for a given search radius. The spot or intersection of interest can be selected by clicking on the map using the mouse or by entering the intersecting route/street names. The end result of this analysis is a report that lists the number of crashes, fatalities, injuries, costs, etc. (as defined by the user) and a graphic that can be output as a hardcopy map (see figure 1) depicting the spot, search radius, and selected crashes.

The *Strip Analysis* routine is used to study crashes along a length of roadway rather than a finite location, spot, or intersection. The user must provide the section length to be used for the analysis as the program traverses the route (e.g., every 1.0 km) and the name/number of the route. The end result of this analysis is a report that lists the number of crashes and other user-defined attributes, and a graphic that

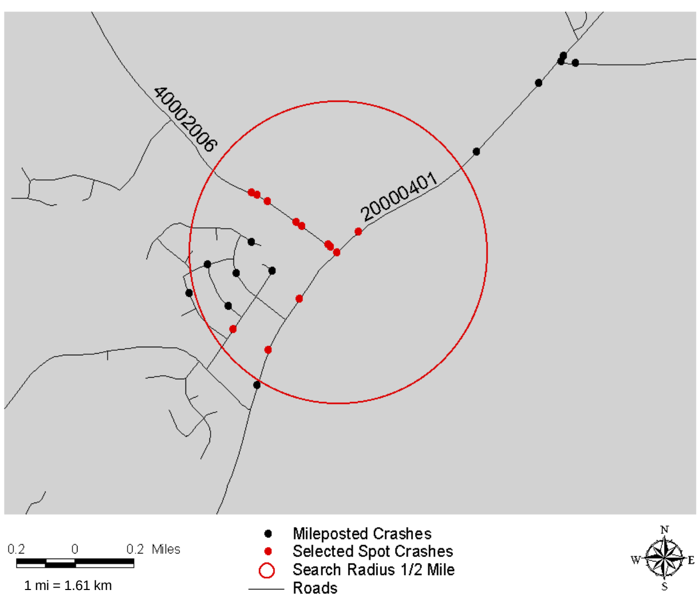


Figure 1. Results from the Spot/Intersection Analysis program can be graphically displayed as shown here.

can be output as a hardcopy map depicting the buffer that makes up the strip, selected crashes, and roadway identifiers.

The *Cluster Analysis* routine is used to study crashes clustered around a specific roadway feature, such as a bridge or railroad crossing. Crashes are identified that fall within a given distance on all selected routes. Again, the output is a report that lists various summary statistics selected by the user and a map depicting the high crash locations.

Network Analysis

Unlike proximity analysis that searches in all directions from a point, line, or area, network analysis is restricted to searching along a line, such as a route, or throughout a network of linear features, such as the road network. Network analysis can be used to define or identify route corridors and determine travel paths, travel distances, and response times. For example, network analysis may be used to assess the traffic volume impact of a road closure on adjacent roadways.

GIS networking capabilities can also be used for the selection of optimal paths or routes. The Safe Route

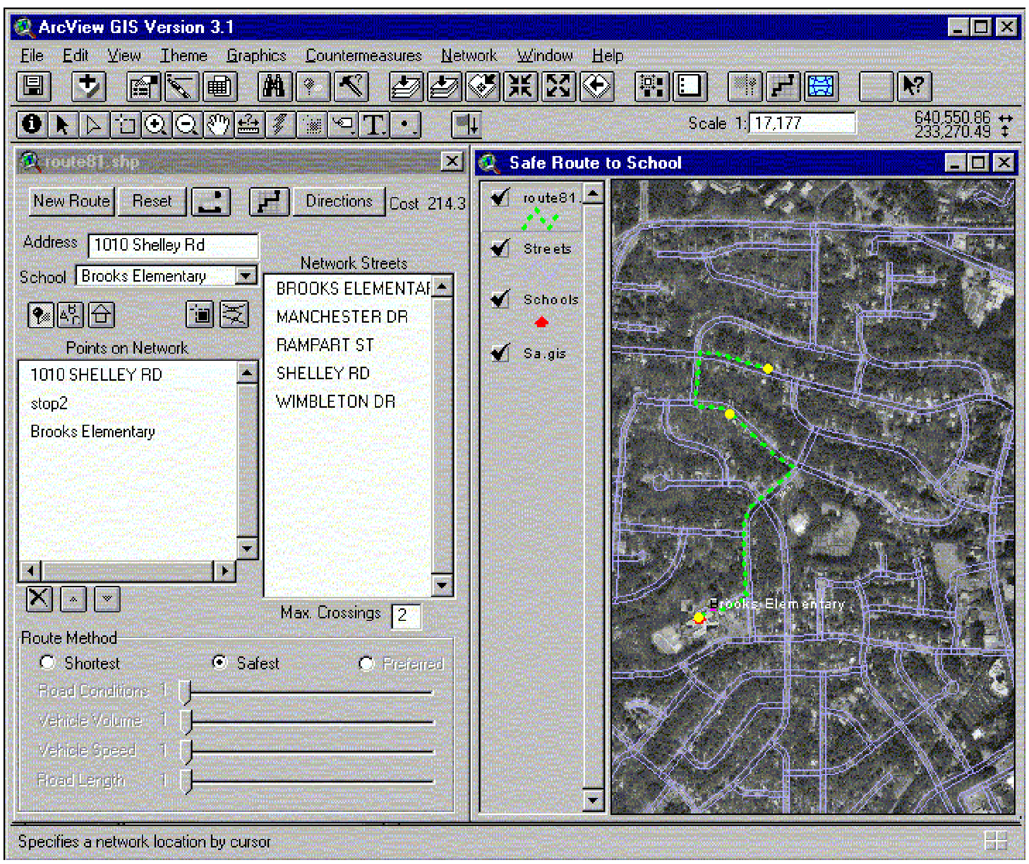


Figure 2. The Safe Route to School application selects the best route between an origin and a school based on roadway and traffic conditions.

to School application (see figure 2) on the GIS Safety Analysis Tools CD is an example of this type of application. The user inputs the origin and destination, and the program produces a map and walking directions for the preferred route, which is based on the level of hazard associated with the various roadway and traffic elements.

To improve the network model and provide the capability of automated route selection, the road network can be developed to include turning points, avoid improper turns onto one-way streets, represent posted traffic control restrictions, and include impedance factors to travel (such as mean travel speeds, number of travel lanes, and traffic volumes) to enhance the network analysis. *Note: Network routing capability is not available with all GIS, some GIS vendors offer network capabilities as an extension or additional modules to their software products at an additional cost.*

Other examples of network analysis tools that have been developed and are available on the GIS Safety Analysis Tools CD include the Sliding-Scale Analysis and Corridor Analysis programs. The *Sliding-Scale Analysis* routine is used to identify roadway segments with a high crash occurrence. This program differs from the Strip Analysis program in that the analysis segment is not fixed, but rather slides along the route in an incremental fashion. The user defines the segment length and the increment length for analysis. The end result of the analysis includes a table showing the high crash locations that exceeded a calculated or user-defined threshold, along with a variety of summary statistics and a map showing these locations.

The *Corridor Analysis* routine provides a visual means to locate high crash concentrations within a corridor. Using traditional methods, segments along a specific route could be examined (e.g., by using the sliding-scale analysis), but multiple routes within a corridor could not be easily linked and analyzed as a group. This program allows routes to be linked together in a manner that allows the analyst to assess the overall safety performance within a transportation corridor. In a recent evaluation, the program was used to examine truck crashes along designated truck corridors in a county in North Carolina. [5] In this case, State laws permit trucks to drive on any designated truck route and along any intersecting routes for a distance of up to 3 mi (4.8 km). The Corridor Analysis program was subsequently developed to identify truck crashes on roadways within the 3-mi (4.8-km) driveable zones. The output of the analysis included crash statistics and a variety of roadway characteristics for each high crash zone in the corridor. In addition, several plots depicting high crash segments and zones were also produced. The plot shown in figure 3 shows the high truck crash segments, including three such segments that were not on designated truck routes and were outside the 3-mi (4.8-km) driveable buffer.

Cell-Based Modeling

Cell-based modeling, also referred to as “grid-based” analysis, uses a grid or cells to aggregate spatial data for discrete distribution. In cell-based modeling, the spatial data are developed as tiles of a given dimension, or points of a uniform distribution, as defined by the user, for display and analysis. Cell-based modeling is effective in displaying patterns over larger areas, such as representing the sum total of crashes that are located within a cell. This capability provides a quick means to view spatial clustering of crash data. This technique is favored among DOTs and MPOs that assign crash data to street midpoints and

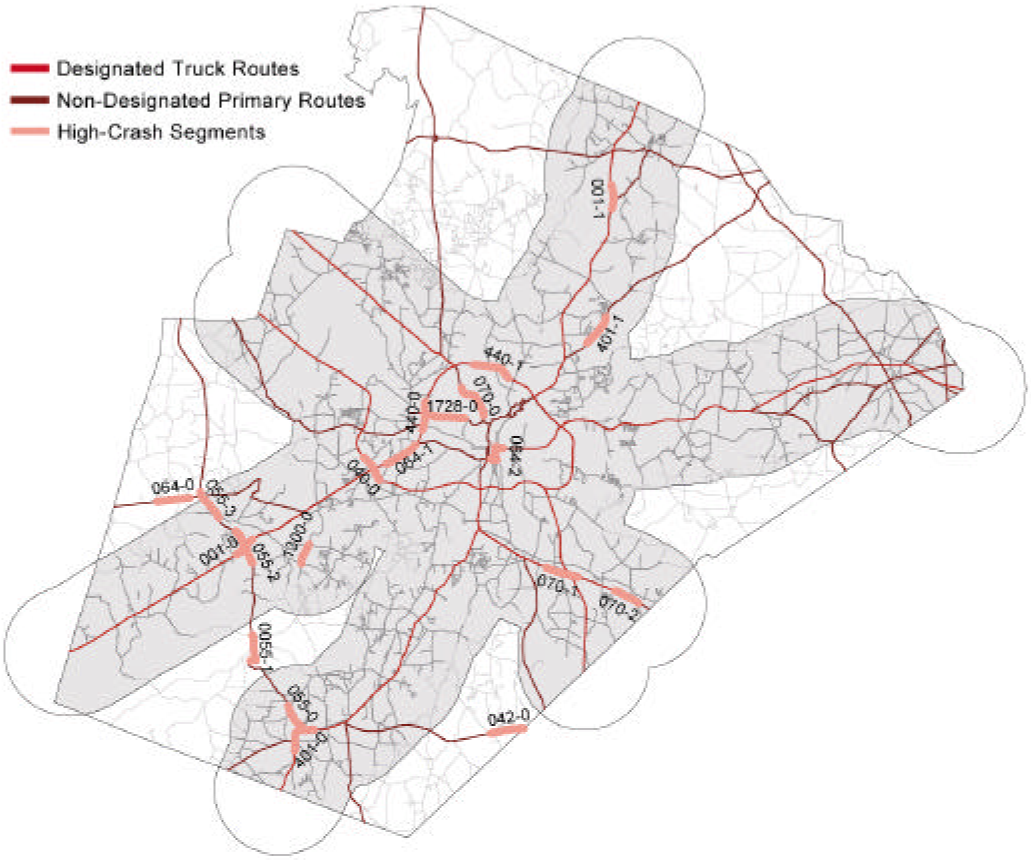


Figure 3. High truck crash segments that were identified using the Corridor Analysis program, including three segments outside of the designated truck zone.

street intersections, a method that in and of itself forms data clusters. Since cell-based modeling aggregates data at a specified grid resolution, it *would not* be appropriate for site-specific spatial analysis.

In cell-based modeling, special tools are available to merge grid data for overlay analysis. Cell-based overlay analysis is similar to the GIS overlay analysis previously discussed; however, the techniques and functions available in cell-based modeling are somewhat different. When the cells of different data sets have been developed using the same spatial dimensions, they can be merged on a cell-by-cell basis to produce a resulting data set. The functions and processes used in cell-based modeling to merge grid data are referred to as “map algebra,” because the grid data sets in cell-based modeling are merged using arithmetic and Boolean operators called “spatial operators.”

The High Pedestrian Crash Zone application on the GIS Safety Analysis Tools CD makes use of this technique. The program uses a discrete point file to calculate the density of selected crashes and generates a contour map identifying areas of high crash occurrence (see figure 4). Summary statistics of the various zones can also be produced in tabular or graphical formats.

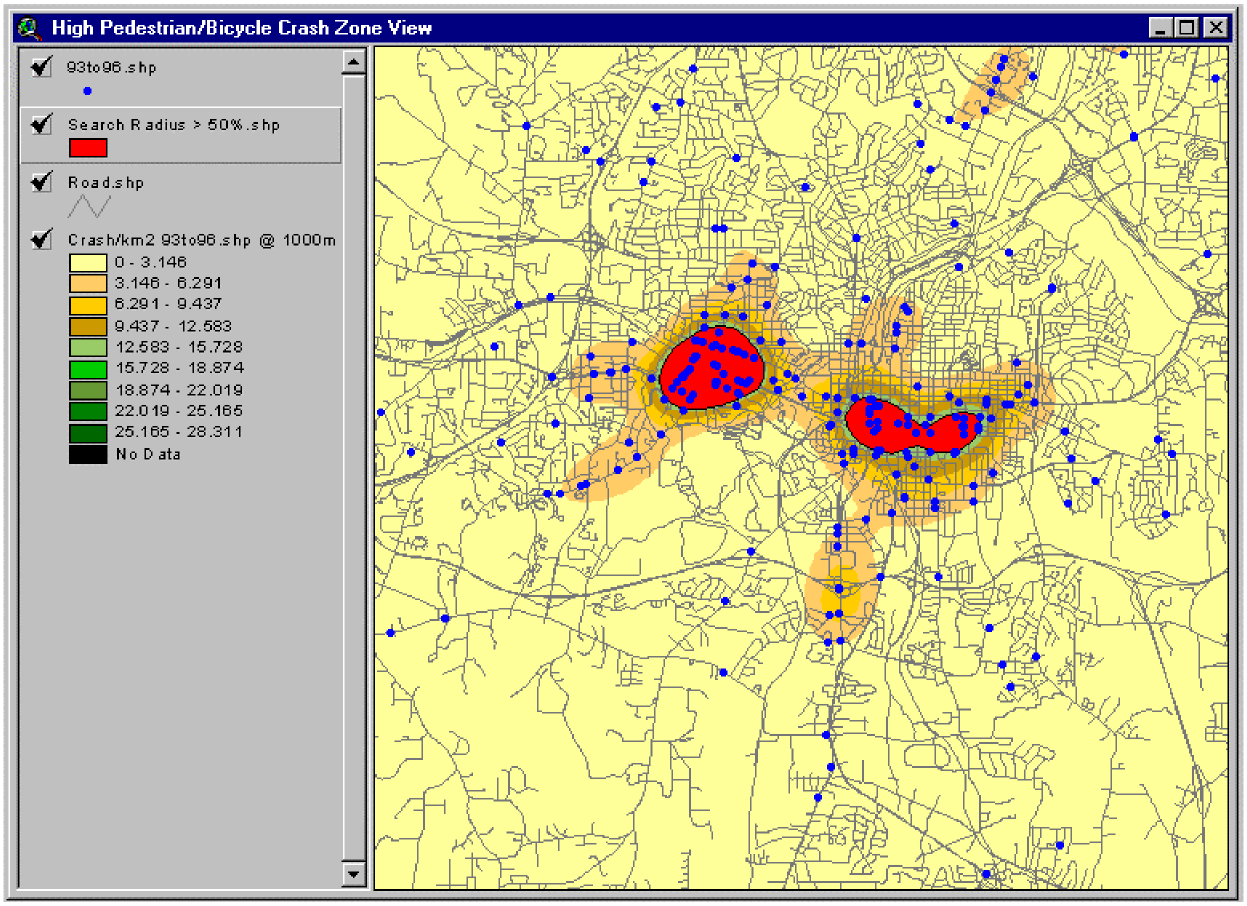


Figure 4. A view of crashes/km^2 grid with high crash zone of 50 percent and greater created in the High Pedestrian/Bicycle Crash Zone application.

Chapter 3. Understanding Traditional Data Collection Methods

Prior to integrating the GIS and safety analysis efforts, it is important to understand how data used in traditional safety analyses are collected and how GIS interprets and makes use of these data. This chapter provides an understanding of the former, while subsequent chapters explore GIS and data interpretation.

Locating crashes and roadway features is a process that traditionally has been accomplished using either references to the roadway or references to monuments along the roadway. This method is known as "linear referencing." Many different variations of linear location referencing systems (LLRS or LRS) have been defined and implemented by States and municipalities, each using various linear referencing methods (LRMs), and various designations and naming conventions. For clarification, the distinction between an LRS and an LRM is as follows:

> *Linear Location Referencing System (LLRS or LRS)* is the total set of procedures for determining and retaining a record of specific points along a [highway]. The system includes the location referencing method(s), together with the procedures for storing, maintaining, and retrieving location information about points and segments on the highways. [(6)]
>
> *Linear* Referencing *Method (LRM)* is the technique used to identify a specific point (location) or segment of highway, either in the field or in the office.

At times, the reference to the type of LRM or LRS is used interchangeably. However, it is important to recognize the difference when discussing route systems and to understand that the LRS is developed from the LRM.

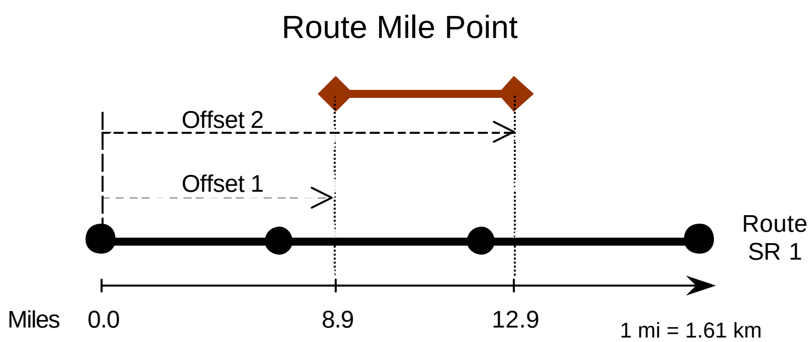
The most common location methods generally fall into one of five categories, with the last one being relatively new with the increasing use of global positioning system (GPS) technologies:

- Route-Milepost (RMP).
- Route-Reference Post (RRP).
- Link-Node (LN).
- Route-Street Reference (RSR).
- Geographic Coordinates.

> *Note: LRMs are supported in a variety of ways by the different GIS vendors. Not all of the GIS software products support all route systems, and the necessary functionality to support a particular route system may require development on the part of the user. This may be particularly true for the RRP or LN systems.*

Route-Milepost System

The Route-Milepost (RMP) system is, perhaps, the most common method used, particularly at the State DOT level. It is sometimes referred to as the "Route Mileage" system because mileage is typically the unit of measurement. In the RMP system, distance is measured from a given or known point, such as the route beginning or a jurisdictional boundary (e.g., a county line), to the referenced location. The distance is usually specified to the nearest hundredth of a mile, although some States may only specify crashes to the nearest tenth of a mile. The point of interest (i.e., crash or roadway feature) is always offset in a positive direction from the zero milepoint, and is not referenced to other intermediate points along the route. This point is illustrated in figure 5 using roadway surface condition as the roadway feature of interest.



Surface Condition Database			
Route	Begin MP	End MP	Condition
SR1	8.9	12.9	Good
...			

Figure 5. A roadway surface condition specified as "good," located along an RMP between mile point (MP) 8.9 and 12.9 as an offset from the route beginning, not referencing intermediate points in the LRM.

Route-Reference Post System

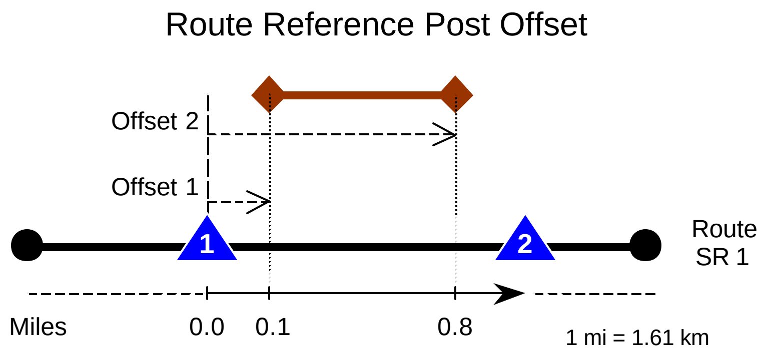
The Route-Reference Post (RRP) system is a method that uses signs posted in the field to indicate known locations. These signs, known as "reference posts," may or may not reflect mileposts. All crash and roadway feature data collected in the field are referenced to these markers in terms of distance and direction. These field-recorded events can later be converted to corresponding mileposts using cross-referencing tables and maps. The advantage of this system over an RMP system is the elimination of the problems caused by changes in route length that may be the result of realignment. Figure 6 illustrates the RRP LRM and uses roadway median type as an example.

Link Node System

In a Link-Node (LN) system, specific physical features, such as intersections, are identified as nodes. Each node is considered unique and is assigned a unique identifier or node number. Links are the logical connection between nodes and may vary in length. Links also have unique identifiers that are often derived from the associated pair of Node identification (ID) numbers. All crashes or roadway features are measured as an offset distance from the nearest or lowest node number along a link. Figure 7 illustrates the LN system and shows a schema for assigning Link IDs.

Route-Street Reference System

The Route-Street Reference (RSR) system is more commonly used in many municipalities and relies on the local system of streets to locate crashes and roadway features. In this system, an event is typically recorded as occurring on one street at a specified distance and direction from another street that is used as a reference. A variation of this system is the use of two reference streets and no distance measurement. For example, a crash may be coded as occurring on Street A between Streets 22 and 23. This option results in a loss of detail with regard to precise location, but still provides enough information to determine sections of roadway that may have a high number of crashes.



Median Database				
Route	Reference Post	Offset 1	Offset 2	Median Type
SR1	001	0.1	0.8	Barrier
...				

Figure 6. A roadway median type specified as a "barrier" is located along an RRP at 0.1 and 0.8 mi (0.16 and 1.29 km) offset from reference post 1, while other reference points in the LRM are not considered.

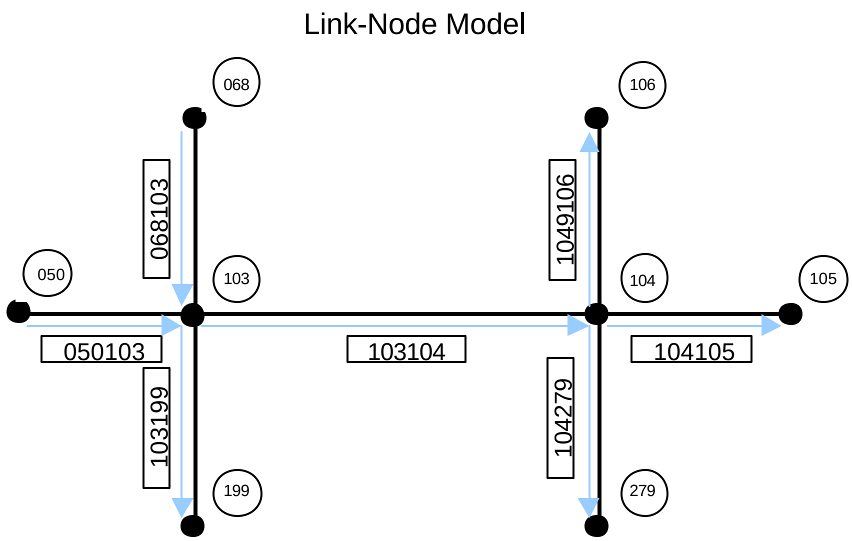
Geographic Coordinate System

Newer methods of reporting crash location information using GPS and other technologies are now available or are being developed. Unlike linear systems, coordinate systems use two or more spatial references that have equal significance. Cartesian coordinates use x and y (x-y) to measure distance along perpendicular axes of a coordinate plane. Geographic coordinates use latitude and longitude to measure distance in degrees along the axes of the sphere of the earth.

Crashes (and beginnings and endings of route segments) can also be located using GPS technology to reference, by latitude and longitude, a location on the earth's surface. Local transportation authorities may use State plane coordinates to measure (in meters or feet) the distance east and west or north and south along a State origin or datum.

Developing the LRS for Crash Data and Roadway Inventory Data

Regardless of which LRM is used, the procedures used by State and local DOTs to collect and process the crash and roadway inventory data are generally the same. A brief overview of these generic procedures and the resulting data formats that are available for safety analysis efforts are provided in this section. The example provided refers to an RMP system, but would be applicable to any LRM.



Route Section Database Table		
Link ID	Begin Node ID	End Node ID
050103	050	103
068103	068	103
103104	103	104
103199	103	199
104105	104	105
104106	104	106
104279	104	279

Figure 7. LN LRM showing links composed of node pairs, with each Link ID being unique and composed of unique Node IDs.

Collecting roadway inventory data, such as number of lanes, shoulder type, and pavement surface, is often accomplished in the field by driving along the roadway. As the inventory item is located, its attributes are recorded, along with the road name (or Route ID) and the mileage driven (or milepost). Mileage attributes for the various elements are generally recorded in one of two ways. For *point features*, such as a signpost or a culvert, a *single* mileage attribute is recorded. For an item located along a stretch of roadway, such as the number of lanes, shoulder type, or pavement surface, a *beginning* mileage and *ending* mileage is recorded. In GIS, the data and attributes associated with the LRS are known as “events,” i.e., point events or linear events. The result of this type of roadway inventory data collection can be represented graphically as shown in figure 8, where each inventory item along the route is associated with specific beginning and ending milepoints.

Most States collect and maintain attribute data on roadway characteristics as a single table containing records representing homogeneous sections of highway, such as represented in table 1, depicting pavement and shoulder type. This information may also be entered into a relational database

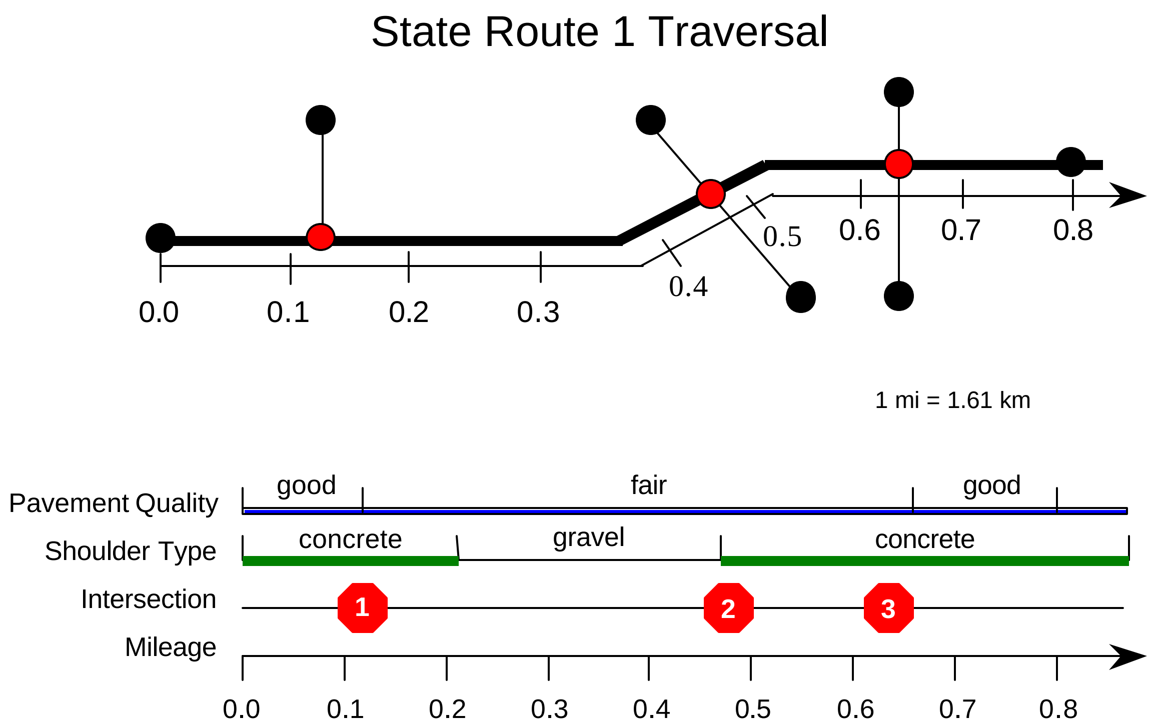


Figure 8. The traversal of State Route 1 graphically shows collection of pavement, shoulder type, and intersection items as the roadway is measured.

management system (RDBMS). Each record in the database would be entered for each observed and recorded occurrence. The attributes for roadway inventory would include Route ID, Mileage, and Inventory Type. Each data type could be entered into database tables, such as a Pavement File, Shoulder File, and Intersection File, as illustrated in figure 9.

Collecting crash location information is somewhat different in that no planning usually takes place to measure a crash location from the route beginning. Instead, crash location is usually measured from the nearest reference (e.g., milepost or intersection). However, crash locations are brought into the same LRS as roadway inventory through the coding process.

Table 1. Depiction of roadway characteristics showing pavement quality and shoulder type (each record represents a homogeneous section of highway).

Section File

Route ID	Beginning Mileage	Ending Mileage	Section ID	Pavement Quality	Shoulder Type
SR1	0.00	0.13	1	good	concrete
SR1	0.13	0.21	2	fair	concrete
SR1	0.21	0.46	3	fair	gravel
SR1	0.46	0.65	4	fair	concrete
...					

The officer at the scene of a crash usually cites observable features and states crash location as route, direction, and

LLRS as Database Tables

Pavement File

Route ID	Begin Mileage	End Mileage	Condition
SR1	0.00	0.13	Good
SR1	0.13	0.65	Fair
SR1	0.65	0.80	Good
...			

Shoulder File

Route ID	Begin Mileage	End Mileage	Type
SR1	0.00	0.21	Concrete
SR1	0.21	0.46	Gravel
SR1	0.46	0.90	Concrete
...			

Intersection File

Route ID	Mileage	Cross-Street Name
SR1	0.13	Main Street
SR1	0.47	Maple Street
SR1	0.64	State Route 5
...		

Figure 9. Partial illustration of a route system roadway inventory data model.

offset. Then a DOT "Coder" interprets the officer's location description and assigns a route code and mileage attributes to the crash location. An exception may be an MPO or urban area authority that might use street intersection coordinates or a street mid-point designation instead of a standard LRS. GPS use for crash location is also being used in some instances, but for the most part, it is not in widespread use by the enforcement community at the time of this writing.

The Coder puts the crash into the LRS by interpreting the location information from the crash report and determining or interpolating a precise linear location. Coders rely on additional information sources, such as roadlog reports that provide a listing of route mileage for cross-streets, roadside features, etc., and assist with correctly locating a crash. Crash attributes would include Crash Case Number, Route ID, and Mileage, and would be entered as crash records within the LRS. For example, a crash report may describe the location for Crash Case No. 2000-0954 as "Interstate 65, 50 ft north of intersection with U.S. 10." The Coder may translate this information into a linear location crash event as "Case No. 2000-0954, Route ID I-65, Milepoint 2.71."

Chapter 4. Understanding the Roadway Within GIS

In GIS, the roadway is represented as a collection of lines with endpoints defined in coordinate space. A combined collection of graphical links form a roadway network, but this representation alone is considered as having no “intelligence.” That is, connectivity and designated route topology are not present. “Routes” are special feature types constructed from the roadway line features (e.g., route number) and can be designated in GIS using relational database tables to identify those lines that make up each route. Routes also have a location method associated with them that allows event locations, such as a crash location milepoint, to be positioned on the route. To implement this capability in GIS, route “measures” are assigned as attributes to the route at the starting, ending, and intermediate points along the route. The intermediate route measures are used to control location placement accuracy for events along the route (see route calibration discussion later in this chapter).

> *The development of routes in GIS varies by vendor and available GIS software. In general though, routes can be: (1) wholly or partially coincident with other routes, (2) disjointed or disconnected, and (3) defined with sections containing route measure attributes.*

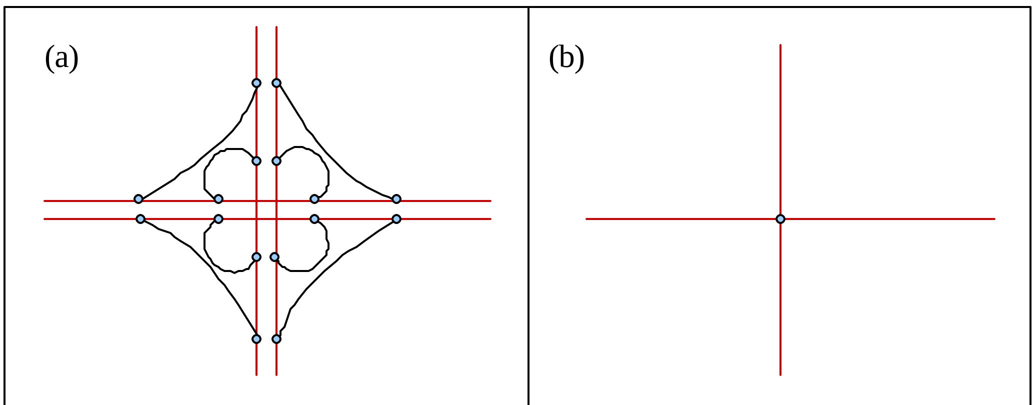
Resolution and Generalization

GIS has historically relied on a cartographic data model, similar to Computer-Aided Design (CAD) systems, to represent roadway and other feature elements. Like CAD, GIS uses a coordinate system to store and display primitive feature elements of points, lines, and areas. In the cartographic data model, the roadway is represented as a “line” feature.

Unlike purely graphical software applications however, GIS builds and manages topology in the cartographic model. A “road network” is the connection of a series of roadway “line” features having the same defined attributes. This interconnectivity of line features is important for using routing applications or in network data modeling. The GIS network applications may be used for optimal routing analysis to find efficient travel routes, closest facility analysis to determine which roadway or other facility is closest, or service area analysis to learn what is near a particular site. Appropriate connectivity and related information can be designed and built into the GIS data to develop the network data model for the support of network analysis.

Modeling the road network as a spatial or graphical layer in GIS is a planning exercise that needs to be compatible with and reflect the needs and requirements of the DOT, as it might support the daily operations of the organization. Mainline, secondary routes, collectors, and interchange features can be represented in GIS at various levels of detail. In the discussion that follows, it is important to understand the difference between small-scale and large-scale mapping. The smaller the scale, which is represented by a larger number in the ratio (e.g., 1:500,000), the less detail that can be represented. Small-scale mapping of State-maintained roads could be represented as a simple roadway or right-of-way centerline. This depiction would show intersections of mainline and secondary roads as a point, thereby showing no interchange features, such as ramps. On a moderately large mapping scale, such as 1:24,000, roadway features could be resolved at a spatial accuracy of ±40 ft (12.2 m) to fully depict the road network for all directions of travel, showing ramps and collectors. On this mapping scale, lanes of travel, i.e., lanes for the same direction of travel, would be generalized to a single pavement centerline and would be sufficient for most LRS application needs. On much larger scales, such as 1:600 (1 in = 50 ft), features such as pavement markings, actual lane designations, and specific design elements could be graphically depicted. The latter is the level of resolution often used in roadway design work.

Consider, as an example, how an interstate interchange, having roadway mainline and collector features, would be depicted on a map. On a larger scale mapping (e.g., 1:24,000), each feature of the interchange (e.g., the mainline roadways, ramps, and intersecting collectors) would be depicted in GIS as a separate line feature, as illustrated in figure 10(a). With small-scale mapping, typically 1:500,000 or smaller, the GIS cartographic data model *would not* support the depiction of ramps and collectors, and the mainline roadway features would be generalized to a centerline representation. At this level of generalization, the interchange would be represented as a single-point feature, as illustrated in figure 10(b).



(a)	(b)
Depiction of interchange generalized as large-scale representation of mainline travel lanes and collectors, representing greater detail.	Depiction of interchange generalized as small-scale representation of roadway centerline intersection, having a reduced level of detail.

Figure 10. Same roadway interchange represented in GIS at two levels of generalization and detail.

Scale and Accuracy

In GIS, scale and accuracy are important considerations, but often these aspects of data collection are overrated when dealing with routes and highway safety analysis. Most site location analyses can be performed with nearly any scale mapping. This should not be misunderstood to mean that knowledge of the scale and accuracy of the base map or linear referenced data is not important. Generally, the source material and the standards of data development determine both the scale and precision of geospatial data sets. As an example, the U.S. Geological Survey (USGS) publishes accuracy standards for Digital Line Graph (DLG) data, such as follows:

As applied to the U.S. Geological Survey 7.5-minute quadrangle topographic map, the horizontal accuracy standard requires that the positions of 90 percent of all points tested must be accurate within one-fiftieth of an inch (0.05 cm) on the map. At 1:24,000 scale, one-fiftieth of an inch is 40 ft (12.2 m). [(7)]

For highway safety analysis, GIS brings together data from various sources – the GIS roadway network, the LRS crash database, and the LRS roadway inventory data. When merging data from different sources, the least accurate data source (i.e., least common denominator) is used for determining overall data accuracy. In many cases, the location of the crash within the LRS crash database will be that least common denominator. Generally, a crash is recorded by the police officer and subsequently coded by an analyst to within 0.01 mi (0.016 km) or approximately 50 ft (15.2 m). Thus, for highway safety analysis, the USGS standard of ±40 ft (12.2 m) would be considered acceptable.

The accuracy of linear referenced data is relative to, and thus mostly dependent on, the calibration of route measures along the road network, and less dependent on the accuracy of the road network data in GIS. However, the accuracy of spatial data will come into play in two ways: (1) when the road network data is overlaid with other spatial data, and (2) as the LRS is linked with the GIS route system through dynamic segmentation (the latter being the degree to which calibration needs to be performed to improve placement of crashes given the spatial resolution of the road network data set).

Another important consideration in positional accuracy is the distinction between locations referenced in different dimensions, i.e., a location referenced relative to the linear distance versus x-y coordinate space versus x-y-z spherical space. Consider the LRM that uses the RMP system as an example. Measurements are taken from the beginning of the route (or perhaps from the beginning of the route in each county) and are used to specify the offset of a feature or event along that route. Since only the length of the roadway geometry is taken into consideration, these offsets are accurate in only one dimension (i.e., linear accuracy). The linear distance for any given point along the route has accounted for the roadway curvature and grade since it essentially represents the driving distance on the roadway.

In the GIS software, roads and other features are referenced using a minimum of two dimensions. In two dimensions, roads appear as if they were in plan view (i.e., being seen from above). Curves, turns, tangent sections, and intersections appear as they would on an aerial photograph or map, i.e., an orthogonal view. Curves and turns in a road obviously impact the measured distance between two nodes, as illustrated in figure 11.

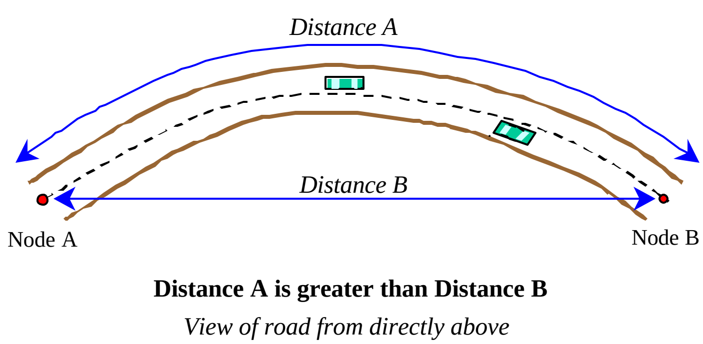
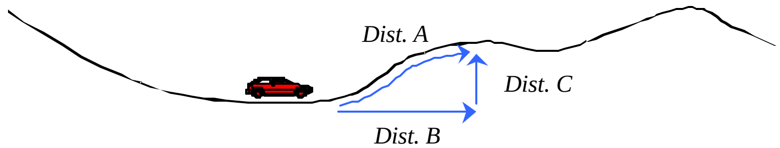


Figure 11. Location references in two dimensions.

The real world, however, is three-dimensional. In three dimensions, topography also affects the distance along a roadway the same way horizontal curves do in two dimensions (see figure 12). The measurements taken by field personnel are obviously made in the real world and, therefore, accurately record distances as they are measured along both horizontal and vertical curves. These measurements reflect the geographic accuracy of a roadway and are the distances used to reference features in an LRS. It is important that GIS properly reflect this level of geographic accuracy.



Distance A is greater than the two dimensional Distance B because of vertical component C

Cross Sectional View of Roadway

Figure 12. Location references in three dimensions.

While some of the GIS systems accurately capture data in three dimensions, most do not. This creates a problem when comparing distances calculated in a two-dimensional GIS with distances measured in the three-dimensional real world. However, this problem is fairly insignificant in most transportation applications, since the slope of most roadways has a minimal effect on distance. For example, on a 10 percent slope, the difference between horizontal and surface distance is just 0.5 percent. [8] This problem will, however, be compounded further down the length of a route, especially on roadways in mountainous terrain.

Another source of error that should be noted with regard to the length accuracy of the GIS links appears when digitizing road features from a paper map or aerial photograph. Most base maps are created from two-dimensional maps, and precision in the road network database is determined by two factors: (1) the scale of the source data, and (2) the skill and abilities of the person digitizing the road network. The use of scanning, character recognition, and raster-to-vector conversion technologies has aided in the task of converting hardcopy to digital data and has mitigated operator-introduced errors.

Route Calibration Using Control Points

Regardless of the source of the error, differences between the distances measured in the field and those calculated by the GIS software will make it difficult to precisely locate attributes and events referenced by an offset from a node or from the route's origin on a two-dimensional map. When the GIS lengths differ from the actual distances as measured in the field, events can "float" away from their actual linear location. The process of adjusting the two-dimensional GIS link lengths based on three-dimensional field measurements taken at control points (points at known distances along the route) is known as "route calibration."

This concept of "float" is illustrated by the following example. A crash occurs at milepoint 6.3 along a route, which is measured as being 6.5 mi (10.5 km) long based on accurate measurements in the field. The road has a number of vertical and horizontal curves and was digitized using a 1:24,000-scale paper map or aerial photograph. Because of the difference between two-dimensional and three-dimensional distances (as described in figures 11 and 12), GIS only calculates a distance of 6.1 mi (9.8 km) for the route. When the GIS system tries to place the crash that occurred at milepoint 6.3 on a route that it believes is 6.1 mi (9.8 km) long, the crash "floats" off the end of the route and cannot be located (illustrated in figure 13).

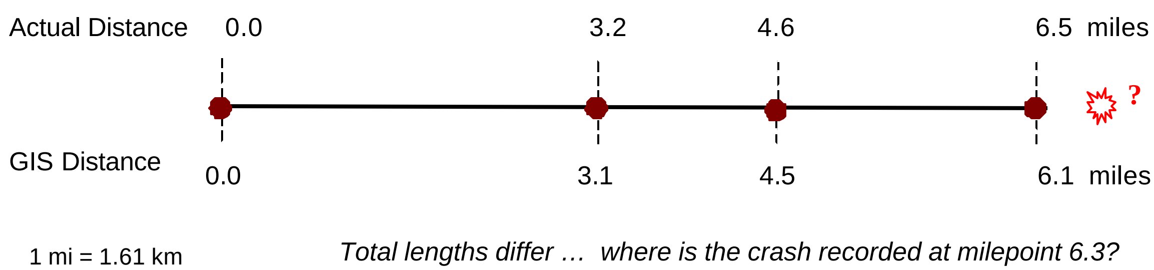


Figure 13. Non-calibrated roadway causes events to shift from their actual locations.

The process of calibration effectively shifts points referenced by an offset along a specified route closer to their actual location. Associating accurate cumulative distance measurements of known, observable features to points on the graphic representation in GIS causes GIS to "know" the three-dimensional mileage rather than simply its calculated two-dimensional measurements. These points in GIS of known three-dimensional cumulative measurements are known as "control points." Thus, a control point is one with a known set of coordinates and a known real-world distance from another control point (e.g., the beginning of a route or an intersection). The GIS software automatically shifts points in between control

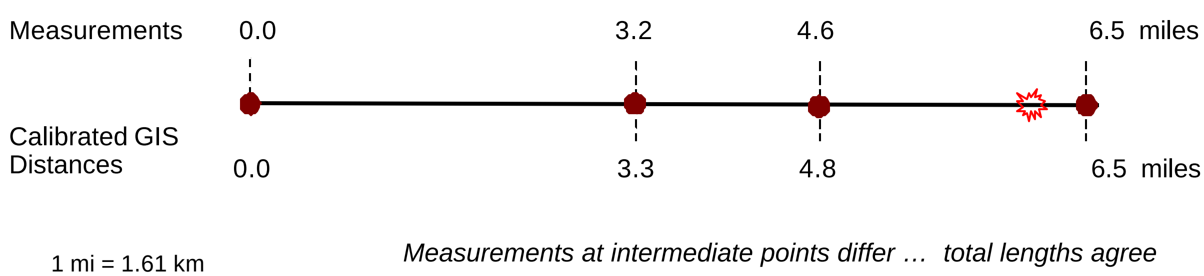


Figure 14. Calibrated GIS distances along section of roadway correct "float."

points, or intermediate points, proportionately to the shift of the node to its control point. This result of the calibration for the above example is illustrated in figure 14, assuming milepoints 0.0 and 6.5 were used as control points.

The more control points used in the calibration, the more accurate the GIS link lengths become and the more precisely event data can be located. GPS, combined with measured distances from vehicles, is

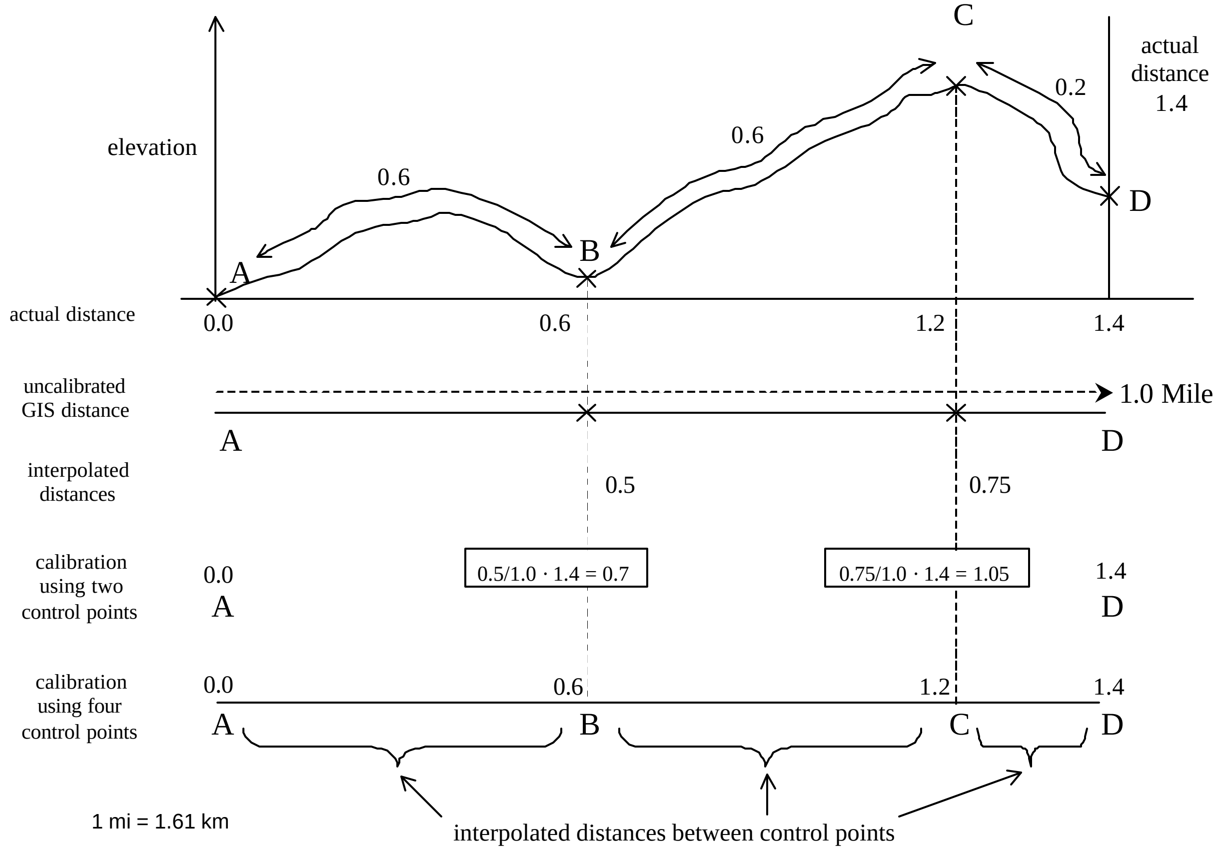


Figure 15. An increase in the number of control points increases relative accuracy.

making it possible to calibrate roadways in GIS at very short intervals, thus further removing the two-dimensional versus three-dimensional distance problem. The following example (illustrated in figure 15) shows how adding control points between two existing control points on a particularly hilly segment of the roadway can dramatically improve the accuracy of the references along this stretch of road. The actual linear distance between Point A and Point D is 1.4 mi (2.3 km) as measured in the field. The distance as measured in GIS is only 1.0 mi (1.6 km). The difference of 0.4 mi (0.64 km) is caused by the inaccuracies of digitizing the map and by the accumulation of distance traveled going up and down the hills, which is lost in the two-dimensional representation of the road in GIS.

Without calibration, the GIS software interpolates between points A and D using the computed 1.0-mi (1.6-km) two-dimensional length of the section, placing Point B at milepoint 0.5 (actual distance 0.6) and Point C at milepoint 0.75 (actual distance 1.2). References along this roadway would be highly inaccurate, and the amount of error increases at points further down the road. Calibration using measured distances at points A and D would improve the accuracy of the intermediate references by adjusting the interpolated lengths based on the actual length of the segment by using the following formula:

$$d_i = D_i \cdot (d_{cp}/D_{cp})$$

where:

	d_i	= Calibrated distance for Point i.
	D_i	= Calibrated distance between control points.
	d_{cp}	= Measured GIS distance for Point i.
	D_{cp}	= Measured GIS distance between control points.

Using this equation and the calibration points A and D in the above example, the calibrated distance for points B and C would be calculated as follows:

$$d_B = 1.4 \cdot (0.5/1.0) = 0.7$$
$$d_C = 1.4 \cdot (0.75/1.0) = 1.05$$

In this case, the accuracy of Point B is not improved, but simply changed. It moved from an uncalibrated distance of 0.5 to 0.7, when the real-world value was 0.6. However, the accuracy of Point C was greatly improved. It changed from an uncalibrated distance of 0.75 to 1.05, with a real-world distance of 1.2. Taking accurately measured distances at points B and C (which have known coordinates) would enable these locations to be used as additional control points. Attributes or events specifically at these points would be located precisely (i.e., at milepoint 0.6 and milepoint 1.2, respectively). References between these new sets of control points would then be interpolated using the above formula and would be much more accurate when compared to using fewer or no control points.

Chapter 5. Integrating GIS and Safety Data

With a basic understanding of how roadway inventory data and crash data are traditionally referenced and how GIS interprets these data, one can turn to the integration of GIS and safety data for analysis. Linking highway safety data to GIS will provide challenges for State DOTs, MPOs, and other agencies. To make this integration a reality, three steps must be taken:

1. The LRS for the crash and roadway inventory data must be developed and made available for integration. In most cases, this development step has been completed by State DOTs in setting up their traditional systems, as previously described in chapter 3. Therefore, a good understanding of how the LRM has been implemented is necessary to plan for the development of an appropriate GIS that avoids linkage-related issues.

2. The spatial data model and a GIS route system must be developed. The GIS road network and the GIS route system are the foundation of GIS and are critical to the long-term success of any GIS for transportation applications.

3. The GIS is then used to process LRS data for display and spatial analysis.

This chapter discusses the technical issues associated with these steps as components of a GIS-based safety analysis system. Also provided is a discussion of other issues that must be considered when implementing GIS for safety analysis, including potential problems related to linkage of GIS and the LRS.

Planning for Integration

Perhaps the most critical step in developing a GIS safety analysis system is understanding the existing database design and planning for the development of the model that will integrate the newer technology into the older, well-established computing environments. The linkage between the existing LRS and GIS is dependent on several factors, including adhering to the naming convention and data type in use by the LRS, in particular for key attributes and data standards that may be in place for specific systems. Both of these issues are discussed below.

Attribute Coding Issues

Key field names are required to establish a database linkage for the crash and roadway inventory database. Small mistakes, such as improperly defining field names, data widths, or data type, could add unnecessary hurdles and delays in GIS development and linkage to linear referenced data. The key fields and items for linking the LRS to the GIS route system are Route ID and Route Measures for both point and linear event data (including crash and roadway inventory data). Other LRS items may also be key, depending on the scale and level of generalization of the LRS, such as with the implementation of multiple-route systems (refer to section at the end of this chapter).

For some agencies, a fully functional LRS linkage may also depend on additional key fields. A County-Route-Milepoint (CRM) is one example of an LRM common to some State DOTs, where an additional LRS attribute would have a "functional dependency" in the GIS LRS data model. In the CRM LRS, County (or jurisdiction) is a key attribute, because the route beginning mileage measurement is reinitiated for each county the route passes through. In this case, were GIS modeled using *only* Route ID and Milepoint, an incorrect linkage would probably occur. This is because without the use of the County attribute (where milepoint is measured independently for each county), the CRM data model would function like an RMP data model (where the route measurement runs continuously across the entire State). In such a case, several crashes occurring in *different* counties, but having the same route and mileage attributes, would probably be improperly mapped to the same point location, thus placing them in the wrong county.

Standardization Issues

Data standardization is a fundamental consideration in developing GIS for integration with existing databases. All working groups depend on standards being established within and outside of organizations and agencies to allow for cooperative efforts. Standards should be established for the LRS and GIS, and should address simple integration and processing of data within GIS. The DOTs have established the LRS based on standards that should include linear referenced data modeling, the data file naming convention, attribute coding, and the design of relational database tables. For placing crashes on the map, spatial data standards are less of an issue. However, the GIS route system standards should include spatial data modeling and considerations for scale, accuracy, resolution, and generalization. Standards for datum and projection mapping should also be considered.

The standards for hardware platforms, operating systems, network environments, database systems, and applications software are generally *not* an issue. The interoperability evidenced by the success of the Internet has proven this point. The standards for data definitions are much more important in order to provide reliability and portability in developing and maintaining systems and applications. The GIS software standards can also add to the complexity, since not all GIS share a common route system that is easily transferred from one vendor-specific application to another.

Developing the GIS Road Network Data Set

A GIS route system, based on a GIS road network data set, is required to display linear referenced data such as crashes or roadway inventory. Each route in the LRS coded in the GIS route system will be used as a reference for the display and analysis of the LRS data associated with that route. A GIS road network data set is produced from a transportation base map, which is developed through one of three means of digital data acquisition:

- Digitization of Source Materials – This method is a common, widely used, and well-tested means of data acquisition in which the road network line features, such as roadway centerline, are digitized from aerial photographs (considered to be primary source material) or from hardcopy maps (considered to be secondary source material). The process involves collecting the x-y coordinate values of the line features by tracing over each one using a digitizing tablet with a cursor or puck as the input device to locate and input map features into the computer. This type of manual production requires planning, source material preparation, and production setup, in addition to digital data post-processing. The costs for this type of data acquisition are significant and can represent the majority of system startup costs. Semi-automated methods using map scanning and line tracing technologies are being used to lower the cost and improve the accuracy of the digitization process.

- Acquiring Existing Digital Data From Other Sources – A cost-effective alternative to digitizing is to acquire digital data from a third-party source, such as USGS Digital Line Graphs (DLG). *Note: Large-scale 7.5 transportation overlay data from USGS may not be widely available for a given State and possibly may require updates to meet the completeness or accuracy standards for DOT use.*

- Directly Collecting Road Centerline Data Using GPS Technology – While this method is gaining popularity with DOTs, it is not widely used at this time due to some limitations in technology and an overall high cost for statewide coverage.

Correcting Data Topology

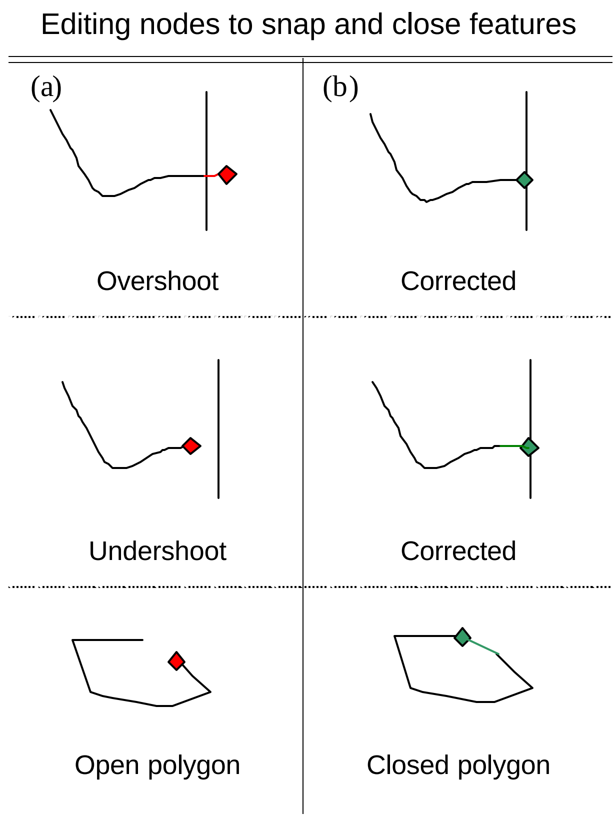


Figure 16. Line features require editing to correct for nodes that overshoot or undershoot, and polygons that do not close.

Road network data are often developed from a hardcopy map source or acquired from other sources of digital data and must undergo quality assurance and quality control procedures. To complete the development of the road network data, the spatial data must be cleaned up and edited to eliminate line overshoots, line undershoots, and to close all open polygons. Figure 16 depicts lines that overshoot or undershoot, and polygons that require node editing to close the links. The features represented in figure 16(a) may look like those of figure 16(b) when viewed on smaller scales. But in performing analyses, GIS can distinguish between differences of less than an inch in measured ground distance. Thus, these unedited links can cause significant GIS-T problems.

Special GIS tools are available to correct overshoots, undershoots, open polygons, and other types of topological problems. The connectivity problems illustrated above are cleaned up or corrected using system capabilities to properly connect the overshoot node to the neighboring line feature. When precision in placement along the neighboring line feature is required, the neighboring feature is split at the appropriate location and the dangle node is snapped to the newly created node feature.

Edge-Matching Map Sheets

Spatial data are often developed independently as map sheets. These map sheets must be post-processed or edge-matched to ensure that features on adjoining map sheets spatially match. Figure 17 illustrates how the edge-matching process adjusts features so that they are coincident with adjacent features that do not align on the adjoining map sheets.

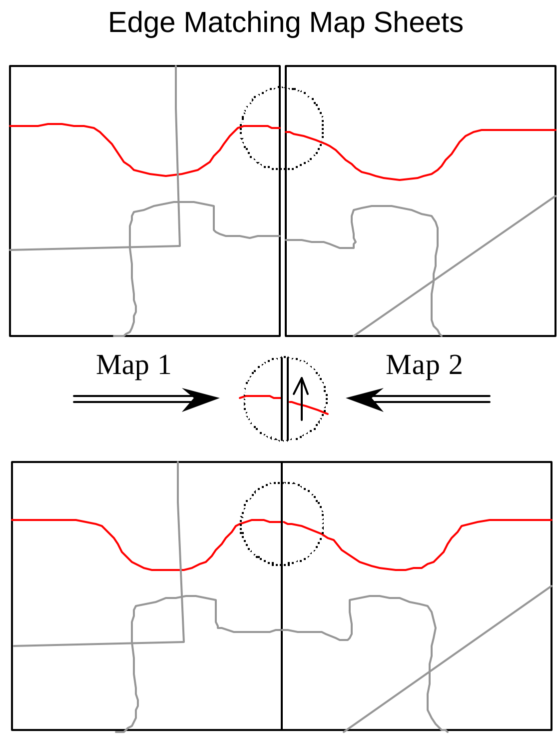


Figure 17. Features along the map sheet edge are aligned to match the location of an adjoining feature.

Creating Route Feature Data Types

Using available GIS tools, route features are created using the developed road network, following a two-stage process. First, the road network line feature elements are identified and coded as route feature elements. Then, route measures are added to the route feature at the beginning and end of the route and at additional locations along the route that will serve as control points. A simplified example of the process used by a safety engineer to create a GIS route system is as follows:

> *The safety engineer interested in linking an RMP LRS with the GIS road network database for safety analysis would require that the road network database have a route system representing the LRS. Given that the GIS route system had not been created, the safety engineer would go about that task using the GIS tools to create route features. Each route in the LRS that is of interest would first be identified. This could represent all routes for a DOT or only those represented in a particular study area. Next, using the tools available in GIS, the road network line features that represent each unique route are first selected to define the route features. Then, the appropriate Route ID is assigned to that route feature. This process would be repeated for each route.*
>
> *Next, the safety engineer might want to map crash locations using LRS data and the routes defined in GIS. Route measures in GIS will ensure good positional accuracy. However, route features will not contain measures until they are coded to the route. To complete the route development, the safety engineer would add route measures in the GIS route system. In the typical example, routes start at zero mileage and the length of each route is known and specified in the LRS. The route mileage of other locations, such as where routes intersect, may also be specified in the LRS. The safety engineer would first*

select a route, next select one of the known point locations, and then assign a correct route mileage to that location using the GIS tools for that purpose. Again, this process would be repeated for each route.

Figure 18 shows a route defined as a single feature composed of four sections defining the measures of the route. Routes do not have to coincide with the start and end of existing lines; they can be disjointed, but should not branch.

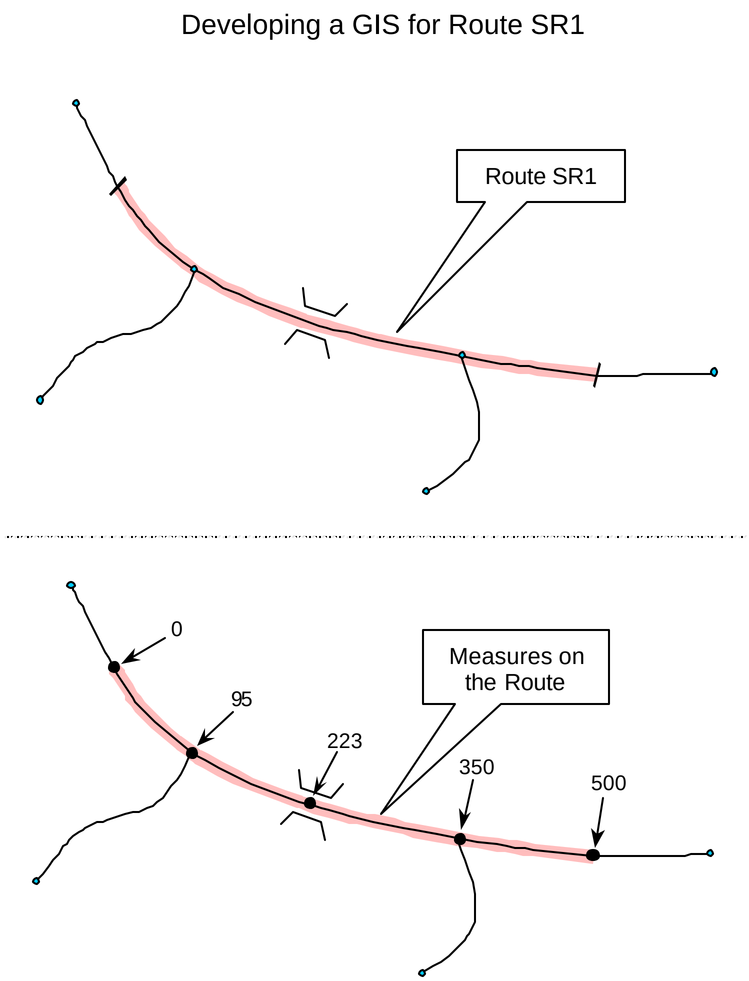


Figure 18. Routes are first defined in GIS and then a measuring system is defined along the route.

Use of Control Points

The process of adding route measures is often automated in GIS using control point data having route measure attributes. The control point data are determined from the LRS as the critical points on the road network, where linear control is required, then used to develop and maintain the GIS route system. For a route model that uses sections as a measurement control, each section's beginning location and ending location are coded with route measure values. For a route model that uses milepost markers, all milepost markers could be added to the GIS route system (as a route measure) to provide a highly accurate linear calibration of the GIS route system (see chapter 4 for a discussion of route calibration), although that additional effort is not necessarily required for GIS functionality.

A GIS route system can be developed using only the beginning and ending route measures and can still support the capabilities that GIS has to offer. This is because GIS maps linear referenced data to a single route feature (the linear reference), relative to its linear measurements. This becomes significant for a point location of cross-streets (or other significant features). To accurately map the linear location of the linear referenced data of one route relative to a cross-street, the intersection of the two routes would have to be developed as a control-point location. This would ensure that events such as intersection-related crashes are mapped at that point location in GIS.

Processing the LRS Data Using the GIS Route System

Once the GIS route system has been developed with a linear location referencing data model, the LRS data (crashes and roadway inventory) can be displayed in GIS using the GIS capabilities and functionality. The spatial accuracy will depend on the spatial accuracy of the road network database and the linear accuracy will depend on the use of control points. This section provides a brief overview of how the GIS capabilities manage the LRS data.

With the GIS route system developed, crashes and roadway inventory can be displayed in GIS without having to perform further data conversion or data development as long as the route measures that have been applied to the GIS route system are inclusive of the measurements in the LRS data. This is accomplished by establishing a database connection and relating the linear referenced route attributes, found in the crashes and roadway inventory data sets, to the route and measure attributes in GIS. This linkage between the LRS and the GIS route system is established during dynamic segmentation (also known as Dyn Seg), which is a set of GIS tools and processes that permit linear referenced data to be placed along a measured line or route system, and spatial attributes to be derived from that location placement. Figure 19 illustrates segmentation by the placement of linear referenced data along a measured route. The LRS is what allows dynamic segmentation to take place in GIS.

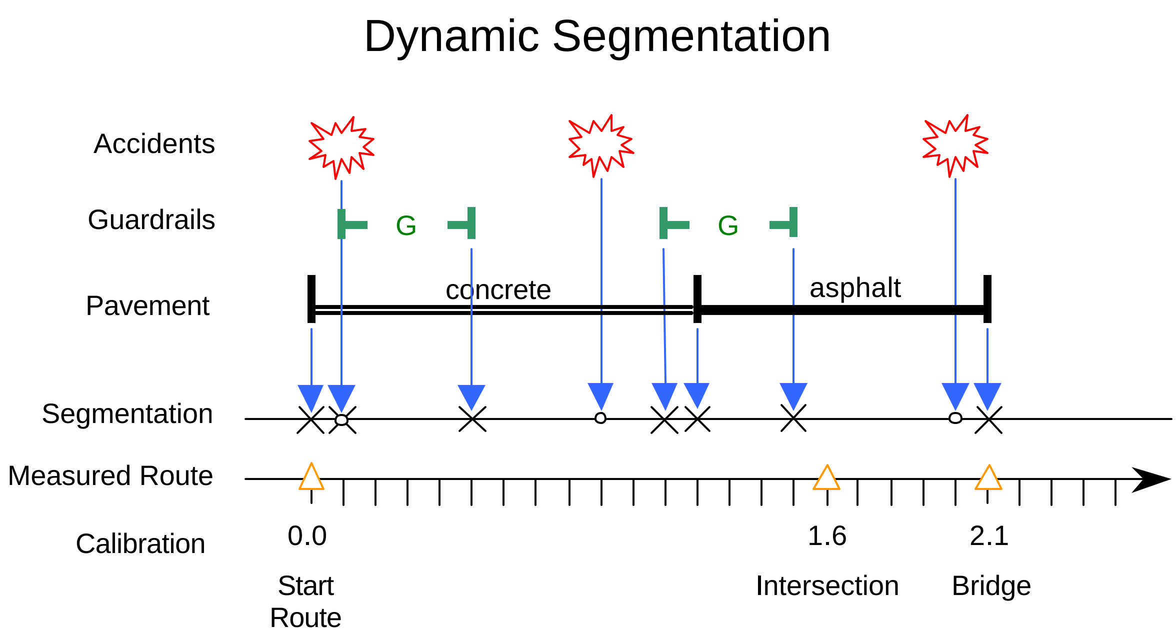


Figure 19. Using Dynamic Segmentation, point events and linear events are located along a measured line that has been calibrated at the measurements for an intersection and a bridge.

Dynamic segmentation was implemented as a means of modeling linear features and point events independently of the route feature type. No longer was it necessary to statically store route information as line feature elements representing homogeneous sections. Rather, the LRS could be organized as database tables (similar to figure 9), and the routes and measures could be used to dynamically display the LRS as linear events along the route feature element. If the linear referenced data attributes happen to change in the LRS, GIS could redisplay the linear referenced data using the same route system without having to redevelop the GIS routes.

Although implementation of dynamic segmentation will vary by GIS vendor, GIS uses dynamic segmentation to locate and display linear features along a route and/or to segment the route itself. This definition of “dynamic segmentation” has taken on a generic meaning of locating linear event data along a measured route. In either case, dynamic segmentation is used in GIS to produce linear referenced data that can then either be displayed on a monitor or produced as hardcopy. These data may also be converted to a spatial data file. In other words, the linear referenced data can be mapped to geographic coordinates, and the coordinates and linear attributes can then be stored as a spatial data set. This process is done by interpolating the distance along the measured line of the GIS route from the beginning measure to the ending measure of the line. Figure 20 illustrates both crashes and roadway surface conditions located along a route using dynamic segmentation.

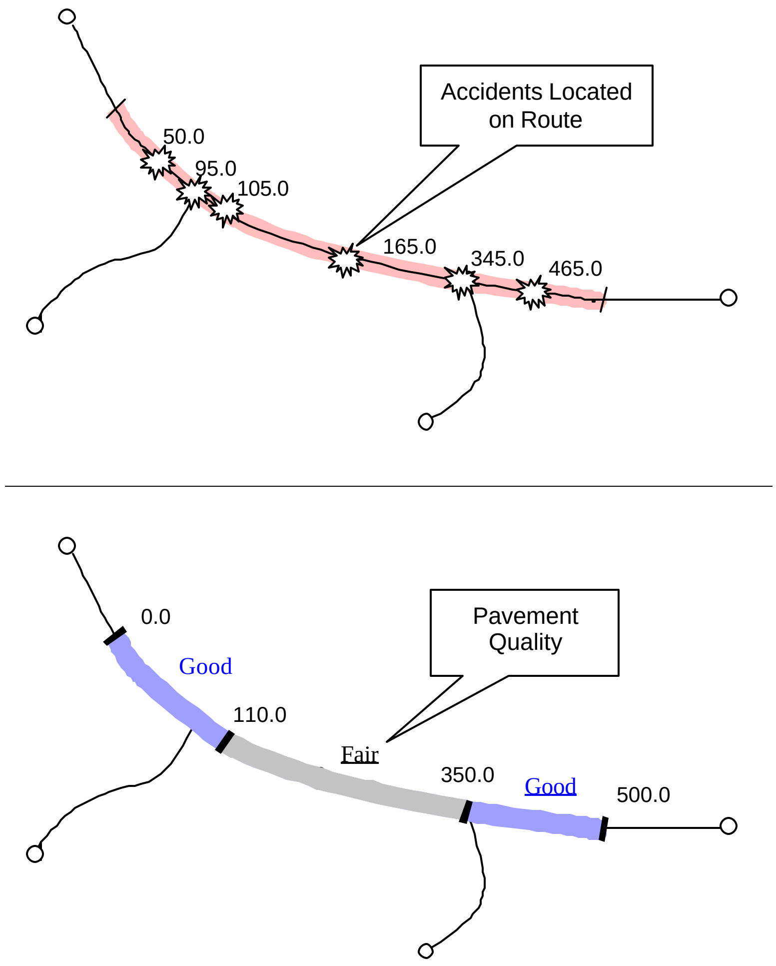


Figure 20. Both crash and pavement data are located on a route using Dynamic Segmentation.

When each discrete event is located and displayed in real time, the process is said to be “dynamic.” That is, the GIS road network is used and segmented with the selected linear referenced data set. However, dynamic segmentation is often performed once in batch processing for reasons of system performance or to fix a spatial coordinate to an event for historical reference (see the section below on Historical Linear Reference). This is especially true for data warehouse applications where all events along the route system are used to segment the route for a transactional database.

The safety engineer can use the same road network to analyze crashes by any tabular attribute or data column (e.g., year, crash severity, etc.). Linear referenced data can be joined to create new linear referenced data sets and can be used for dynamic segmentation. In fact, it may be useful to join two or more linear referenced data sets, such as crashes and roadway inventory, for statistical analysis for thematic map display (color-coded by data values). However, there may be cases where the safety engineer would require more than one LRM or LRS. In such a case, additional consideration should be made for using GIS to bring together linear referenced data, which is also discussed below.

Other Considerations

Discussed below are several additional issues that will need to be addressed in order to properly develop and maintain a GIS-based safety analysis system.

Multiple LRMs

Some agencies may have more than one LRM. For example, an MPO may share its LN linear referenced data with a DOT that uses an RMP LRM for all State routes. Of the eight HSIS States, three currently use multiple LRMs. As a result, the DOT would have to support both LRMs to use the MPO data. Whether it is the case that a DOT has to support disparate LRS's, or that the DOT wants to utilize an externally supported LRS, the DOT will have two issues to consider when integrating the LRS into GIS. First, the two LRS's will have been developed from two differing road network databases. This means that the LRS data will have to be attached to a common road network to make it useable. One means of accomplishing this feat is to use geometrical and rotational transformation techniques to match and merge the LRS attributes developed in one GIS route system into the other GIS route system. This process is called “conflation.” Second, not all vendors have fully implemented all LRS types. This means that each LRM must be supported and implemented in GIS, or the DOT will be required to perform data conversion or develop custom GIS programming.

LRS Versus Coordinates

As noted above, dynamic segmentation is used to develop spatial coordinates of crashes and other linear referenced data. GPS technology is beginning to be used to assist in the crash data collection task by providing x-y coordinates of the crash site location, and will be an improvement far superior to most current collection methods for crash locations. However, it would be expected that when overlaid with linear referenced crash data, the LRS data would not align well with the GPS data due to the difference in the datum, or set of parameters and control points used to accurately define horizontal or vertical measurements. Data derived from different sources can be resolved for accurate display and meaningful analyses if the datum is known. It is suggested that metadata be available for all coordinate data and include projection, datum, and unit of measure information.

As GPS data are more widely used, the precision of the road network layer (developed from digitization or another non-GPS method) will be questioned relative to the precision of the GPS crash data. The solution for the road network data to spatially “fit” other data having a higher spatial precision, such as GPS data, is to conflate one data set to the other more precise data set. Conflation is used to rectify spatial accuracy between two data sets by adjusting all coordinates of the data points in the less precise data set to allow for a better match between selected data points and their more accurate locations. This process is also referred to as “rubber sheeting.”

Another technique is to adjust the GPS data positional accuracy to the linear datum or snap the GPS data to the linear features in the GIS route system. Coordinate-based crash data derived from GPS or other sources, such as a different road network, will require adjustment to snap to the roadway as depicted in GIS. This will be expected for site location analysis mapped against the road network data. For States having GIS-located crash data, the buffering distance along routes, available in the GIS Safety Analysis Tools, will have to be considered to allow for the spatial margin of error in crash x-y placement relative to the GIS-defined roadway feature.

Address Geocoding

Crashes located by street address require a special set of GIS tools and a different GIS road network data. The process of linear location referencing by street address is called “Address Geocoding” or “Address Matching.” GIS does this in a manner similar to dynamic segmentation (except not dynamically). First, all streets in the GIS road network database are attributed by street name. Then the GIS street network is further developed to include a beginning and ending address for each street block.

After the street network has been processed to contain beginning and ending address ranges for each street section, crash locations can be displayed using the GIS street network and the crash street address designation using GIS tools for address geocoding. The GIS tools do this by first parsing the address into its parts: number, street name, street type, etc., with each address part stored in the crash address data fields. GIS can then locate the crash by street name and interpolate the location of the street number as a distance along the street block using the street network beginning address and ending address. The spatial location along the road network is assigned to the street address and the coordinate values are used for display.

The numeric value of the address need not be an actual postal address, but merely a legitimate address value within a range of beginning and ending addresses assigned to each block face. Thus, this is not really a true address location in that each address does not have an accurate location associated with any cadastral survey or postal assignment.

Historical Linear Reference

Over time, roads change. New highways are built, roads are realigned, roads are abandoned, routes are renamed, and roadway inventory continually changes. Route identifiers and road measurements may change in the process, and the system that maintains this linear information would be updated accordingly. As changes in the LRS occur, changes in the spatial representation of linear features in the road network layer need to be updated also. Often this synchronization of databases requires an interdepartmental cooperative effort.

As early as 1985, the HSIS States have provided crash and other related data to HSIS. Each year's data set represents an annual snapshot of the linear representation and events for the State's roadways. The annual data sets are adjusted to correct for changes in linear measurements for that year.

There will always be uncertainty in spatial accuracy in locating linear events using a method that relies on a current GIS data set to map historical linear referenced data. The best way to initially locate historical data is to have a separate view of the LRS for each year of data, both in the linear referenced database and in the GIS roadway network. This approach would provide a snapshot of the LRS and would ensure complete and accurate LRS linkage of the linear referenced database with the GIS route system for that time period. This method assumes that the data model and all roadway realignments and other similar changes are fixed in the LRS for that period of time.

It becomes a challenge for agencies to develop procedures and methodology for GIS to adopt for the accurate representation of the road network over the life of the system. It may be that all historical data cannot be confidently located. However, the key is to plan for the future and use old data as well as one can. For those States that implement a data warehouse approach to their LRS or linear referenced data, spatially enabling the data warehouse will provide a solution to historical data reference by generating coordinate locations for linear referenced data within the data warehouse.

Chapter 6. Conclusions

This report was written to discuss GIS/safety integration in terms that can be understood by both safety engineers and GIS specialists, and to describe the issues and solutions involved in this type of systems integration. This report is intended to serve as an educational document for both audiences to initiate a common dialogue. Hopefully, the content of the report will begin to bridge the gap between the desire to implement GIS highway safety analysis within an organization and the development of a GIS-T infrastructure to support that effort. The specific topics discussed included:

- The benefits that GIS technology offers in general analyses, including display, spatial, and network evaluations, as well as cell-based modeling. The applications from the already-developed GIS Safety Analysis Tools are discussed as examples (see chapter 2).

- A description of how historical safety data (crashes and roadway inventory) are acquired, why such data are collected as linear referenced data, and how linear referenced data are different from spatial data. Definitions of common route systems are provided with illustrations to show how each is different (see chapter 3).

- General background information on Linear Location Referencing Systems (LLRS or LRS), which includes an explanation of routes and measures, common types of LRS's, how linear referencing methods (LRMs) are used to locate crashes and roadway inventory, and how GIS uses LRS's to locate linear features (see chapter 3).

- A general understanding of how GIS manages road network data, and how route features are developed using the road network feature data. The impact of resolution, scale, and route calibration are discussed as related to data accuracy (see chapter 4).

- A detailed discussion of the process of integrating GIS and safety data, including the need to plan for the integration and development of the GIS road network and route system, and processing the LRS data within GIS (see chapter 5).

References

1. Federal Highway Administration, *HSIS – Highway Safety Information System: The Essential Information System for Making Informed Decisions About Highway Safety*, Publication No. FHWA-RD-00-033, Washington, DC, 2000.

2. Federal Highway Administration, GIS Safety Analysis Tools, Version 2.0 (CD format), Washington, DC, July 2000.

3. Federal Highway Administration, *GIS-Based Crash Referencing and Analysis System*, HSIS Summary Report, Publication No. FHWA-RD-99-081, Washington, DC, February 1999.

4. Federal Highway Administration, *GIS Tools for Improving Pedestrian and Bike Safety,* TechBrief, Publication No. FHWA-RD-00-153, Washington, DC, July 2000.

5. Federal Highway Administration, *Using GIS in the Analysis of Truck Crashes*, HSIS Summary Report, Publication No. FHWA-RD-99-119, Washington, DC, June 1999.

6. Transportation Research Board, *Highway Location Reference Methods,* NCHRP Synthesis 21, Washington, DC, 1974.

7. National Map Accuracy Standards, http://www.usgs.gov/fact-sheet/map-accuracy/map-accuracy.html.

8. GIS/Trans, Ltd., *Linear Reference Practitioner's Handbook* (Draft), November 1997.

9. *TIDE: A GIS-Linked Data Warehousing Approach for Building an Integrated Transportation Information Environment*, American Association of State and Highway Transportation Offices, GIS-T '98, June 1998.

Appendix A. Case Studies

In an attempt to better understand some of the issues associated with integrating GIS and safety analysis, two case studies were conducted using data from the HSIS States of Maine and Washington. Spatial data sets were acquired from both States and were integrated with HSIS data currently maintained in the system. These two studies are described below and provide examples of the successes and problems of developing GIS linkage to an LRS.

Maine Case Study

In HSIS, crash, roadway, traffic volume, and interchange data files are maintained for Maine back to 1985, representing 22,000 roadway miles (35,405 roadway kilometers), and an average of 38,000 crashes per year. The Maine Department of Transportation (MeDOT) relies on the Transportation Integrated Network Information System (TINIS) to bring together data for crashes, roadway inventory, bridges, railroads, and project history/maintenance, and to support their LRS. Recently, MeDOT, with the assistance of GIS/Trans, Ltd., implemented the Transportation Information for Decision Enhancement (TIDE) system as a data warehouse to integrate their legacy systems with GIS and to augment the LRS to provide new system-wide access and capabilities. (9)

One of the many benefits of TIDE is in the area of historical data referencing. Using a process referred to by MeDOT as "static segmentation," the GIS coordinates for all data linked to the LRS are managed on a periodic weekly basis, such that any changes occurring in the LRS during that time period are reflected in GIS. For historical data referencing, this process addresses the issue of linear referencing and fulfills the department's goal of providing historical analysis capabilities for crashes. Thus, when road realignment takes place, the crash *will not* be imprecisely placed in an improper location along the new alignment, but will be located more accurately to a coordinate position that matches the location of the roadway at that point in time.

Maine uses a link-node (LN) system for their LRS, which means that the Link ID is a key variable for routes and is defined as a composite field made up of beginning Node ID plus ending Node ID. The system has been fairly stable in Maine, but over the years, new links were created that required additional Node IDs to be added. These additions also resulted in changing the four-digit Node ID number to five digits, which, in turn, increased the link number by two digits. All of these changes were implemented in TINIS. The TINIS data was then migrated to TIDE – source of the MeDOT GIS route system made available to HSIS. This seemingly small change to the Node ID number had a great impact on the ability to integrate HSIS data and the GIS data. These additional digits in the link numbers had not been changed within HSIS. In order to link the two systems, the Link IDs were changed in all 12 years of Maine data residing in HSIS. This problem clearly illustrated a key point that even with a well-managed GIS, such as the one Maine has, integration solutions will need to be found for existing incompatibilities.

In summary, the following conditions and situations, both advantageous and problematic, were encountered:

- Maine's data warehouse approach implemented in the TIDE system provided a very reliable and desirable approach to historical data referencing and mapping crash locations.

- The Maine road network and LN route system implemented in GIS provided a solid basis for mapping crashes and other linear referenced data.

- Changes for key linear feature data formats that had migrated to GIS had not been applied to existing Maine data in HSIS, which resulted in linkage problems until the HSIS data were brought up to the new Maine data standard.

- A business decision long established by Maine – to round up the crash data linear location reference – found a different solution in TIDE than the solution implemented years earlier in HSIS. HSIS had to be reconciled with the Maine DOT source data to permit complete mapping of available data.

- The Maine LN LRS spatial data were found not to adapt well with the GIS Safety Analysis Tools, which anticipates an RMP LRS route system.

Washington State Case Study

The Washington State DOT (WSDOT) relies on the Transportation Information and Planning Support (TRIPS) system to bring together data for crashes, roadway inventory, bridges, curve/grade/features data, roadway crossings, roadside facilities, special-use lane information, railroad grade crossing index, and traffic data to support their LRS. In HSIS, crash, roadway, traffic volume, curve/grade, and interchange data files are maintained for WSDOT for all years since 1993, representing 8,400 roadway miles (13,518 roadway kilometers), and an average of 35,000 crashes per year.

Washington State has a great investment in developing their GIS road network and route system data. They have developed GIS route systems on two scales of resolution – 1:500,000 (good for small-scale mapping) and a higher resolution GIS road network based on 1:24,000-scale maps. The WSDOT GIS route systems contain route measures based on the TRIPS system's State Routes and Accumulated Route Mileage (ARM), a type of RMP LRS. Although both route systems contain the same Route ID and similar ARM values, they must be treated differently in the linkage with the LRS.

In developing GIS capabilities for use with Washington State HSIS data, the WSDOT GIS route system was used for linkage to HSIS. Working with the two GIS road networks available, it was easily discerned that road features are depicted differently, as would be expected. For example, at the 1:500,000 scale (small-scale mapping), a highway interchange containing ramps and collectors is generalized as a simple intersection of mainline routes. This generalization and reduction of detail is adequate for the mapping of mainline features and crashes, but the lack of ramp features degrades the accuracy of the GIS linkage to the LRS. At the 1:24,000 scale, the roadway has been modeled differently using additional details. Ramp features and divided roadways are present. This spatial data model represents a truer depiction of the roadway, where each lane of travel is represented as a route with an increasing or decreasing direction and ramps are represented as other routes.

The effort to map Washington State crashes and roadway inventory data (referred to as "WSDOT Roadlog" data) represented the first attempt, by anyone, to use the WSDOT GIS route data for that purpose. After a clear understanding of the linear referenced data model deployed by WSDOT, the linkage with the HSIS data was established for four years (1993-1996). As shown in table 2, the linkage with the roadway inventory data across all routes to each corresponding year of WSDOT 1:500,000-scale route data was achieved with an average success rate of 86 percent for all route types. When broken down by road type, a 98.9 percent success rate was achieved for linkage of mainline roadway inventory data to the 1:500,000-scale route data. This linkage would be equivalent to mapping to the road centerline. The difference between mapping all data and mapping mainline data is thought to be the result of ramps not having a representation in GIS at that scale.

Table 2. Summary of success in mapping Washington State Roadlog data to 1:500,000-scale route system.

All Data				Mainline Data			
Data Set	Roadlog	Miles	% Mapped	Data Set	Mainline	Miles	% Mapped
1993	8,583	7,314	85.2%	1993	7,265	7,207	99.2%
1994	8,659	7,321	84.5%	1994	7,265	7,209	99.2%
1995	8,352	7,317	87.6%	1995	7,265	7,204	99.2%
1996	8,397	7,240	86.2%	1996	7,265	7,122	98.0%
Average	8,498	7,298	85.9%	Average	7,265	7,186	98.9%

1 mi = 1.61 km

Washington State data mapped at the 1:24,000 scale presented several challenges in terms of geographic division of data and functional dependency. First, the WSDOT route systems were developed independently for the 39 counties in the State, which provided a technical challenge to working with HSIS data that are maintained on a statewide basis by year. Scripts had to be developed to handle the multiple-route systems and the geographic division of the data by county jurisdiction. Secondly, the large-scale mapping permitted greater feature resolution and less generalization, and depicted ramps and divided highways not shown in the smaller scale mapping.

As previously noted, the WSDOT TRIPS system uses ARM values for locating crashes and features. These values, computed from the State Route Milepost (SRMP) equations, contain measurements for increasing and decreasing directions on the roadway. For undivided highways, the ARM values would be the same, regardless of direction. But for divided highways, the increasing and decreasing side of the same route section can have different ARM values. To manage the differences in measurements for increasing and decreasing directions, WSDOT represented the TRIPS LRS data model in the GIS spatial data model by developing separate route systems within the same GIS – one for increasing routes, a second for decreasing routes, and a third for ramps (each as separate route systems). This solution preserved the functional dependency inherent in the TRIPS data for direction of route measurement.

Unfortunately, the HSIS Washington State roadway inventory files do not contain a key variable for direction of mileposting on the State route, which separated features by increasing and decreasing the direction of travel. As a result, the GIS linkage of the Roadlog data with the 1:24,000-scale route system data model could not be achieved. Subsequently, comparisons between the two mapping scales for the Roadlog data could not be made.

For crash data, however, both scales could be linked. Taking advantage of a little-used crash data variable for direction of crash impact, crash locations were mapped with great success. The larger scale mapping allowed better accuracy in mapping events for divided roadways and interchanges by mapping crash data to the proper side of the roadway or to a specific ramp. The results for the 1:24,000-scale route data (see table 3) show that linkage with HSIS crash data was achieved for 97.1 percent of all available crashes and 98.9 percent of all mainline crashes. The 1.8 percent difference in mapping all crash data and mainline data is attributed to being able to accurately map crashes occurring on interchange ramps and couplets.

Table 3. Summary of success in mapping Washington State crash data to 1:24,000-scale route system.

All Data				Mainline Data			
Data Set	Records	Crashes	% Mapped	Data Set	Records	Crashes	% Mapped
1993	33,837	32,972	97.4%	1993	30,315	30,017	99.0%
1994	36,784	35,806	97.3%	1994	32,933	32,525	98.8%
1995	38,935	37,660	96.7%	1995	34,711	34,284	98.8%
1996	42,141	40,801	96.8%	1996	37,737	37,365	99.0%
Average	37,924	36,810	97.1%	Average	33,924	33,548	98.9%

For comparison, the use of the 1:500,000-scale model resulted in 89 percent of all crashes being properly linked (see table 4), which is 8 percent lower than achieved with the larger scale model. This lower value is due to the generalized representation of the roadway within GIS at this scale, where only mainline roadway features are represented, and ramp and collector features are not shown. Thus, one cannot map crashes to a roadway feature not depicted. This phenomenon was previously illustrated in figure 10 and described in chapter 4. Note that there was a very small increase (0.6 percent) in the mapping of mainline crash data for the 1:500,000-scale model over the 1:24,000-scale model. This increase is believed to be caused by the complexities and possible inaccuracies in the spatial data sets of the larger scale mapping.

Table 4. Summary of success in mapping Washington State crash data using the WSDOT 1:500,000-scale route system.

All Data				Mainline Data			
Data Set	Records	Crashes	% Mapped	Data Set	Records	Crashes	% Mapped
1993	33,837	30,214	89.3%	1993	30,315	30,214	99.7%
1994	36,784	32,741	89.0%	1994	32,933	32,741	99.4%
1995	38,935	34,525	88.7%	1995	34,711	34,525	99.5%
1996	42,141	37,588	89.2%	1996	37,737	37,588	99.6%
Average	37,924	33,767	89.0%	Average	33,924	33,767	99.5%

In summary, this case study highlighted the need for a complete understanding of the LRMs in use in order to develop GIS linkage to safety data for highway safety analysis. Below is a summary of the conditions and situations, both advantageous and problematic, that were encountered in the Washington State case study:

- The Washington State road network and ARM route system implemented in GIS provided a solid basis for mapping crashes and other linear referenced data.

- The quality of development and the completeness of the Washington State GIS provided a high degree of success (99 percent) in mapping crash data from mainline roads.

- To fully exploit the complexity of the Washington State GIS route system, a thorough understanding of the LRMs and the LRS was required.

- Newly discovered methods for using HSIS data for GIS integration – exploiting key linear and spatial data field attributes – provided opportunities and challenges in GIS development.

- GIS linkage of the HSIS Roadlog data with the WSDOT 1:24,000-scale spatial data could not be achieved due to a critical variable not being available in HSIS. This variable supported the functional dependency for direction of travel in Washington State roadway inventory data.

- The Washington State spatial data, developed using geographical coordinates and mileage as a route measurement, were found not to adapt well with the GIS Safety Analysis Tools, which anticipates State plane coordinates having units of measure in meters.

Glossary of Terms

Cartesian Coordinates:

A two-dimensional x-y location of a point on a plane (planar) in relation to two intersecting straight lines (axes). If the axes are perpendicular to each other, the coordinates are rectangular; if not, they are oblique. The x-axis measures the horizontal distance and the y-axis measures the vertical distance from the origin. An x-y coordinate defines every point on the plane. Relative measurement of distance, area, and direction are constant throughout the Cartesian coordinate plane.

Conflation:

A process by which two digital maps, usually of the same area at different points in time, or two different thematic maps of the same area, may be matched and merged into one through geometrical and rotational transformations. (Association for Geographic Information (AGI), the AGI GIS dictionary, http://www.agi.org.uk/pag-es/dict-ion/dict-agi.htm).

Coordinate:

Pairs of numbers expressing horizontal distances along orthogonal axes; alternatively, triplets of numbers measuring horizontal and vertical distances. Any of a set of numbers used in specifying the location of a point or position.

Coordinate System:

A framework used to define the position of a point, line, curve, or plane, and derivative map features within a two- or three-dimensional space. A reference system for defining points in space or on a particular surface by means of distances or angles, or both, with relation to designated map projection, datum, one or more standard parallels, and a central meridian.

Datum:

A set of parameters and control points used to accurately define the three-dimensional shape of the Earth (e.g., as an ellipsoid). The corresponding datum is the basis for a planar coordinate system. A reference surface for horizontal or vertical measurements.

A base reference level for the third dimension of elevation for the earth's surface. A datum can depend on the ellipsoid, the earth model, and the definition of sea level.

Divided Highways:

A divided highway is a roadway where the opposing directions are separated by a median that restricts movement between the two directional roadbeds. Note that some GIS installations consider highways to be divided only if the scale of the map and the size of the median are such that the two roadbeds can be mapped separately.

Dynamic Segmentation:

Dynamic segmentation of lineal spatial objects provides a means by which new point or line objects can be created by relating the distance-referenced attributes with a manageable set of distance-referenced linear objects. Dynamic segmentation removes the need for a set of spatial objects for each attribute. Spatial objects and distance referencing of routes are used to create attribute-based spatial objects as needed. A method of referencing attribute data on demand, based on variable segmentation of a single route or network structure.

Generalization:

A reduction of detail and a transformation of cartographic data into a representation at a reduced scale. The process of moving from one map scale to a smaller (less detailed) scale, changing the form of the features by simplification, etc.

Global Positioning System:

A satellite-based navigational system allowing the determination of any point on the earth's surface with a high degree of accuracy given a suitable GPS receiver. In the past the U.S. Department of Defense has intentionally degraded the accuracy of the satellite signal for non-U.S. military users. The error introduced into the signal is known as "selective availability." Error in the accuracy of GPS-derived positions can also be introduced through the nature of local conditions, for example, multipath. These errors can be greatly reduced using a technique known as "differential GPS." (Modified from the Association for Geographic Information (AGI), http://www.agi.org.uk/).

Linear Feature:

A geographic feature that can be represented by a line or set of lines. For example, rivers, roads, and electric and telecommunications networks can all be represented as linear features.

Linear Location Referencing Method:

A mechanism for finding and stating the location of an unknown point along a network by referencing it to a known point. All linear referencing methods consist of traversals and associated traversal reference points that together provide a set of known points, a metric, and a direction for referencing the locations of unknown points. No attributes are assigned to linear referencing methods.

Linear Location Referencing Systems:

The total set of procedures for determining and retaining a record of specific points along a linear feature. The system includes the location reference method(s), together with the procedures for storing, maintaining, and retrieving location information about points and segments on the highways.

Linear Referencing:

Process of identifying a location(s) on a network or specific link in a network by specifying a start position, direction, and distance.

Mileage (mileage measurement):

A given distance expressed in miles.

Milepoint:

The name given to the numerical value of the mileage displacement from a base point to any location.

Milepost (mileage marker):

One of a series of posts or markers set along a highway or other thoroughfare to indicate distance in miles. A physical entity, ordinarily a sign, placed beside a highway that contains a number that indicating the mileage to that point from some zero point on the highway.

Reference Markers:

Physical objects along roads that may or may not have a simple relationship to the length of roads and that form control points with a route and milepost measurement.

Reference Point:

A fixed identifiable feature, such as an intersection, railroad crossing, or bridge, from which a location can be measured or referenced.

Reference Post:

A physical entity, ordinarily a sign, placed beside a highway that contains a number that does not reflect a mile point (MP), but is an identification number for the location of the post. The identification number is associated with the actual MP of the location in office records.

Scale:

The proportion between two sets of dimensions.

In relation to maps, the best scale for your map depends on the resolution of the original data, as well as the level of detail you want your map to include. For example, 0.25 in^2 on a 1:250,000-scale map represents approximately 1.0 mi^2 (640 acres) on the ground. But 0.25 in^2 on a 1:63,360-scale map represents 0.25 mi^2 (160 acres).

State Plane Coordinate System (SPCS):

The plane-rectangular coordinate systems developed by the U.S. Coast and Geodetic Survey (now known as the National Geodetic Survey or NGS), one for each State in the United States, for use in defining positions of geodetic stations. Each State is covered by one or more zones, over each of which is placed a grid imposed upon a conformal map projection. Zones having limited north-south dimension and indefinite east-west extent have the Lambert conformal conic map projection with two standard parallels as the based for the State plane coordinate system. Zones in which this sequence is reversed (i.e., limited east-west dimension and indefinite north-south extent) have the transverse Mercator projection as the basis.

Traverse:

A method of surveying in which lengths and directions of lines between points on the earth are obtained by or from field measurements, and are used in determining the positions of the points.