Objective

This project was designed to evaluate the potential benefits and limitations of integrating transportation safety-related databases in a spatially-referenced geographic information systems (GIS) environment. The project focus was the analysis of crash records and driver records.

Problem Statement

Information on crashes and crash characteristics is collected using various methods, for various purposes, and across several separately developed and managed databases. As a result, key relationships spanning different databases may go undiscovered, and transportation safety analyses may not be sufficiently comprehensive and informative. Moreover, many database attributes, such as crash locations, vehicle owner addresses, and medical facility locations, can be more useful if the data are spatially referenced. Linking the spatially-referenced data among several transportation safety-related databases can ultimately improve the analysis of transportation safety issues.

Integrated Safety-Related Databases

The National Highway Traffic Safety Administration (NHTSA) outlines six database types that are related to motor vehicle crashes and that are typically included in transportation safety information systems:

### Six database types typically included in transportation safety information systems

<table>
<thead>
<tr>
<th>Database Type</th>
<th>Data Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crash information</td>
<td>Crash time/location, drivers involved, vehicles involved, injuries, crash circumstances</td>
</tr>
<tr>
<td>Driver information</td>
<td>Name, address, date of birth, license number, driving restrictions, traffic violations, previous crashes, etc.</td>
</tr>
<tr>
<td>Vehicle information</td>
<td>Vehicle ownership, registration, make/model/year, VIN</td>
</tr>
<tr>
<td>Citation/adjudication information</td>
<td>Driver citations, convictions, sentencing information</td>
</tr>
<tr>
<td>Roadway information</td>
<td>Structure, classification, geometry, pavement type, traffic volume, roadside features, etc.</td>
</tr>
<tr>
<td>Statewide injury surveillance information</td>
<td>EMS data, hospital ER data, hospital stays, outpatient services, death certificates</td>
</tr>
</tbody>
</table>

In addition, transportation safety information systems may include insurance, land use, topography, and U.S. census data.
To help integrate and analyze spatially referenced crash-related data, GIS software can work with different location referencing systems across multiple information sources. Data elements may include, for example, the location of the crash, the location of the nearest hospital to the crash, and the addresses of the drivers involved.

The figure below, from the NHTSA’s 1998 Traffic Records Advisory but updated regularly, illustrates the ways data from various sources are linked to crash data.

### Key Benefits

The integration of spatially-referenced data can help improve transportation safety analysis:

- Spatially-referenced data are useful for displaying spatial patterns, integrating data from different sources, and generating new research questions.
- Linking spatially-referenced GIS safety data can improve information about crash types and frequencies and about the locations associated with drivers and crashes.
- Spatially-referenced data collected and stored for non-transportation purposes, such as demographic or land use data, can be used to explore causal relationships in crash patterns.

### Implementation Issues

Before implementation, any integrated spatially-referenced database will need to contend with four main issues:

#### Technical issues

These include concerns about the accuracy, completeness, or format of spatial data. Most issues may be resolved by making spatial data GIS-compatible and by accurately recording crash position and other locations (e.g., citations or hospitals) using geospatial coordinates.

#### Methodological issues

These include the types of conclusions that can fairly be drawn from spatial data. Methodological issues can be addressed primarily through an understanding of the constraints of spatial data and through the use of newer statistical methods.

#### Administrative issues

These include ways to integrate data that have been collected and managed by agencies with multiple and varying objectives. Administrative issues may be mitigated through intra-and inter-agency cooperation on data collection and sharing standards.

#### Legal/Ethical Concerns

These include confidentiality issues raised by using legal records, medical data, or other information that may identify individuals. Legal/ethical issues may be resolved by “masking” individual data points, e.g., aggregating individual driver data into larger units (e.g., zip codes) or randomizing the data. Additionally, a “tiers of risk” approach can grant researchers access to different levels of confidential data for different research needs.