

NDOT Research Report

Report No. 528-14-803



**Development of a Nevada Statewide Database
for Safety Analyst Software**



February 2017

**Nevada Department of Transportation
1263 South Stewart Street
Carson City, NV 89712**

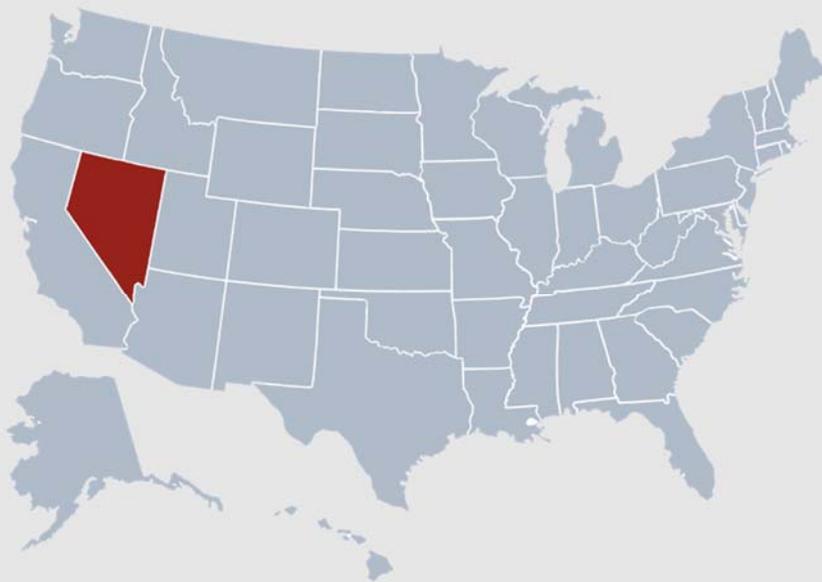


Disclaimer

This work was sponsored by the Nevada Department of Transportation. The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the State of Nevada at the time of publication. This report does not constitute a standard, specification, or regulation.

2015-2017 NDOT Research Project

Development of a Nevada Statewide Database for Safety Analyst Software



Center for Advanced Transportation Education and Research

University of Nevada, Reno

TABLE OF CONTENTS

Table of Contents	2
List of Figures	4
List of Tables	5
List of Acronyms	6
Executive Summary	8
1 Introduction	9
2 Safety Analyst	12
3 Literature Review	16
3.1 Safety Analyst Applications	16
3.2 Methodologies and Tools to Prepare Datasets	21
3.3 Findings and Conclusions	23
4 Data Requirement and Management	25
4.1 Overview of Data Requirement	25
4.2 Data Management	26
4.2.1 Step 1: Data collection	27
4.2.2 Step 2: Creating a dataset	27
4.2.3 Step 3: Importing data into a dataset	28
4.2.4 Step 4: Post processing a dataset	29
4.2.5 Step 5: Calibrating a dataset	29
4.2.6 Step 6: Serving a dataset	30
4.2.7 Copying Datasets	30
4.2.8 Reporting and Exporting Datasets	31
5 Preparation of Nevada Safety Analyst Database	32
5.1 Data Sources with NDOT	32
5.1.1 Roadway segment data	32
5.1.2 Intersection Dataset	41
5.1.3 Ramp Dataset	43
5.1.4 Crash Data	45
5.2 Data Processing	48
5.2.1 Road Segment Data Processing	48
5.2.2 Ramp Data Processing	49
5.2.3 Intersection Data Processing	50
5.2.4 Crash Data Processing	51
5.2.5 CSV File Import	52
5.3 Dataset Validation	53

Development of a Nevada Statewide Database for Safety Analyst Software

5.3.1	Import Validation	53
5.3.2	Post-Processing Validation	54
5.4	Data Accuracy of the Statewide Safety Analyst Database	55
6	Summary	56
	References	61

LIST OF FIGURES

Figure 1-1. Major data sources, users and relationship of the product database and the project database.....	11
Figure 2-1. General safety management process	14
Figure 3-1 Interface of Florida’s SADC Tool.....	19
Figure 3-2 Interface of Data Import Utility Tool in Ontario	23
Figure 5-1. Nevada HPMS road network 2014	36
Figure 5-2. NDOT urban/rural boundary data	37
Figure 5-3. NDOT historical HPMS AADT data layer	38
Figure 5-4. TRINA AADT data layer.....	39
Figure 5-5. Clark County Safety Analyst road layer	40
Figure 5-6. NDOT intersection data layer	43
Figure 5-7. NDOT statewide ramp layer	44
Figure 5-8. Map of statewide crashes 2006-2015.....	47
Figure 5-9. Example input for searching ramps in the road segment layer	50
Figure 6-1. Intersection type distribution chart of the Nevada Safety Analyst database.....	56
Figure 6-2. Crash distribution in different years.....	57
Figure 6-3. Distribution of crash severity	58
Figure 6-4. Distribution of crash types	58

LIST OF TABLES

Table 5-1. Roadway segment data from NDOT	32
Table 5-2. Data Items Required for the Full Extent and/or Sample Panel Sections	34
Table 5-3 Mandatory Roadway Elements Required by Safety Analyst	41
Table 5-4. Received Intersection Data from NDOT Divisions.....	42
Table 5-5. Mandatory Intersection Elements Required by Safety Analyst	42
Table 5-6. Mandatory Ramp Data Elements.....	45
Table 5-7. Crash Data for the Safety Analyst Crash Dataset.....	46
Table 5-8. Mandatory Crash Data Required by Safety Analyst	46
Table 5-9. Road Segment Data Gaps and Data Extraction to Fill the Gaps	49
Table 6-1. CSV Files in the Nevada Safety Analyst Database	56
Table 6-2. Numbers of Intersections in Different Control Types.....	57

LIST OF ACRONYMS

AADT - Annual Average Daily Traffic

Caltrans - California Department of Transportation

CAR - Crash Analysis Reporting

CATER - Center for Advanced Transportation Education and Research

CSV - Comma Separated Value

DBMS - Database Management System

DDL - Data Definition Language

DOT – Department of Transportation

EB - Empirical Bayes

EPDO - Equivalent Property Damage Only

FHWA - Federal Highway Administration

GDOT - Georgia Department of Transportation

GIS - Geographic Information System

HPMS - Highway Performance Monitoring System

HSIS - Highway Safety Information System

HSM - Highway Safety Manual

IDOT - Illinois Department of Transportation

IHSDM - Interactive Highway Safety Design Model

JDBC - Java Database Connectivity

JRE - Java Runtime Environment

LRS - Linear Referencing System

MeDOT - Maine Department of Transportation

NDOT - Nevada Department of Transportation

NHS - National Highway System

SADC - Safety Analyst Data Converter

SHSP - Strategic Highway Safety Plan

SPF - Safety Performance Function

UNR CATER

Development of a Nevada Statewide Database for Safety Analyst Software

SQL - Structured Query Language

TDM - Travel Demand Models

TIDE - Transportation Information for Decision Enhancement

TRINA - Traffic Records Information Access

TSE - Traffic Safety Engineering

UDOT - Utah Department of Transportation

UNLV - University of Nevada, Las Vegas

UNR - University of Nevada, Reno

VDOT - Virginia Department of Transportation

WisDOT - Wisconsin Department of Transportation

XML - Extensible Markup Language

EXECUTIVE SUMMARY

Safety Analyst is a software package developed by the Federal Highway Administration (FHWA) and twenty-seven participating state and local agencies including the Nevada Department of Transportation (NDOT). The software package implemented many of the analytical methods in Part B of HSM 2010. NDOT is to utilize Safety Analyst to establish a data-driven and self-sustaining program for highway safety management. Safety Analyst requires comprehensive data in particular formats, which are not available in Nevada. In this project, the Center for Advanced Transportation Education and Research (CATER) at the University of Nevada, Reno (UNR) collected road, intersection, ramp, traffic and crash data from different Nevada data sources and generated a statewide database meeting requirements of the Safety Analyst software. This statewide database contains the mandatory data of roadways, intersections, ramps, and crashes for utilizing Safety Analyst, so it allows NDOT to benefit from this advanced software to improve traffic safety in Nevada.

The completed database includes CSV data files that can be directly loaded into Safety Analyst and the GIS data layers for extended data applications. The final database includes:

- 5,152 road segments with a total length of 10849.37 miles
- 6,847 intersections including 749 intersections from the Clark County Safety Analyst database
- 1,178 ramps with a total length of 353.56 miles, including 571 off-ramps, 561 on-ramps, and 46 freeway-to-freeway ramps. Among the 1,178 ramps, 786 of them are diamond ramps, and 218 of them are partial clover leaf loop ramps
- 11-year AADT data from 2004 to 2015 with minor road AADTs estimated with the NDOT AADT default spreadsheet
- 423,126 crashes in 10 years from 2006 to 2015, in which 249,198 crashes are road-segment-related, 32,419 crashes are crashes on ramps and 141,509 crashes are intersection-related

The completed dataset will help the NDOT Traffic Safety Engineering (TSE) to identify the sites with high priority for safety improvement. Various network-screening methods can be conducted on the state road network or a sub network selected by safety engineers. All the data have been re-located with the Linear Referencing System (LRS) that is used in the Nevada HPMS road network. Therefore, the data in this statewide database can be integrated with any other dataset using the same LRS. The completed database is based on the HPMS road network. It has not covered all public roads in Nevada, such as some streets and county roads. This database can be updated with new crash data and AADT data, and the road network can be extended when more road/intersection data are available. The data update and extension can be implemented by using the data processing procedures documented in this report.

1 INTRODUCTION

It is the top priority for federal and state agencies to reduce traffic fatalities and serious injuries, which is also emphasized in the Strategic Highway Safety Plan (SHSP) of Nevada. The Nevada Department of Transportation (NDOT) works to reduce traffic crashes throughout the state by placing traffic safety at the forefront of their priorities. The advanced traffic safety tools (such as Safety Analyst [1], Interactive Highway Safety Design Model (IHSDM) [2], Highway Safety Manual (HSM) [3], etc.) were developed to support traffic agencies to identify and solve road safety problems [4]. NDOT is to utilize the advanced tools to establish a data-driven and self-sustaining program for highway safety management. Safety Analyst, outstanding one of these advanced tools, has significant capabilities by implementing cutting-edge analytical methods for traffic safety analysis and management. However, the software package typically requires comprehensive data in particular formats, which are not available in most states. In this NDOT research project, the Center for Advanced Transportation Education and Research (CATER) at the University of Nevada, Reno (UNR) collected road, intersection, ramp, traffic and crash data from different Nevada data sources and generated a statewide database meeting requirements of the Safety Analyst software. This statewide database contains the mandatory data of roadways, intersections, ramps, and crashes for utilizing Safety Analyst, so it allows NDOT to benefit from this advanced software to improve traffic safety in Nevada.

Safety Analyst is a software package developed by the Federal Highway Administration (FHWA) and twenty-seven participating state and local agencies including NDOT [5]. The software package implemented many of the analytical methods in Part B of HSM 2010 [4]. The Safety Analyst toolkit consists of four major tools for *Administration, Data Management, Analysis and Implementation of Countermeasures* and six software components for a traffic safety management process. The six modules help transportation agencies analyze the safety performance of specific sites, suggest appropriate countermeasures, quantify their expected benefits, and evaluate their effectiveness. The software can be used to proactively identify and analyze sites that have the highest potential for safety improvement, and then it can suggest countermeasures for the identified sites.

Extensive traffic and crash data are needed for traffic agencies to employ Safety Analyst, including properties of roadway segments, intersections, ramps, historical traffic volumes, and historical crash records. The Safety Analyst data are classified as required properties, conditionally required properties, and desired properties with consideration of how the software uses the data. For successful use of the software, the required and conditionally required variables are essential to the applications. If an agency would like to adopt this software tool, it has to compile a minimum set of data elements (required data) in the specific format. Recent nationwide survey reveals major hurdles for the adoption of Safety Analyst software. Some of the deterrents include non-availability of comprehensive data sources and tedious data importing and processing. Many of the data elements required by Safety Analyst are readily available from multiple sources at NDOT, but the data structure and format do not meet the requirements of the

software. Therefore, additional efforts are needed to assemble, review, process, and format the existing datasets.

Local or remote data can be imported into the software database using the *Data Management Tool* of Safety Analyst [6]. Safety Analyst supports two basic mechanisms for data import: *File Import* and *Database-to-Database Mapping*. For traffic safety agencies that maintain their complete data inventory in a Structured Query Language (SQL)-compliant database management system (DBMS), the Database-to-Database Mapping mechanism is the preferred alternative to load data into Safety Analyst. With the other alternative, the standard input files for Safety Analyst can be in either the comma separated value (CSV) format or the extensible markup language (XML) format. The File Import provides a flexible mechanism for agencies maintaining data without the SQL-compliant DBMS. The Database-to-Database Mapping does not support the data conversion when data of crashes, roads, and traffic are stored in different Geographic Information System (GIS) layers with different linear reference methods. This challenge exists in the data of many traffic agencies, including NDOT. Therefore, UNR CATER applied GIS data processing methods to integrate statewide data from different sources and formats, generated the dataset in CSV files, and then imported the data into the Safety Analyst database. The final statewide database is in the required format of Safety Analyst and can be used for other transportation applications. The database contains the statewide roadway data, intersection data, ramp data, and crash data from NDOT. The database includes intersections of the Clark County Safety Analyst data prepared by the University of Nevada, Las Vegas (UNLV) in a previous project. In addition, UNR CATER staff extracted road, intersection, and ramp data by reviewing the infrastructures in Google Earth. All of the obtained or generated datasets are stored and managed in a project database. Although this research project database was used to serve the research team, the NDOT project manager and other interested NDOT engineers can access all the data obtained for this project. For distinguishing the two databases in this project, the statewide database for Safety Analyst is called “Product Database” in this document, and the

project database for the research team is called “Project Database.” The relationship of the Product Database, the Project Database, the data users, and sources are demonstrated in Figure 1.

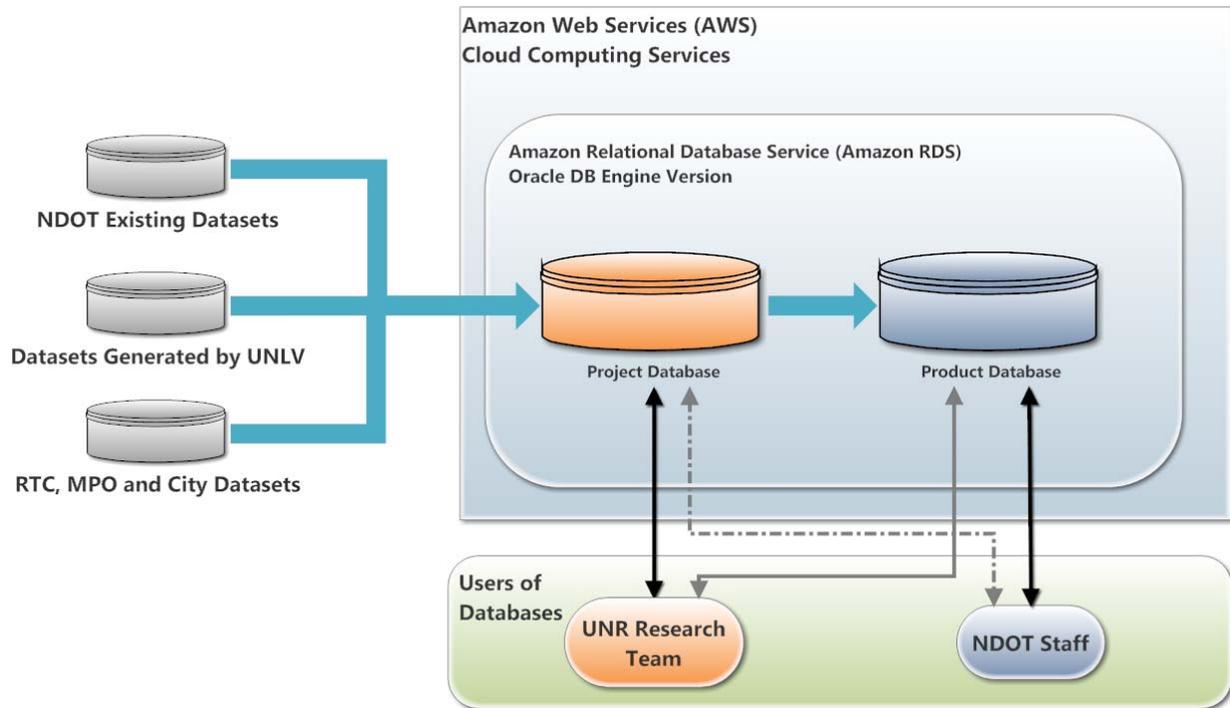


Figure 1-1. Major data sources, users and relationship of the product database and the project database

In this project, UNR CATER worked closely with MRIGlobal and Exelis during the development of this statewide Safety Analyst database, under the technical support agreement of NDOT and the Safety Analyst contractors. MRIGlobal and Exelis offer NDOT with special fixed-fee increments or units of contractor-provided service for consultation and support to assist the agency in preparing data and using Safety Analyst.

Chapter 2 of this report introduces the Safety Analyst software package by summarizing the information from multiple Safety Analyst manuals and reports. Chapter 3 presents the literature review findings of applications of Safety Analyst in different states, and how the software users prepared their datasets. Chapter 4 documents the Safety Analyst data requirements, how the data is managed and tools for implementing the Safety Analyst data management. Chapter 5 documents the methods and procedures for UNR CATER to prepare the Nevada statewide Safety Analyst database. At the end of this report, Chapter 6 summarizes the efforts and achievements of this project.

2 SAFETY ANALYST

Safety Analyst is a set of automated analytical tools for site-specific safety improvements. Safety Analyst guides the decision-making process to identify safety improvement needs and recommend cost-effective countermeasures. It needs to be noted that Safety Analyst is not intended for non-site-specific highway safety programs such as vehicle design improvements, graduated licensing, occupant restraints, or alcohol/drug use programs [5]. Safety Analyst was implemented to identify crash patterns at specific locations, including the frequency and percentage of particular crash types. These capabilities allow identification of potential engineering improvements at a site. The analysis of crash patterns can also be used to guide enforcement and public education efforts in an area.

A safety management process includes six main steps [3]:

Step 1: Identification of sites with potential for safety improvement

Step 2: Diagnosis of the nature of safety problems at specific sites

Step 3: Selection of countermeasures at specific sites

Step 4: Economic appraisal for sites and countermeasures under consideration

Step 5: Priority rankings of improvement projects

Step 6: Safety effectiveness evaluation of implemented countermeasures

Safety Analyst implemented the process with four software modules:

Module 1 - Network screening

The basic purpose of the network-screening module is to review the entire roadway network or portions of the roadway network. It identifies and prioritizes those sites for potential safety improvements. Therefore, the model merits further investigation (i.e., sites to which the other Safety Analyst modules should be applied). The network-screening procedure uses the following data elements as input:

- Geometric design features
- Traffic control features
- Traffic volumes
- Crash history
- Crash characteristics
- Safety performance functions (SPFs)

Detailed engineering studies of candidate improvement sites are expensive (time and resources). A highway agency can investigate only a limited set of sites in a period with the limited resources. Network Screening identifies sites (roadway segments, intersections, or ramps) highest ranked regarding cost effectiveness. Sites for network screening may include all roadway segments, intersections, and ramps of an agency or a subset of the network. Once a

site list for network screening is selected, the analyst specifies the screening method. Different network screening methods are available in Module 1:

- 1) Basic network screening (with Peak Searching on roadway segments and CV Test)
- 2) Basic network screening (with Sliding Window on roadway segments)
- 3) Screening for high proportion of specific crash type
- 4) Sudden increase in mean crash frequency
- 5) Steady increase in mean crash frequency
- 6) Corridor screening

The first five screening methods are on a site-by-site basis, while corridor screening performs an analysis across a group of sites that are treated as a single unit or entity. A corridor may include all site types (i.e., roadway segments, intersections, and ramps). The output from Module 1 is a report with a list of the highest-ranked candidates (sites or corridors) for further investigation with Safety Analyst. The list varies depending on the selected screening method.

Module 2 - Diagnosis and countermeasure selection

Module 2 is for the diagnosis of safety problems and the selection of possible countermeasures for a specific site. This module combines the second and third steps of the safety management process. A site input to Module 2 is normally selected by the network-screening module or can be selected by the analyst on other basis.

To diagnose the nature of safety problems at a specific site, Module 2 provides the following functions:

- Generate collision diagrams
- Generate crash summary statistics
- Conduct statistical tests on crash frequencies and proportions

With an implemented expert system, this module guides the analyst through appropriate investigations to identify particular safety concerns at a site. The result of this diagnosis process is a list of recommended countermeasures that could mitigate particular crashes at the diagnosed sites. Countermeasures recommended by this module will be considered for further economic analysis in Module 3.

Module 3 - Economic appraisal and priority ranking

Module 3 is for an economic appraisal of countermeasures at a site or countermeasures across a network. Several scenarios exist for an analyst to use Module 3. When an analyst selects a countermeasure for a site based upon the output from Module 2 or personal experience/knowledge, Module 3 can be used to perform an economic appraisal for that particular countermeasure at that specific site. When an analyst selects several countermeasures or combinations of countermeasures for a specific site, Module 3 can evaluate the cost-effectiveness of each countermeasure and combination of countermeasures to determine countermeasure(s) with top priority. When an analyst selects

candidate countermeasures for multiple sites, Module 3 can suggest countermeasures for each site to maximize the net benefits, given budgetary constraints.

The economic appraisal functionality within Module 3 estimates effectiveness in economic terms. The priority ranking functionality within Module 3 provides recommendations on which countermeasures should be implemented across numerous sites for the maximum benefits given certain budget constraints.

Module 4 - Countermeasure evaluation

Module 4 estimates the safety effect of countermeasures implemented at specific sites: a single countermeasure at a specific site or the collective effectiveness of a group of countermeasures at multiple sites. The effectiveness measures are expressed as a percentage change (decrease or increase) in crash frequencies or specific target crash types or a shift in the proportion of specific collision types. The effectiveness of countermeasures is determined through before-after evaluations. The primary statistical approach to performing the before-after evaluation is the Empirical Bayes (EB) technique [5].

The safety management process implemented by the 4 modules can be described by the block cycle diagram shown in Figure 2.1.

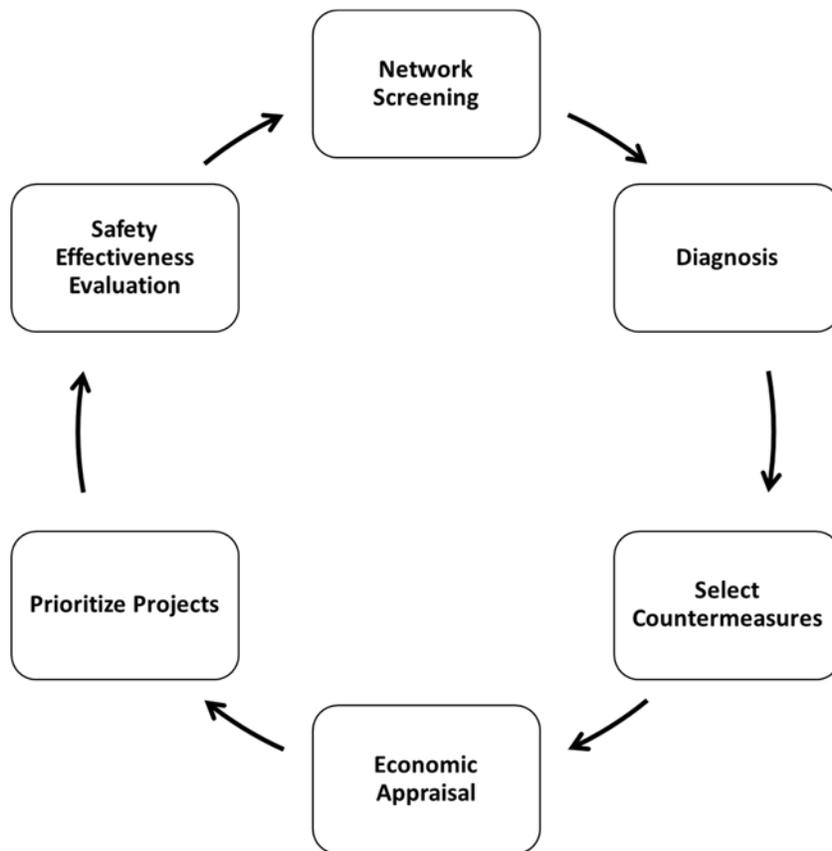


Figure 2-1. General safety management process

Development of a Nevada Statewide Database for Safety Analyst Software

Safety Analyst includes an embedded JavaDB (a.k.a. Apache Derby) database that allows the software to operate as a desktop application. It can also work as a client-server deployment by connecting to a separated database server. Safety Analyst defines user roles of Administrator, Data Manager, and Analyst. Administrators and data managers prepare Safety Analyst and the agency data for use by safety analysts. Administrators install Safety Analyst and configuring system attributes, collision distributions, countermeasures, and diagnostic scenarios. Data Managers configure, import, post process, and calibrate the agency's site data (segments, intersections, ramps, traffic, and crash data). Analysts use the Safety Analyst Analytical Tool to conduct safety analysis on an agency's inventory. While there may be many analysts using the Analytical Tool, it is envisioned that there will be only a few (perhaps one) personnel serving in the Administrator/Data Management role.

Safety Analyst is with five manual documents for detailed information regarding the Safety Analyst architecture and data relationships concerning the individual applications within the Safety Analyst tool set.

- **Administration Tool Manual** - A detailed reference describing the installation and configuration of Safety Analyst, use of the Administration Tool, and customization of Safety Analyst data and components.
- **Data Management Tool Manual** - A detailed reference for using the Safety Analyst Data Management Tool for importing and transforming data into a format usable for Safety Analyst.
- **Data Import Reference** - A detailed reference that describes the data that can be imported into a Safety Analyst database from agency-supplied ASCII import files or agency-maintained databases using the Safety Analyst Data Management Tool.
- **Implemented Countermeasure Management Tool Manual** - A detailed reference for using the Safety Analyst Implemented Countermeasure Tool for importing and transforming countermeasures that have been implemented in the inventory into a format usable for Safety Analyst.
- **Analytical Tool User's Manual** - A detailed reference for using the Safety Analyst Analytical Tool and provides engineering information necessary to make appropriate use of Safety Analyst evaluation capabilities and interpretation of results.

3 LITERATURE REVIEW

This chapter presents a summary of the Safety Analyst utilization status in other states as well as the existing database applications for the deployment of the software. Applications in the investigated states indicate that Safety Analyst provides a comprehensive framework that automates the safety analysis process and supports efficient decision making within an agency. Although the data requirements of Safety Analyst are intense, once the data is imported, the software is able to perform analyses relatively easily requiring minimum statistical expertise. The objectives of the literature review were: (1) to summarize existing Safety Analyst applications in other states and lessons learnt; and (2) to investigate the database prepared for the software and methodologies or tools used to prepare the database. Previous research and states' applications provide valuable information and reference for this research project.

3.1 SAFETY ANALYST APPLICATIONS

As of 2014, eight states including Nevada and eight universities have Safety Analyst licenses. Nevertheless, not all these agencies have the database needed to take full advantages of the software.

Nevada

To assist the implementation of the software in Nevada, the UNLV research group developed a database system for Clark County. The database integrated data from various sources, including the Highway Performance Monitoring System (HPMS), Linear Referencing System (LRS) of road network, Travel Demand Models (TDM), Nevada Accident and Citation Tracking System, and Traffic Records Information Access (TRINA). The research team collected most of the missing data (e.g., ramp type, ramp configuration, type of control at intersections) using Google Earth.

The research team found several issues in the datasets:

- There are spatial gaps among GIS data files of multiple datasets including HPMS, CDS road network and TDM layers;
- There is no unique route master ID among datasets; and
- There is no common ID among the HPMS, CDS road network and TDM layers.

The team developed several customized tools using ArcGIS ModelBuilder to solve the issues found in the original data sources. Three primary tools used include:

- A mapping tool that maps road network segments spatially to data elements in HPMS when there is geometry shift and no common field between them;
- A linear referencing tool that creates a milepost for each crash with respect to roadway segment, ramp, or intersection mileposts; and
- A dynamic segmentation tool that breaks/join the segments at required locations such as at intersections and freeway influence zone.

The developed database system is relatively comprehensive and for NDOT staff to conduct countywide safety analysis. However, there are several concerns in regards to the developed database. First, the data quality in the developed database needs to be assessed before running the software. Some inconsistency was found among the congregated datasets. Additionally, the database is limited to Clark County and not all facility types in the county were included in the database. Further data collection efforts are required.

Ohio

The Ohio Department of Transportation (ODOT) is one of the agencies leading the effort to integrate Safety Analyst into the state safety programs [7]. ODOT uses Safety Analyst to assist all steps of its safety management system including network screening, diagnosis, countermeasure selection, economic appraisal, prioritization, and countermeasure evaluation. Similar to other DOTs, most of ODOT data is already available and maintained in-house. ODOT staff customized tools to automate data mapping transformation to ensure that the DOT data is properly formatted for the software. Currently, ODOT is putting significant efforts into obtaining additional data and improving data quality to increase the reliability of Safety Analyst results. ODOT also extends the utilization of Safety Analyst to its Central and District offices, which allows the advanced safety methods to be employed for the local roadway system.

ODOT's experience proved that Safety Analyst is useful for the safety management process. Nevertheless, ODOT revealed several hurdles when deploying Safety Analyst for the safety program.

- 1) develop a systematic process to aggregate all of the required data into the software;
- 2) create a process to address data consistency and use between offices as database updates occur;
- 3) address the logistics and cost to make the software available in all the District offices; and
- 4) offer more training so District personnel can apply the software correctly and interpret the results [7].

These are the common issues for all other states to deploy Safety Analyst.

Wisconsin

The University of Wisconsin-Madison conducted a project to evaluate Safety Analyst software for the Wisconsin Department of Transportation (WisDOT). The goal was to implement and test Safety Analyst using Wisconsin data, evaluate the analytical functions and validate the results [8]. The project included two parts. The first part compared the data elements required by Safety Analyst and data availability at WisDOT. It suggested the best data import procedure in many options available in Safety Analyst, as well as the step-by-step guidance to import data into the software. The second part focused on the evaluation and assessment of individual modules and their functionalities. The analysis compared the results from Safety Analyst and the results of WisDOT practices. WisDOT data (e.g., road, traffic and crash data) were all stored in excel files with the CSV format. The data was uploaded to an Oracle database, and then imported into

another schema using the Safety Analyst Database-to-Database mapping. More specifically, data in CSV files were loaded onto an input loader Oracle database called “SALDR” which would be placed in an output production database “SAPRD” by the Data Management Tool in Safety Analyst using the custom-built XML Mapping Scheme. In the process, missing data elements from the initial CSV files were identified. The main issues encountered during the data import process included (1) software issues; (2) missing data issues; and (3) data structure issues. The missing data elements caused issues in data importing process and affected the performance of analytical tools in Safety Analyst. Gaps of intersection data included the missing or incomplete traffic control information and Minor Street AADT values. Missing AADT data, median type, and access control information were found in the roadway data. The original data with WisDOT were not in the format required by Safety Analyst. Thus, the research team created a database in Oracle to conform the data to requirements of Safety Analyst [8]. The data structure and formatting issues require careful verifications.

In short, the data initiation process resulted in some problems and challenges mainly because of the shortcomings in the data. The research team also claimed that successful data import does not automatically guarantee that all the data is usable by the software until post processing and calibration procedures. Safety Analyst evaluation for WisDOT indicated that Safety Analyst provides excellent and diversified results for evaluating countermeasures to improve traffic safety. The analyst is provided with many options and choices to analyze proposed countermeasures and select the best options comprehensively.

Florida

Florida is one of the first states that employed Safety Analyst. In 2009, the research team at the University of South Florida developed an interface between Florida DOT’s Crash Analysis Reporting (CAR) system and the Safety Analyst software [9]. The interface is known as the Safety Analyst Data Converter (SADC). The tool converts the CAR data to the data format required by Safety Analyst. SADC contains four parts as shown in Figure 3-1. Available source databases are in the CSV format. Florida also performed SPF case studies to evaluate the accuracy of the Safety Analyst SPFs. The studies found that state-specific SPFs fit the data better than Safety Analyst default ones that were calibrated with the local data [9].



Figure 3-1 Interface of Florida’s SADC Tool

In 2012, the research team at the Florida International University published the research work of standardization of crash analysis in Florida [10]. In this study, the authors conducted a statewide survey about crash analysis methods and tools. Most of the districts in Florida were not confident in utilizing the HSM and Safety Analyst due to the extensive requirements of data and statistical/software expertise. Nevertheless, local agencies are interested in adopting Safety Analyst and wish that the software to be provided for free with low-cost training courses.

California

The California Department of Transportation (Caltrans) assessed the cost for Caltrans to use Safety Analyst. Caltrans considers Safety Analyst as a potential safety analysis tool for California. In the application of Safety Analyst, the California Highway Safety Information System (HSIS) data were formatted to meet the requirements of Safety Analyst. The assessment concluded that it would be desirable for Caltrans to adopt Safety Analyst to improve the efficiency of their network screening and their safety management process as a whole [11]. The project report documented the Caltrans data elements matching the required input of Safety Analyst.

Oregon

The Oregon Department of Transportation evaluated whether Safety Analyst is compatible with the state databases [12]. The evaluation found that the Oregon crash data could be converted to the required Safety Analyst format, although many variables cannot be completely populated as the software definitions. The data requirements of Safety Analyst is more extensive than the state data. Nevertheless, the Oregon data is reasonable compatible and can be used in Safety Analyst with certain formatting adjustments. The evaluation report emphasized the importance of the minor road traffic data and concluded that it is a sensitive data element for intersection evaluation. The report also documented in detail whether a data element in the Oregon database

can be adopted in Safety Analyst and if not what type of adjustment is needed. The research team developed a subset of the entire state database according to the data requirements of Safety Analyst. The subset database was used to evaluate the practical application of the software. Oregon's application confirmed that Safety Analyst is a powerful tool and recommended to the state agencies. Nonetheless, the project did not thoroughly introduce the data import process.

Georgia

The Georgia Department of Transportation (GDOT) performed case studies to evaluate the feasibility of applying Safety Analyst throughout the state [13]. Similar to the experience of other DOTs, GDOT also concluded that data preparation was very tedious, as it required a significant amount of data and significant preparation effort to meet the stringent requirements of the software. Nonetheless, the report did not provide detailed discussions about the methodologies or tools used to recode the data. Several errors were encountered during the data import and processing. For instance, some crashes were not located on any roadway segment, no traffic data was associated with some roadway segments, traffic data or growth factors were unrealistic, and segments were not assigned to any subtype. Some of the errors are due to incorrect or missing data elements. The researchers concluded that Safety Analyst produces much more reliable results and is a superior method compared to the traditional methods. One recommendation is to start with one county or one road type to use only segment data because segment data is easier to manage and prepare for importing. Also, it recommends using state-specific SPFs when available.

Virginia

The Virginia Department of Transportation (VDOT) conducted research to develop SPFs for Virginia using the annual average daily traffic (AADT) as the most significant causal factor and compared the results to the default ones in Safety Analyst [14]. Safety Analyst calibrates SPFs automatically with the state data. It was found that the Virginia-specific SPFs fit Virginia data better than the Safety Analyst models. The study suggested that the DOT should use Virginia-specific SPFs when using Safety Analyst tools. Nevertheless, this study did not focus on the implementation of Safety Analyst in the state.

Utah

Presently the Utah Department of Transportation (UDOT) is not planning to adopt the Safety Analyst software for the state, but Peter Kelly discussed traffic safety analysis of Utah county highways using Safety Analyst in his thesis [15]. The author concluded that the data obtained from UDOT underwent extensive preparation to be formatted properly for the use with Safety Analyst. The roadway inventory data available in UDOT is mostly in the form of GIS shape files that contain tabular and spatial data. The author used the Linear Referencing Tools and Dynamic Segmentation approach to overlay the shape files into one roadway layer. The dynamic segmentation process yielded a roadway segment dataset that contains all the necessary data elements for importing into Safety Analyst. Besides, the crash data files were aggregated into one file using VBA codes in excel. All the data were imported to Safety Analyst from the CSV files.

Some missing data elements such as median type and interchange influence area were initially input using default values. After manually gathering these data of Utah county highways, the database was updated. The initial default values and the collected data yield very similar results of the network screening in Safety Analyst. The author recommended that the maximum amount of data elements should be prepared and imported for use in Safety Analyst to obtain the best possible results. It also recommended that separated databases for roadway segments and intersections should be prepared.

Summary

Several states have studied the feasibility and challenges of utilizing the Safety Analyst software in their highway safety management procedures. The studies concluded that overall Safety Analyst provides excellent and diversified results for evaluating and recommending countermeasures. It is also desirable for the states to adopt Safety Analyst to improve the efficiency of their network screening and their safety management process. Most states emphasized the importance of the data quality and data collection efforts. Nevertheless, the raw data sources are different in reviewed states, and the tools developed by other DOTs cannot be directly adopted for NDOT data integration.

3.2 METHODOLOGIES AND TOOLS TO PREPARE DATASETS

The federal and state governments have been spending considerable resources to build accurate and timely safety database at both national and state levels. The availability and quality of transportation data is a cornerstone of any data-driven program including Safety Analyst [16]. Safety Analyst requires a significant amount of data, and the data must follow a particular format. For example, the software requires crash severity type in the form of K for fatal, A for severe injury, P for property damage only. Preparing the data to follow a particular formatting is one of the primary barriers for DOTs to use Safety Analyst [17]. Data preparation plays a crucial role in the utilization of the software. Wrong data format or structure, missing data elements, or incorrect data information may cause the study sites to be invalidated, and hence the analysis cannot be performed on the sites.

Data elements from multiple sources require review, processing, and formatting to ensure consistency with Safety Analyst. The data integration process combines and links two or more datasets from different resources to facilitate data sharing, promote effective data gathering and analysis and support overall information management activities [18]. Based on the states' applications summarized in the preceding sections, divisions of state DOTs collect and maintain various datasets based on their corresponding data needs. Consequently, the approaches to preparing and congregating data vary in different states.

Safety Analyst provides several methods to import agency data depending on the format and availability of data. Previous research reported the development of data collection and integration methods for transportation applications, including (1) GIS frameworks; (2) Spatial data warehouse; and (3) customized converting tools and visualizations tools. Below is a summary of these methods and tools.

GIS Frameworks

GIS technologies are widely used in storing, analyzing, and visualizing traffic data [8, 9, 11, 13]. Pendyala et al. developed the GIS-based conflation tools for data integration and matching [19]. This research developed framework to match road networks that come from different sources or have been created at various times. The authors emphasized the fact that the common way to integrate data is to establish correspondences (e.g., shared route ID, etc.) among different networks and reference data to integrate as one data layer.

The linear referencing tools in GIS software are often used to achieve dynamic segmentation and overlay different GIS layers. The GIS shape files cannot be directly imported into the Safety Analyst database, so GIS data need to be converted to CSV files or relational database before being imported into Safety Analyst.

Spatial Data Warehouse

A spatial data warehouse is closely related to GIS technologies and data integration methodologies. For instance, O'Packi et al. introduced Maine DOT's (MeDOT) experience in regards to the development of data warehousing [20]. The study overviewed MeDOT's GIS-linked data warehouse named Transportation Information for Decision Enhancement (TIDE). This tool effectively facilitated the growth of agency integration. Similarly, Hall et al. introduced the spatial information system infrastructure implemented by the Illinois DOT (IDOT) to enable delivery of information to management decision makers in asset management applications [21]. This spatial data warehouse infrastructure makes extensive use of GIS technologies to integrate information from a variety of database structures and formats. The data warehouse was achieved by embedding underlying link-node structure into roadway inventory databases and enabling the direct linkage of data through various system identifiers such as different milepost referencing and project numbering schemes.

Customized Tools

Customized tools can be used to integrate data, transform data to Safety Analyst standards, or perform additional analysis. Florida's SADC Tool as shown in Figure 1 comprises several modules to convert different categories of data to meet Safety Analyst's requirement. Izadpanah et al. developed a Utility Tool to generate input files following the Safety Analyst standards in Ontario, Canada [22]. Infrastructure characteristic data, traffic volume data, and crash data were converted into the input files with this tool. Figure 3-2 shows a view of the tool. The tool is capable of incorporating all the mapping rules to generate the standard data format. As introduced earlier, the standard Safety Analyst input files are CSV format or XML format. This study recommended using the CSV formats considering the sizes of the XML files tend to be large which slows down the import process. In Ontario, the software and its databases were hosted on the Cloud server, to keep the software databases updated for the entire province.

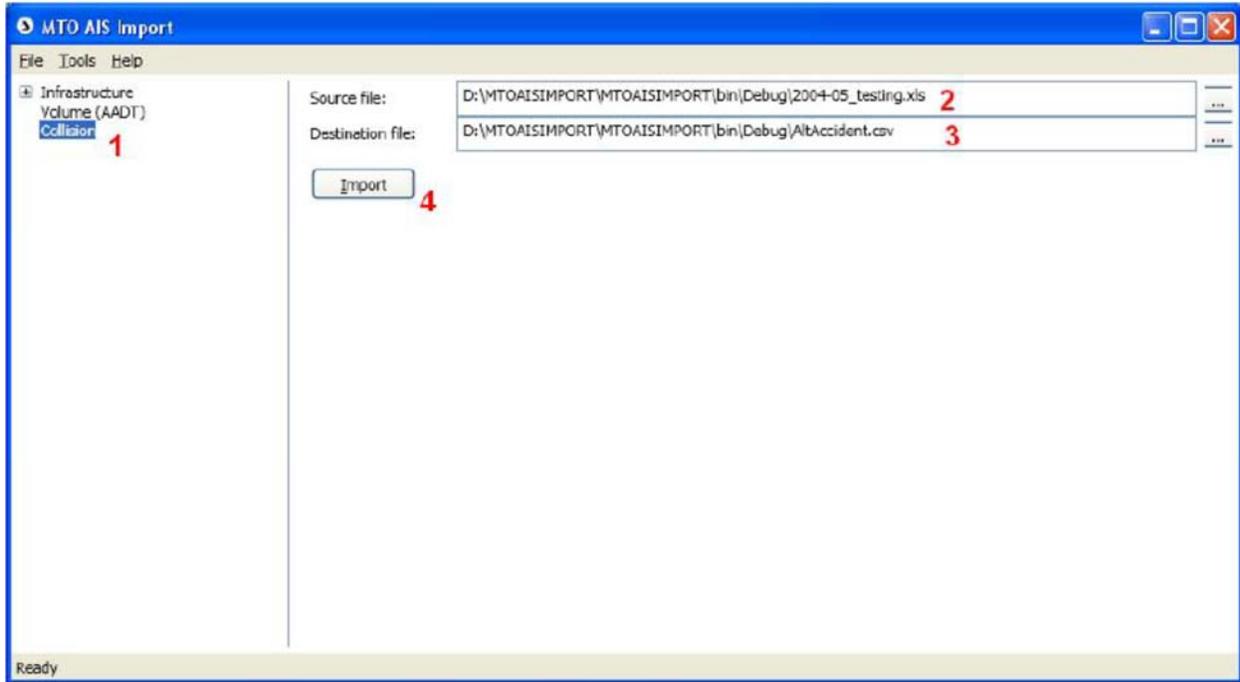


Figure 3-2 Interface of Data Import Utility Tool in Ontario

To conclude, the methods to integrate data from different sources vary among investigated states. Most states utilize GIS software and tools to store and organize their data. The linear referencing method is often used to combine and overlay data elements from multiple sources. The data stored in GIS files or data warehouse system needs to be converted to Safety Analyst standards in order to be successfully imported into the software. The conversion can be achieved using data management tools in GIS, VBA programming in excel, or other customized tools developed to meet states' needs. The data transformation process may be tedious if numerous recoding is required. Therefore, innovative techniques and tools would promote the efficiency of data conversion and update process.

3.3 FINDINGS AND CONCLUSIONS

Safety Analyst as a powerful analysis tool can assist transportation engineers and researchers in implementing HSM and enhance statewide safety. Having a comprehensive database for the entire state facilitates the use of both Safety Analyst and the HSM. Several state DOTs have evaluated the Safety Analyst software and concluded that the software has a significant advantage compared to other tools and can quickly apply the more sophisticated screening methods within the HSM. With the statewide database, the software can analyze sites based on specific facility type and different screening methods.

Safety Analyst requires integrating, matching and merging disparate data sets and road networks. A significant amount of data in a certain format is required by the software. The extensive data gathering effort is expected to increase the reliability of safety analysis in Nevada and determine the sites for safety improvement and countermeasure more cost-efficiently. The tools and data

Development of a Nevada Statewide Database for Safety Analyst Software

processing methods developed in the user states were considered in this project. However, UNR CATER could not directly use any of those tools, which was determined by the tool availability, the data availability, and formats in Nevada. As NDOT manage road, traffic and crash data with GIS software and database, this project employed geoprocessing functions of ESRC ArcMap, especially the linear referencing toolbox.

4 DATA REQUIREMENT AND MANAGEMENT

4.1 OVERVIEW OF DATA REQUIREMENT

The analytical procedures in the four Safety Analyst modules including network screening, diagnosis and countermeasure selection, economic appraisal and priority-ranking module and countermeasure evaluation, utilize several types of data. This section documents the types of data used and maintained within the Safety Analyst database. Safety Analyst uses two types of data – agency data and data for computational purposes.

Agency Data

The agency data include characteristics of sites (roadway segments, intersections, and ramps), traffic, crashes, and implemented countermeasures, which are imported from existing data sources or collected by agencies. For each site, the database record contains geometric, traffic control, and traffic volume data and location identifier data to link these site properties to a location on the highway network. There are four basic systems of location identifier used by most highway agencies and supported by Safety Analyst.

- Route/county/milepost (*Location identifier used by NDOT*)
- Route/milepost
- Route/segment identifier/distance
- Segment identifier/distance

The characteristic data elements are classified as either mandatory (i.e., required) or optional variables:

Mandatory variables represent the minimum and necessary data requirements for utilizing Safety Analyst. Each mandatory variable is used in at least one specific procedure or set of logic. Where mandatory variables are missing, some sites may need to be omitted from consideration by Safety Analyst procedures. If the same mandatory variables are missing for all locations, some types of analyses may not be able to be performed. For example, if no intersection data are available, Safety Analyst procedures cannot be applied to intersections. If data are available for some intersections, then Safety Analyst procedures can be applied only to those intersections with data available.

Optional variables represent a set of variables that are desirable to include in the Safety Analyst database whenever the variable is available. Obtaining and entering these data into the Safety Analyst database enhances the usability of Safety Analyst.

The Safety Analyst Data Management Tool Manual [6] presents a dictionary that includes a description of all the site characteristic data elements stored in the Safety Analyst database. Within the Data Management Tool Manual, all data elements are identified as either being mandatory (i.e., required), optional, or derived by Safety Analyst. The characteristic elements reflect the status of the geometric design features of a site.

The Safety Analyst database includes data elements that characterize the type of crash and data to link the crash to a specific location on the highway system (i.e., the location identifier variables). The crash data elements included in the Safety Analyst database can be broadly categorized as the crash, vehicle, and person data elements.

The Safety Analyst database contains data about the construction or improvement history of sites. Safety Analyst uses the data to determine the crash history to be analyzed at a site. An analyst can limit the analysis period by excluding years before major reconstruction, to avoid miscalculation of the expected crash. The implemented countermeasure data can be used to determine sites for inclusion in a countermeasure effectiveness evaluation. The Implemented Countermeasure Management Tool Manual [6] describes all the data elements related to the construction or improvement history of a site. The countermeasure data is not necessary for network screening or safety diagnosis, and there is no existing state database to maintain the statewide construction data. Therefore, it is not included in the database developed by this project.

Data for Computational Purposes

The data maintained for computational purposes either are default values within the Safety Analyst program or are calculated during the data import process. The data for computational purposes are not about individual sites but pertain to a collection of sites or all sites. The elements are listed as follows:

- Safety performance functions (SPFs)
- Crash proportions
- Countermeasure defaults
- Crash costs
- Equivalent Property Damage Only (EPDO) weights
- Beta distribution parameters
- Other defaults

The key data required for performing safety management analysis are the inventory, traffic volume, and crash data. Without these three sets of data, safety analysis cannot be conducted, and there would be no reason to employ Safety Analyst. Furthermore, these data must be converted from their raw form to a format suitable for use in Safety Analyst. Once converted, these data form a Safety Analyst dataset. A single dataset may contain a substantial amount of data. However, Safety Analyst supports the use of multiple datasets. For example, an agency may choose to separate their inventory into different datasets based on local roads and state roads. An agency may maintain different datasets for different years.

4.2 DATA MANAGEMENT

The workflow for creating a Safety Analyst dataset comprises six steps, five of which are performed with the Data Management Tool of Safety Analyst:

- 1) Data collection

- 2) Creating a dataset
- 3) Importing data into a Safety Analyst database
- 4) Post processing the Safety Analyst database
- 5) Calibrating the Safety Analyst database
- 6) Serving the Safety Analyst database

The descriptions of the steps in the following sections refer to the data manager. If a dataset is stored on a remote DBMS, the data manager must have access to a database account that has privileges to perform insert, delete, update, and select operations on the dataset tables in the database.

4.2.1 Step 1: Data collection

Of the six steps, this is the only step that cannot be performed using the Data Management Tool. In some agencies, this is also the most time-consuming step in the workflow. In some agencies, all of the data required by Safety Analyst may reside in one data store, and conform to a well-documented format. In other agencies, the data may be dispersed across several data stores, each using different formats. In some agencies, the data may not even be stored in electronic format. For a data manager to begin this step, the Data Import Reference Manual [6] is recommended as a reference. This reference describes in detail the data that can be imported into a Safety Analyst data set. Not every data item defined in the Manual is required. The Manual identifies those elements that are required, as a minimum, for importing data into Safety Analyst. Every data item marked as needed for import must be available to create a usable data set.

The collection process results in a collection of all required data, stored in an electronic format that can be imported into Safety Analyst using the Data Management Tool.

4.2.2 Step 2: Creating a dataset

Safety Analyst uses a database to store and access datasets. A database can be an "embedded" one stored locally on the platform hosting the Data Management Tool, or it can be a remote database server. For an embedded database, Safety Analyst uses the JavaDB (a.k.a. Apache Derby) database that is freely available and packaged with the Java Runtime Environment (JRE). This database is included with the Safety Analyst installation. If an agency chooses to serve a Safety Analyst database on a remote database server, the agency can still develop the dataset using the embedded Derby database. Once the dataset is fully defined and calibrated, the agency can then copy the database to the remote server as described in the Serving a Dataset step (Step 6).

For remote databases, Safety Analyst can use any DBMS that supports the Java Database Connectivity (JDBC) standard. Furthermore, this generic database interface has been tested with Safety Analyst on several well-known and widely deployed relational database management systems, including:

- Oracle
- Sybase
- Microsoft SQL Server

- MySQL
- JavaDB (a.k.a. Apache Derby)

When using a remote database to create a Safety Analyst dataset, the data manager needs to know the database connection information, such as host name, port, and database system ID. To accommodate the policies in place at different agencies, the Data Management Tool supports two modes of creating a remote dataset database. If the data manager has data definition language (DDL) privileges on the remote database, the Data Management Tool can be used directly to create the dataset tables in the database. If not, the Data Management Tool provides an option to generate SQL script files that can then be run by an authorized database administrator to create the dataset tables.

4.2.3 Step 3: Importing data into a dataset

The primary function of this step is to import the inventory, crash, and traffic data specified by the agency in Step 1 into the database created in Step 2. Safety Analyst currently supports two basic mechanisms for data import:

Database-to-database mapping

For agencies that maintain their inventory, traffic, and crash data in a SQL-compliant DBMS, Safety Analyst provides a database-to-database mapping capability to extract the required data directly from the agency data stores. Using the map specification interface provided by Safety Analyst, the data manager creates a map that associates each Safety Analyst data item to one or more data elements in the agency data stores. During import, Safety Analyst uses the map to extract data from the agency data stores, convert the data as necessary to create the Safety Analyst data, and then stores the data in the Safety Analyst database.

File import

The file import capability supports agencies that do not use a SQL-compliant DBMS to maintain their inventory, traffic, and crash data (or whose database structures do not conform to the basic requirements for creating a map). The Data Management Tool accepts four file formats for import into Safety Analyst:

- XML files conforming to the Safety Analyst Standard Import schema
- XML files conforming to the Safety Analyst Alternate Import schema
- XML files conforming to an agency-specified schema
- CSV files

Of these file formats, the CSV format is the easiest to use and is the most practical and efficient when importing large amounts of data. Although a set of files, each using a different format, can be imported, it is more common for the set of import files to conform to a single format.

Once an import mechanism is chosen and the map or import files have been created, the data can be imported. The import process supports an option to merge new records into an existing dataset. For example, if the existing dataset contains crash and traffic data for the years 2000-

2007 and the import files contain new data for 2008, the new data can be added to the existing dataset without having to reimport the existing data.

The Data Management Tool provides an export utility that enables the data manager to export an existing dataset into files that conform to one of the Safety Analyst-defined XML schemas or into the CSV format. Furthermore, the Data Management Tool allows the creation of CSV files from a newly created (empty) dataset. These files contain no data but contain the header-row item names. This feature can be used to create file templates for producing the CSV files for data import.

4.2.4 Step 4: Post processing a dataset

Once the agency data are successfully imported into the dataset, the data must be post-processed for use in Safety Analyst. During post processing, the Data Management Tool automatically performs the following tasks:

- connects segments to each other and intersections
- locates crashes on the inventory (segments, intersections, and ramps)
- assigns site subtypes to the inventory
- optionally combines segments into homogeneous segments
- validates traffic volume data
- performs other data validation

The statistical analysis by the Safety Analyst Analytical Tool produces reliable results without very short segments in the database. Thus, the data manager may select an option to aggregate individual, contiguous, homogeneous segments into longer segments. The Data Management Tool provides a user interface to configure when two segments are considered homogeneous and can be combined. A safety engineer within the agency determines whether homogeneous segment aggregation is needed and what values should be used for the aggregation parameters.

During post processing, if an inventory element does not have any valid traffic data, it is rejected (removed from the dataset). If traffic data are available but incomplete (e.g., missing years), the post processing interpolates or apply a growth factor to estimate traffic counts. Also during this phase, crashes not located on one of the valid inventory elements are rejected.

4.2.5 Step 5: Calibrating a dataset

An imported and post processed dataset is still not ready for use until it has been calibrated. During calibration, crash distributions are computed for the inventory based on site subtype and crash severity. Also during this phase, dataset calibration factors are computed for the Safety Performance Functions employed by the Safety Analyst Analytical Tool. The Data Management Tool provides an interface to specify thresholds used in the calibration process. These thresholds control which sites are used to compute calibration factors and distributions, and prevent calibration of site subtypes that are not well represented (too few sites) in the agency's data set.

4.2.6 Step 6: Serving a dataset

Once a dataset has been calibrated, it is ready for use by the Analytical Tool. However, before the Analytical Tool can access the dataset, it must know where it is located. This depends on how the dataset database was created and how it will be served. There are four possible options:

- 1) Created locally (embedded), accessed locally
- 2) Created locally (embedded), accessed via remote DBMS
- 3) Created on remote DBMS, accessed locally
- 4) Created on remote DBMS, accessed via remote DBMS

The Data Management Tool provides a mechanism for generating an installer that can be used to install the Safety Analyst Analytical Tool on analysts' computers. The installer generator can be configured to support each of the aforementioned options. It can also be configured to just install a dataset database or the connection information file for dataset updates on local platforms that already have the Analytical Tool installed.

Option 1: Locally created, locally accessed

For a locally created database that will be accessed locally, the installer generator of Safety Analyst can create a Safety Analyst installation including the Analytical Tool and the specified datasets.

Option 2: Locally created, remotely accessed

For a locally created database that will be accessed via a remote DBMS, the data manager needs to create an empty database on the remote DBMS. The copy utility of the Data Management Tool can copy the locally created dataset to the remote DBMS database. The installer generator can then be used to create a Safety Analyst installation including the Analytical Tool and a file that contains the connection information for the dataset on the remote DBMS.

Option 3: Remotely created, locally accessed

For a dataset database that was created on a remote DBMS and that will be accessed locally, the data manager must create an empty embedded database on the platform that hosts the Data Management Tool. The copy utility of the Data Management Tool can copy the dataset on the remote DBMS to the local database. Then, the installer generator can create a Safety Analyst installation including both the Analytical Tool and the specified datasets.

Option 4: Remotely created, remotely accessed

For a dataset database that was created on a remote DBMS and that will be accessed via the same remote DBMS, the installer generator can create a Safety Analyst installation including the Analytical Tool and a file that contains the connection information for the dataset on the remote DBMS..

4.2.7 Copying Datasets

In many instances, it may be more efficient to use an embedded (local) database to import, post-process, and calibrate a Safety Analyst dataset. This is particularly the case when first using

Safety Analyst to develop a dataset, a process that may require multiple iterations to get the data mapping correct. It is also the case when an agency has a restricted enterprise database where the Safety Analyst data manager has restricted rights on the remote database. For these situations, it may be easier to create a dataset locally and then copy the dataset to the remote database.

The copy utility can be used to copy a dataset from any database to another, regardless of the type (embedded or remote) or location of the database. A dataset can be copied regardless of its state: created, imported, post-processed, calibrated.

4.2.8 Reporting and Exporting Datasets

The Data Management Tool provides two capabilities that are not required steps in creating or maintaining a data set. However, these capabilities may be useful in the process.

Generating a dataset report

Using the Data Management Tool, a set of report files can be generated for a dataset. The report files include a summary of the contents of the dataset. Additional files format text descriptions of each of the inventory elements in the dataset. The tool provides an option for generating the summary only.

Exporting a dataset

Once a dataset has been created, it can be exported to a set of files in one of three formats:

- Comma-separated value files
- XML files using the Safety Analyst standard import schema
- XML files using the Safety Analyst alternate import schema

A dataset in any state (i.e., created, imported, post-processed, calibrated) may be exported, even though a dataset in the created state contains no data. Exported datasets in any one of these three formats can be imported into Safety Analyst using the Data Management Tool.

5 PREPARATION OF NEVADA SAFETY ANALYST DATABASE

5.1 DATA SOURCES WITH NDOT

To build the Nevada Statewide Database for Safety Analyst, UNR CATER collected road, traffic, and crash data from different NDOT divisions. The data collection is to meet the minimum data requirement of Safety Analyst with the mandatory data and fill as much optional data as possible. This section presents the data collected in this project. The available NDOT datasets for Safety Analyst are mainly in the format of GIS layers, including roadway segment characteristics data, intersection characteristics data, ramp characteristics data, traffic volumes, and crash data. The final dataset includes the road segment table, the intersection table, the ramp table, and the crash table. A GIS dataset was also built for the final Safety Analyst database with the ESRI ArcGIS software, which can be used to extend applications of the statewide dataset.

5.1.1 Roadway segment data

Different road network data were obtained from NDOT divisions. The received data types, formats, and sources are listed in Table 5-1.

Table 5-1. Roadway segment data from NDOT

Name	Format	Division/Agency
HPMS 2014 data	Geometric layer	NDOT Traffic Safety Engineering
Urban/rural boundary	Geometric layer	NDOT Traffic Safety Engineering
HPMS AADT layer	Geometric layer	NDOT Roadway Systems
TRINA AADT	Geometric layer	TRINA Website
Clark County road data (UNLV)	Geometric layer	NDOT Traffic Safety Engineering
NDOT supplemental layers	Geometric layer	NDOT Traffic Safety Engineering

HPMS Data

The Highway Performance Monitoring System (HPMS) is a national level highway information system that includes data on the extent, condition, performance, use, and operating characteristics of the nation's highways. The HPMS contains administrative and extent of system information on all public roads, while information on other characteristics is represented in HPMS as a mix of universe and sample data for arterial and collector functional systems. Limited information on travel and paved miles is included in summary form for the lowest functional systems. The HPMS data are used for assessing highway system performance under the U.S. DOT and FHWA's strategic planning and performance reporting process. Each State is required to furnish all data annually per the reporting requirements in the HPMS Field Manual. A total of 11,127-mile roadways are included in the Nevada HPMS dataset.

The data required for the annual submittal of HPMS includes: (1) limited data on all public roadway sections, which includes the Federal-aid system (i.e., Full Extent data), (2) more

detailed data for designated sections of the Federal-aid system (i.e., Sample Panel data), and (3) area-wide summary information primarily for lower functional system roads (i.e., Summary data).

Full Extent Data

Full Extent Data refers to a limited set of data items that are reported for an entire roadway system such as the National Highway System (NHS) or an entire functional system (e.g., Interstate roadways).

Sample Panel Data

Sample Panel Data consists of data items that are reported for a select portions of a given roadway system. The sampled sections are a fixed sample panel of roadway sections that are monitored from year to year and, when expanded, represent the Full Extent of the systems that are sampled. The more detailed information collected for a Sample Panel section is used to represent similar conditions on the associated functional system after expansion.

Partial Extent Data

Partial Extent Data refers to those data items that are reported on a Full Extent basis for some functional systems and on a Sample Panel basis for other functional systems.

Statewide Summary Data

Statewide Summary Data includes information on travel, system length, and vehicle classification by functional system and area type, in addition to land area and population by area type. The area types include rural, small urban, and individual urbanized, non-attainment, and maintenance areas. Pollutant type is also reported as indicators of air quality in non-attainment areas.

Linear Referencing System (LRS) Data

LRS data provides a spatial reference for the Full Extent and Sample Panel data on selected highway functional systems. This spatial data coupling (i.e. representing roadway attribute data in a spatial format) enables the analysis of HPMS data in a GIS environment. Within the HPMS software, the State-provided LRS represents all roadways in a given State's road network for a designated set of functional classifications. Both the geospatial and attribute data contain three referencing elements that are used to perform the linkage for linear features: (1) A unique Route ID, (2) a beginning milepoint, and (3) an ending milepoint. Point features use a route milepoint in place of a beginning and ending milepoint for referencing purposes.

The data items listed in Table 5-2 are the fields in the HPMS dataset.

- Item Number is the number assigned to each data item
- Data Item identifies the type of attribute data to be reported
- Extent indicates if the data item is required for the Full Extent (FE), Sample Panel (SP) sections, or the Full Extent and Ramp sections (FE+R)

Table 5-2. Data Items Required for the Full Extent and/or Sample Panel Sections

Data Item Type	Item Number	Data Item	Extent	
Inventory	1	Functional System	FE + R	
	2	Urban Code	FE + R	
	3	Facility Type	FE + R	
	4	Structure Type	FE**	
	5	Access Control	FE*	SP*
	6	Ownership	FE	
	7	Through Lanes	FE + R	
	8	Managed Lane Operations Type	FE**	
	9	Managed Lanes	FE**	
	10	Peak Lanes		SP
	11	Counter Peak Lanes		SP
	12	Right Turn Lanes		SP
	13	Left Turn Lanes		SP
	14	Speed Limit		SP
	15	Toll Charged	FE**	
	16	Toll Type	FE**	
Route	17	Route Number	FE*	
	18	Route Signing	FE*	
	19	Route Qualifier	FE*	
	20	Alternate Route Name	FE	
Traffic	21	Annual Average Daily Traffic	FE + R	
	22	Single-Unit Truck & Bus AADT	FE*	SP*
	23	Percent Peak Single-Unit Trucks & Buses		SP
	24	Combination Truck AADT	FE*	SP*
	25	Percent Peak Combination Trucks		SP
	26	K-factor		SP
	27	Directional Factor		SP
	28	Future AADT		SP
	29	Signal Type		SP
	30	Percent Green Time		SP
	31	No. of Signalized Intersections		SP
	32	No. of Stop Sign-Controlled Intersections		SP
	33	No. of Intersections, Type - Other		SP

Development of a Nevada Statewide Database for Safety Analyst Software

Geometric	34	Lane Width		SP
	35	Median Type		SP
	36	Median Width		SP
	37	Shoulder Type		SP
	38	Right Shoulder Width		SP
	39	Left Shoulder Width		SP
	40	Peak Parking		SP
	41	Widening Obstacles		SP
	42	Widening Potential		SP
	43	Curve Classification		SP*
	44	Terrain Type		SP
	45	Grade Classification		SP*
	46	Percent Passing Sight Distance		SP
	Pavement	47	International Roughness Index (IRI)	FE*
48		Present Serviceability Rating (PSR)	FE***#	SP*
49		Surface Type	FE***	SP
50		Rutting	FE***	SP
51		Faulting	FE***	SP
52		Cracking Percent	FE***	SP
54		Year of Last Improvement		SP
55		Year of Last Construction		SP
	56	Last Overlay Thickness		SP
	57	Thickness Rigid		SP
	58	Thickness Flexible		SP
	59	Base Type		SP
	60	Base Thickness		SP
	61	Climate Zone		SP
	62	Soil Type		SP
Inventory	63	County Code	FE	
Special Networks	64	National Highway System (NHS)	FE**	
	65	Strategic Highway Network (STRAHNET)	FE**	
	66	National Truck Network (NN)	FE**	
	67	Future Facility (Planned/Unbuilt NHS)	FE**	
Inventory	68	Maintenance and Operations	FE	
Traffic	69	Capacity		SP
Inventory	70	Directional Through Lanes	FE****#	

The received Nevada HPMS data includes the route network in a GIS layer and tables of road characteristics. Each table in the HPMS data includes a road data element and the LSR information for locating the properties onto the specific road segment. The tables of road

characteristics were overlaid to integrate all the road properties. By the end, the integrated road properties were mapped onto the GIS route layer. This integrated HPMS layer is the basis of the generation of the statewide dataset. The other road, intersection, traffic, and crash data were located on to the HPMS road network and integrated with the HPMS data elements. The Nevada HPMS road network is shown in Figure 5-1.

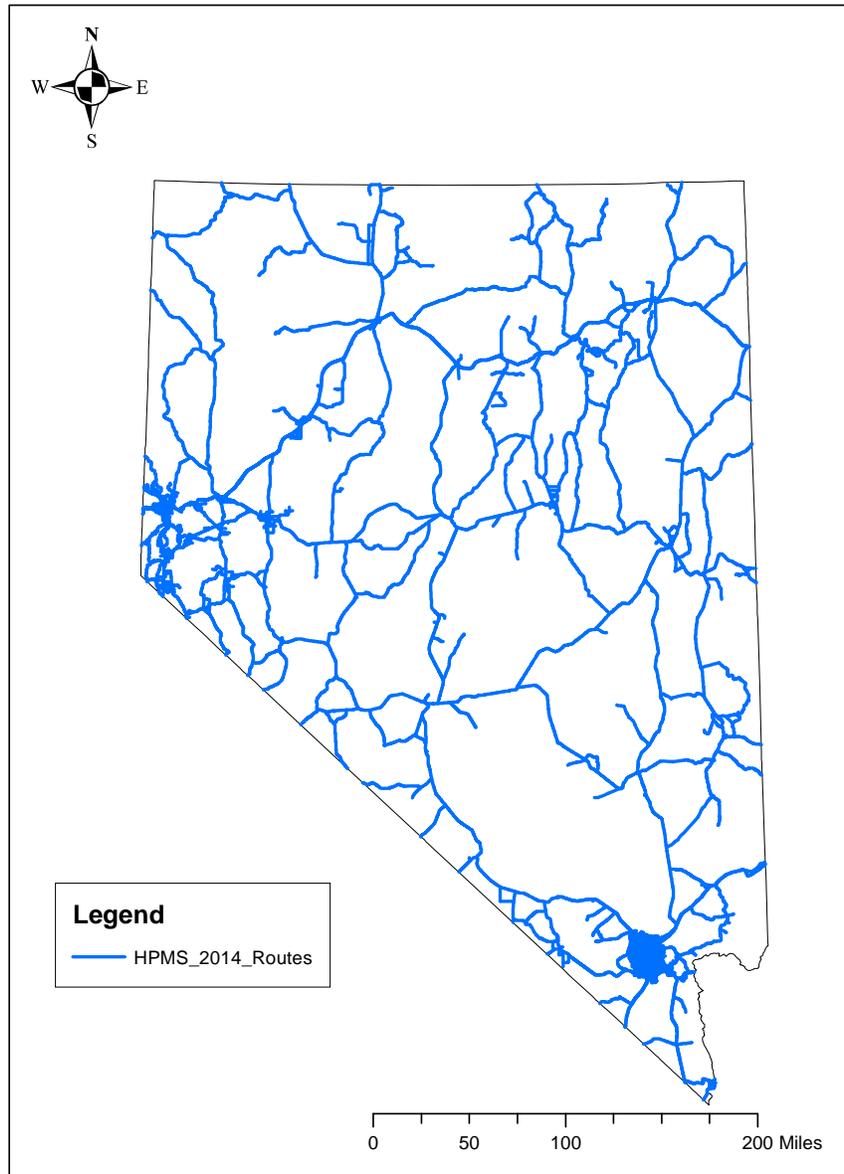


Figure 5-1. Nevada HPMS road network 2014

Urban/Rural boundary

The boundary layer is received from NDOT TSE. This layer contains 17 different urban and rural areas in Nevada. Figure 5-2 shows the map of road boundary layer.

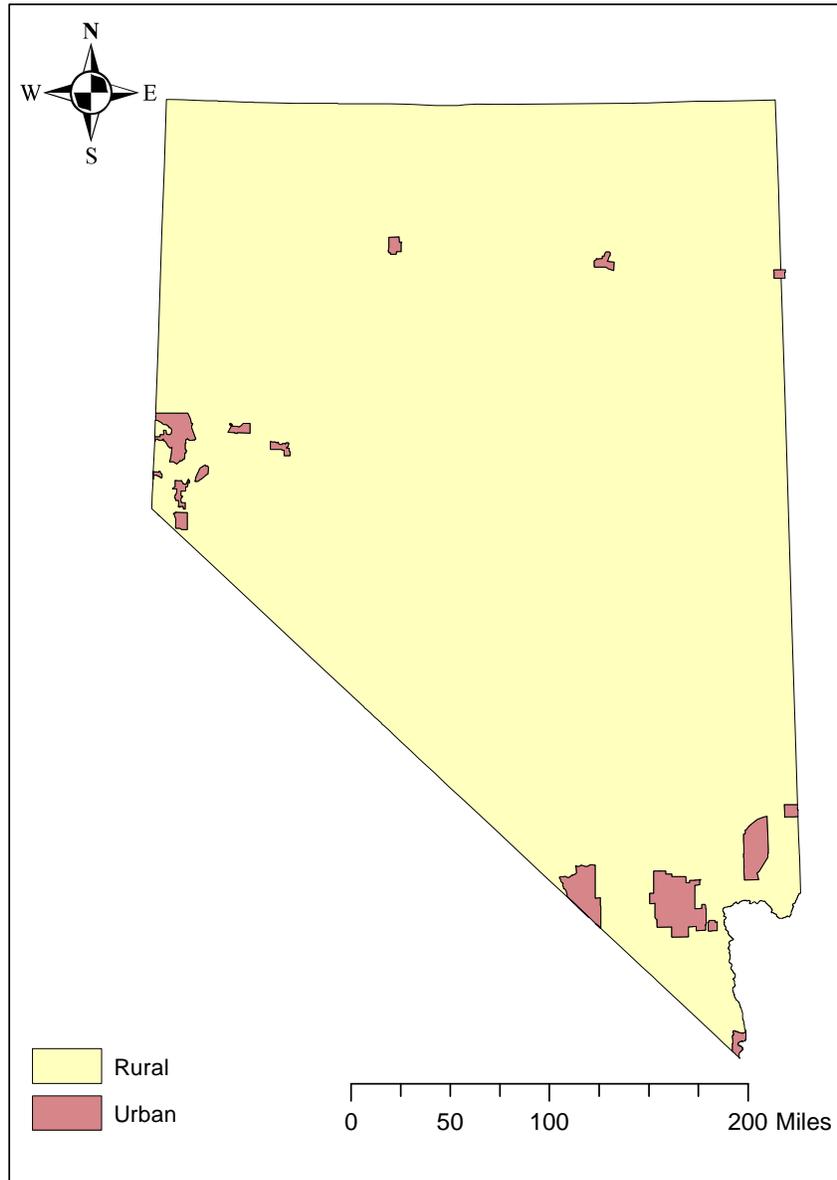


Figure 5-2. NDOT urban/rural boundary data

Historical HPMS AADT Layer

A geodatabase of historical AADT data from 2009 through 2015 was provided by the NDOT Roadway Systems. This historical AADT table can be linked to the NDOT HPMS data using the Route ID, the begin milepost and the end milepost. Figure 5-3 shows the AADT of HPMS route in 2015.

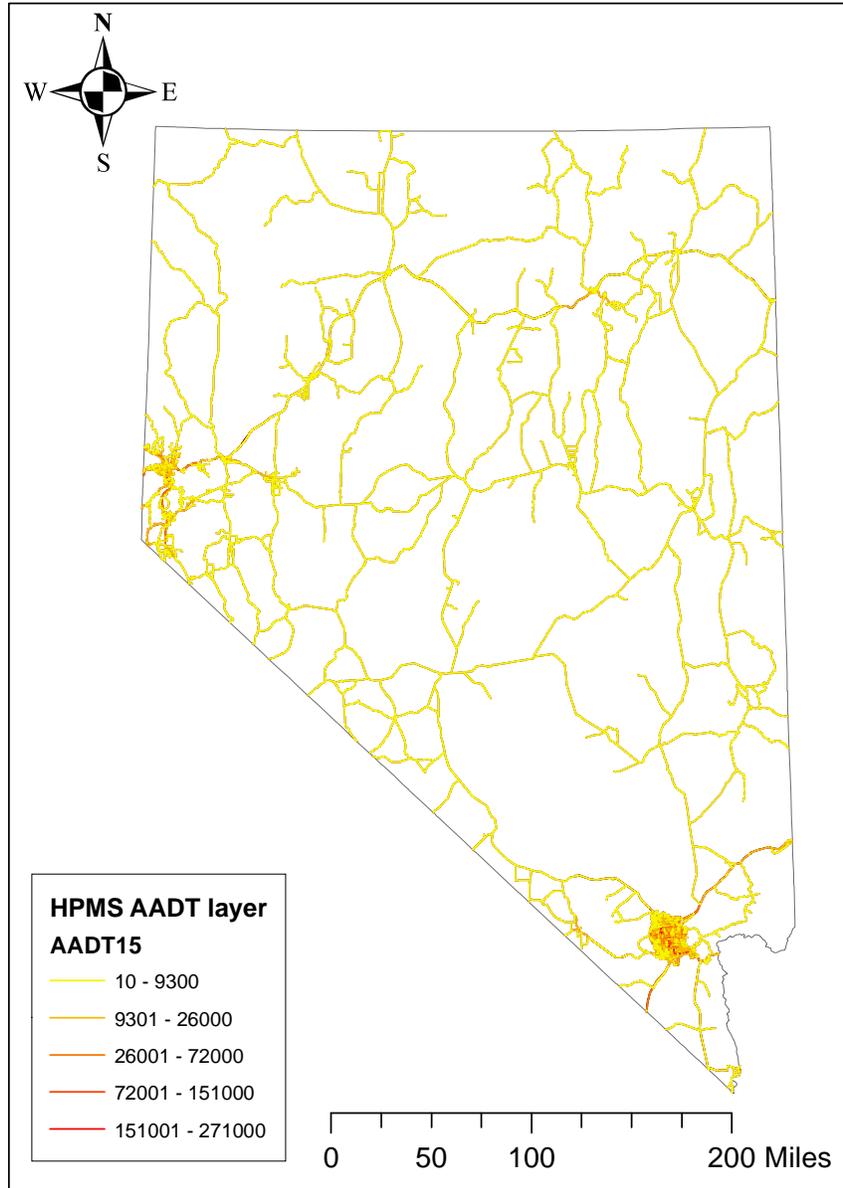


Figure 5-3. NDOT historical HPMS AADT data layer

TRINA AADT

The AADT in TRINA was downloaded from the TRINA website. The AADT data collection stations with TRINA roads were shown in Figure 5-4. The TRIAN AADT route does not include milepost information, and there is offset between the TRIAN road network and the HPMS route layer. These issues lead to uncontrolled errors when TRINA data are used to bring AADT into HPMS layer. So TRINA AADT was not used for AADT assignment in this project.

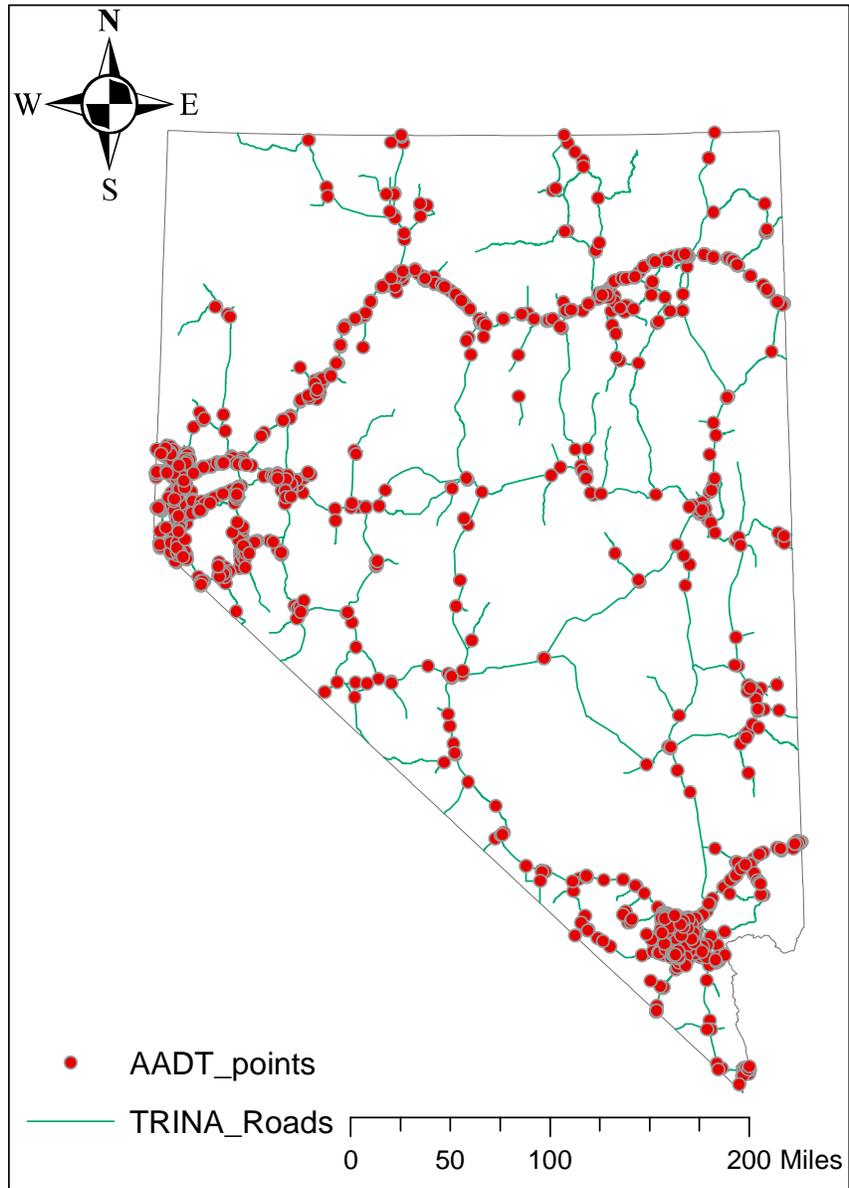


Figure 5-4. TRINA AADT data layer

Clark County Road Data

The Clark County road segment layer was developed by UNLV for the Clark County database of Safety Analyst and was provided by NDOT TSE. Figure 5-5 illustrates this roadway layer. Since the layer does not use the same LRS as the HPMS layer and some data errors found in this layer, the Clark County road layer was not aggregated into the statewide road database.

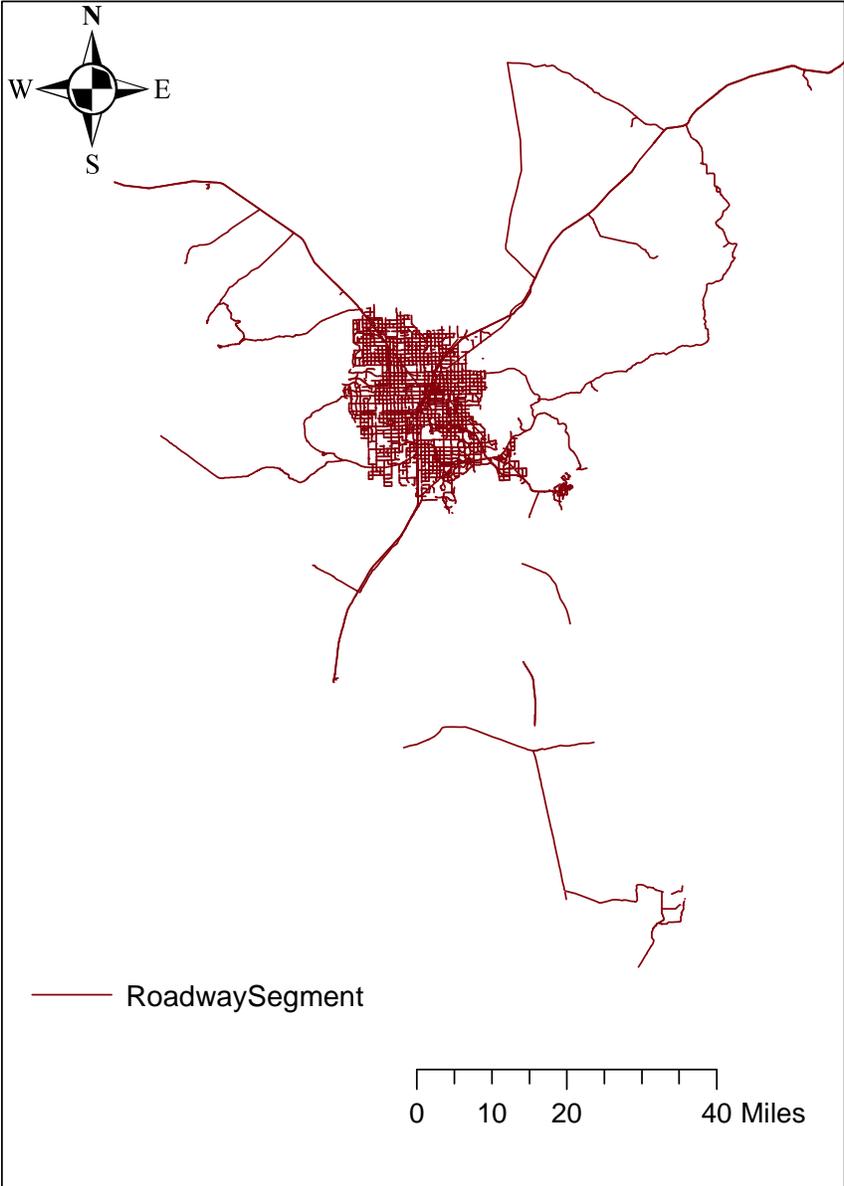


Figure 5-5. Clark County Safety Analyst road layer

Table 5-3 lists how the required data elements being filled by the received Nevada data and additional data collection.

Table 5-3 Mandatory Roadway Elements Required by Safety Analyst

Data element	Covered in HPMS layer (Yes/No)	Data collection/processing	Additional data
routeType	No	Functional class is used as the route type	
routeName	Yes		
county	Yes	The last two letters in RouteID	
startOffset	Yes		
endOffset	Yes		
segmentLength	Yes	Need to be recalculated after the overlay processing	
areaType	Yes	Partly covered	Urban/rural boundary layer
d1numThruLane	No	Manually coding in Google Earth	
d2numThruLane	No	Manually coding in Google Earth	
medianType	No	Manually coding in Google Earth	
calendarYear	Yes		
aadtVPD	Yes	Partly covered	AADT layer

The available AADT data are from 2004 to 2014. Default AADT values were used for some local streets in the NDOT AADT data. The default values, such as 50 vehicles-per-day and 100 vehicles-per-day, could cause errors of Safety Analyst. AADT values were calibrated by the adjacent years' volumes to improve the data reliability. In the data processing, it was found that the M (a property for the linear referencing system) information of some road segments in the HPMS GIS route layer is not correct, which caused problems when integrating different road characteristics. The road segments with M errors were excluded from the final dataset.

5.1.2 Intersection Dataset

GIS layers and spreadsheets of intersection data were obtained from NDOT divisions for the statewide intersection database. The received data are summarized in Table 5-4.

Table 5-4. Received Intersection Data from NDOT Divisions

	Format	NDOT Division
HPMS layers	Geometric layer	NDOT Traffic Safety Engineering
Intersections layer	Geometric layer	NDOT Traffic Safety Engineering
AADT estimation sheet	Spreadsheet	NDOT Traffic Information Systems

The intersections layer was received from NDOT TSE. The layer contains 6,431 intersections. The data of route name, milepost information, intersection type, and traffic control type were available in the layer. Figure 5-6 shows the statewide distribution of the intersections. The AADT estimation sheet is a table for NDOT to estimate the AADT values if the data are not counted on some roads.

The intersection dataset of Safety Analyst is built based on the intersection layer received from NDOT TSE. As the intersection layer uses a different LRS from the HPMS LRS, offsets between the intersection locations and HPMS road crossings exist. Intersections in the intersection layer were spatial-joined to the road crossings in the HPMS layer, so the LRS of the intersection layer was changed to the standard LRS in the Nevada HPMS. The minimum intersection elements required by Safety Analyst are listed in Table 5-5.

Table 5-5. Mandatory Intersection Elements Required by Safety Analyst

Data element	Covered in intersection/HPMS layer (Yes/No)	Data collection/processing	Additional data
routeType	No	Functional class is used to represent route type	
routeName	Yes		
county	Yes	Last two characters are the county abbreviation	
majorRoadOffset	Yes		
areaType	Yes	Partly covered	Urban/rural boundary layer
intersectionType	Yes		
calendarYear(MajorRoad)	Yes		
aadtVPD (MajorRoad)	Yes		
calendarYear(MinorRoad)	No		AADT estimation sheet
aadtVPD (MinorRoad)	No		AADT estimation sheet

The geographic information of many minor roads at intersections are missing in the HPMS routes layer. The imaginary minor roads are then created to make sure the data meets the requirement of Safety Analyst. The AADT of minor road is estimated based on the AADT estimation sheet.

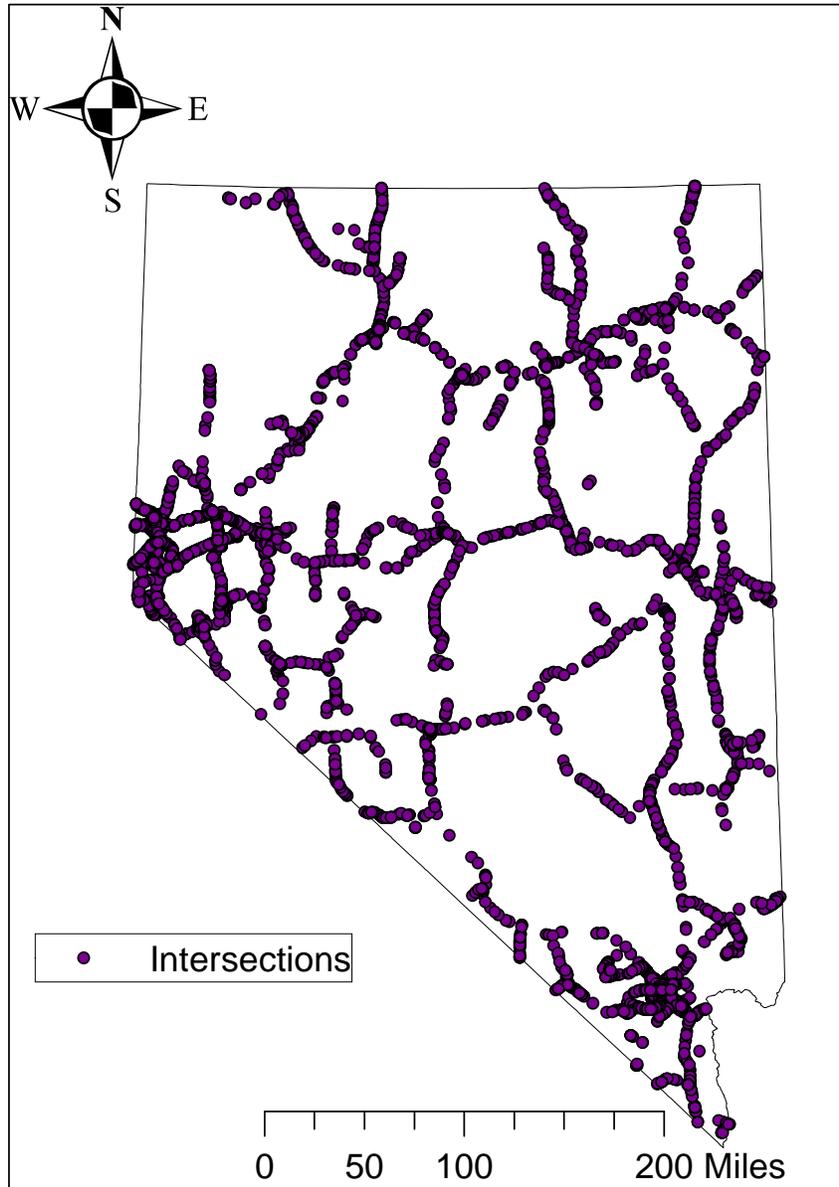


Figure 5-6. NDOT intersection data layer

5.1.3 Ramp Dataset

A ramp layer was received from NDOT TSE. The ramp layer covers 987 highway ramps. It contains the ramp name, milepost, ramp type, ramp configuration, and ramp length. Figure 5-7 presents the ramp layer. However, this ramp layer does not cover all the ramps in the HPMS layer, and the location offset exists between the ramps and the HPMS road network.

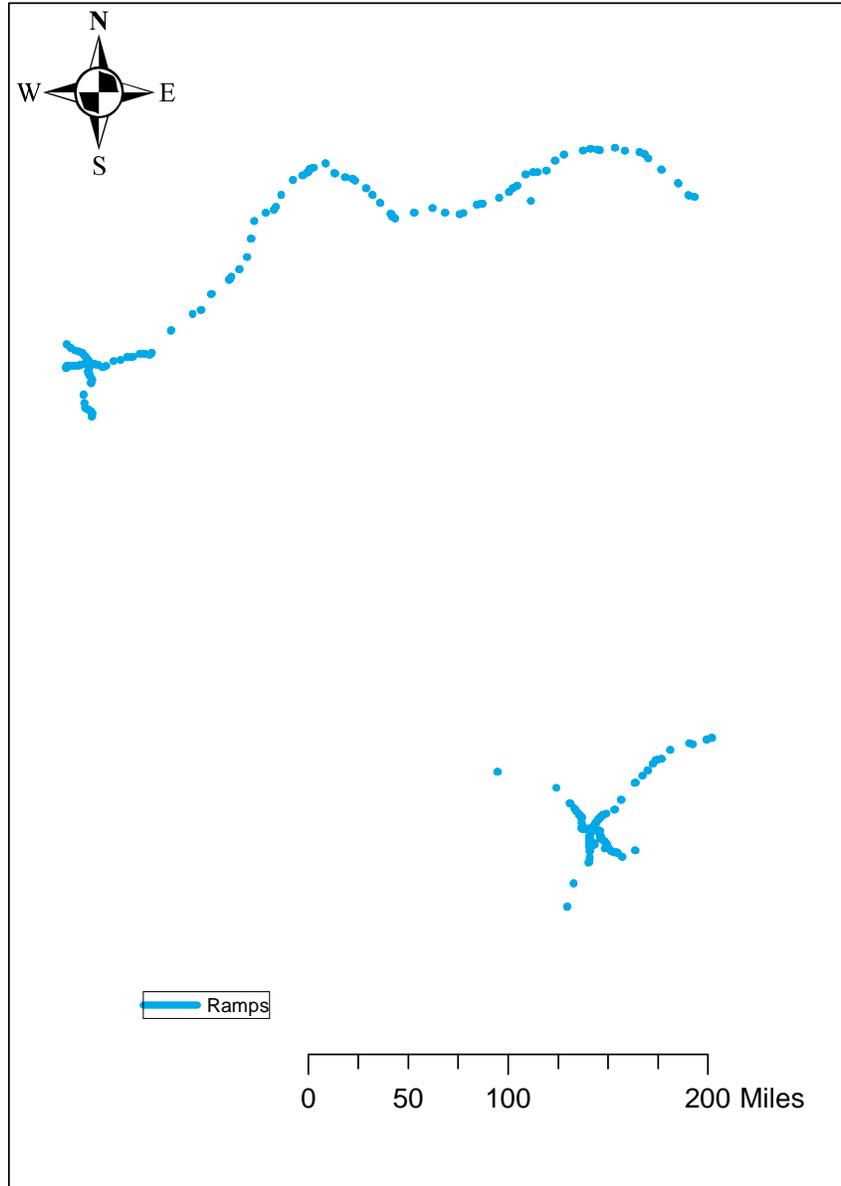


Figure 5-7. NDOT statewide ramp layer

UNR CATER extracted ramps from the road segment dataset for Safety Analyst (Section 5.1.1). A route name with letters of “RM” indicates that the road segment is a ramp, which was used to identify the ramps from the road segments. The minimum ramp data elements required by Safety Analyst are shown in Table 5-6, with the data preparation/processing methods.

Table 5-6. Mandatory Ramp Data Elements

Data element	Covered in road segment dataset (Yes/No)	Data collection and processing
routeType	Yes	
routeName	Yes	
county	Yes	
startOffset	Yes	
endOffset	Yes	
areaType	Yes	
rampType	No	Manually coded in Google Earth / imported from NDOT ramp layer
rampConfiguration	No	Manually coded in Google Earth / imported from NDOT ramp layer
numOfLanes	Yes	
rampLength	Yes	
calendarYear	Yes	
aadtVPD	Yes	

5.1.4 Crash Data

NDOT TSE provided the statewide crash data for preparing the Safety Analyst crash dataset. UNR CATER also obtained original crash records from the Nevada Citation and Accident Tracking System (NCATS) maintained by the Nevada Department of Public Safety. The data received for the statewide crash dataset are listed in Table 5-7. NDOT TSE receives crash data from the NCATS database, geo-locates the crashes onto a GIS layer, and excludes the crash elements not needed by NDOT TSE daily traffic analysis. The GIS layer of crashes provided by NDOT TSE includes 423,126 records with 10-year data from 2006 to 2015. Figure 5-8 presents the map of these crashes. The crash data from NDOT TSE do not contain some data elements that are required by Safety Analyst. The missing data elements are available in the original records in the NCATS database. Therefore, UNR CATER obtained the original crash records in the same period as the NDOT TSE crash data. The NCATS crash records were integrated into

the NDOT TSE crash GIS layer for all the crash elements required by Safety Analyst and the crash locations in GIS. The minimum crash data elements required by Safety Analyst are listed in Table 5-8, with the information of data preparation and processing.

Table 5-7. Crash Data for the Safety Analyst Crash Dataset

	Format	Agency/NDOT Division
Crash layer	Geometric layer	NDOT TSE
HPMS layers	Geometric layer	NDOT TSE
NCATS data form	spreadsheet	Nevada Department of Public Safety

Table 5-8. Mandatory Crash Data Required by Safety Analyst

Data element	Covered in the crash layer/NCATS form (Yes/No)	Data preparation/processing
routeType	No	Functional class in HPMS is used to represent route type
routeName	Yes	
accidentDate	Yes	
accidentSeverity1	Yes	Not the same division criteria, was adjusted
numberOfFatalities	Yes	
numberOfInjuries	Yes	
junctionRelationship	Yes	Partly covered, creating buffers around intersections to determine the values.
numVehicles	Yes	
collisionType	Yes	

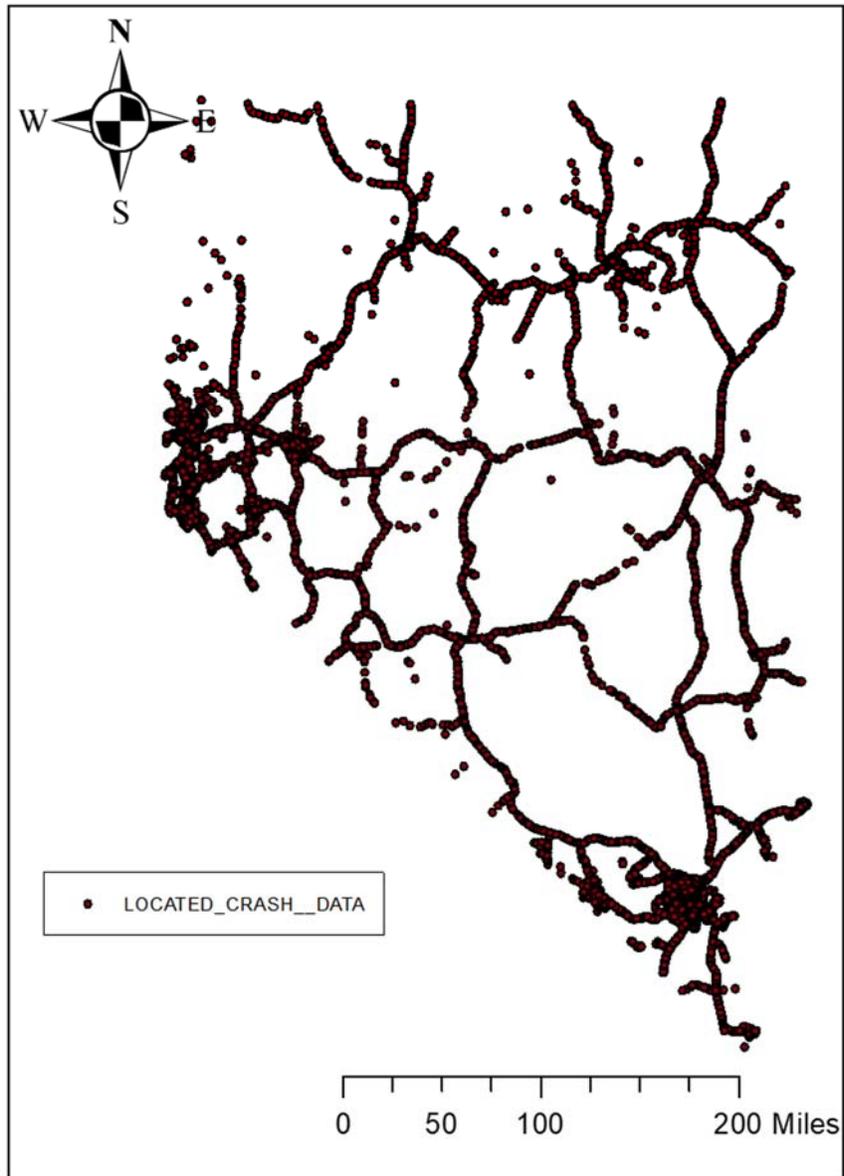


Figure 5-8. Map of statewide crashes 2006-2015

5.2 DATA PROCESSING

As NDOT data are mainly managed as GIS layers, ESRI ArcMap was used as the major tool for building the Safety Analyst database. This section documents data processing to convert the NDOT existing dataset to the Safety Analyst dataset. The final output of data processing includes CSV files of road, intersection, ramp, and crash data that are ready for being imported into Safety Analyst, and related GIS layers of these CSV tables.

5.2.1 Road Segment Data Processing

The major steps of processing road segment data include HPMS route layer revision, HPMS data overlay, collection of missing data, AADT assignment, and formatting data.

HPMS route layer revision

The 2014 Nevada HPMS route layer was used as the base road network of the Safety Analyst database. Some errors with the HPMS LRS were found on a few road segments in the HPMS layer. These errors cause issues in overlaying HPMS layers and crash assignment. The major errors include:

- Error 1. Missed linear referencing information (the value of vertex "M" property shows "NaN").
- Error 2. All points on a road segment have the same milepost information (the "M" values of road vertices are same)
- Error 3. Different points on a road segment have repeated milepost information (the begin and end vertices of a segment have the same "M" values)

The accuracy of the road segment database is determined by the accuracy of HPMS route layer. Some road segments with wrong linear referencing information were fixed.

HPMS data overlay

Each road element data were stored in a table of the NDOT HPMS dataset. The records in the tables were located using LRS. These tables were firstly overlaid using the "Overlay Route Events" function in ArcMap. This tool overlays two or more event feature layers against a target network and outputs a feature class or event table that represents the union of the input. The "Make Route Event Layer" was then used to create a new HPMS layer with road properties using the overlaid table and the HPMS route layer as input.

Collection of missing data

Some information such as county and median type was missing in the HPMS data. The data gaps were filled with data manually coded by CATER staff in Google Earth. The missed data and data extraction are summarized in Table 5-9.

Table 5-9. Road Segment Data Gaps and Data Extraction to Fill the Gaps

Element	Method	Note
Median Type	Extracted from GoogleEarth	
Through Lane	Extracted from GoogleEarth	
Facility Type	Extracted from GoogleEarth	
County	Extracted from Route ID in HPMS route layer	The two leading letters of Route ID are the abbreviation of county
Area Type	Overlaying the urban boundary layer with HPMS route layer	If the road crosses the boundary of urban and rural area, then the area type is determined by the longer part of road

AADT assignment

The HPMS_2014_AADT data covers part of the HPMS road segments, while Safety Analyst needs AADTs of all routes. The TRIAN AADT data does not have milepost information, and there are offset between counting stations and HPMS routes. These issues led to uncontrolled errors when TRINA data were integrated with the HPMS AADT. Therefore, the additional AADT data and related HPMS routes were obtained from the NDOT Roadway Systems Division. This table is segmented using the “Make Route Event layer” function in ArcMap to locate the data geospatially. Then the AADT table was joined into the overlaid HPMS.

Formatting data

The data format in HPMS layer is different from the data format required by Safety Analyst. The original Data format was then converted into the format required by the Safety Analyst in ArcMap. The final GIS layer containing the mandatory data of Safety Analyst with the required format.

5.2.2 Ramp Data Processing

The Ramp database is generated from the road segment dataset. The “altRouteNames” element of ramps in the road segment dataset is named beginning with “RM.” This feature was used to extract ramps from the road segment dataset. The detailed ArcMap operation steps are in the follows:

- 1) Load “Roadsegment_Database” into ArcMap.
- 2) Open the attribute table of Roadsegment_Database, click”Table Options-Select by Attribute.” The input is shown in Fig.1, Click “Apply” button.

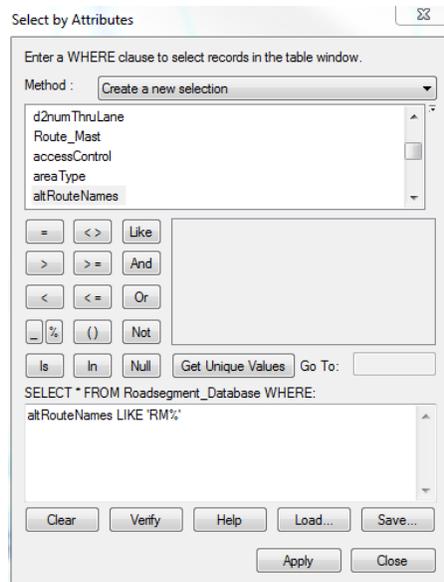


Figure 5-9. Example input for searching ramps in the road segment layer

- 3) Right Click “Roadsegment_Database” in Table of Contents, select “Data-Export Data...”, save the selected data into a new feature class. The exported feature class is the ramp database.

Some information such as ramp type and ramp configuration was missing. The data gaps were filled with data manually coded by CATER staff in Google Earth. The final GIS layer containing the mandatory data of Safety Analyst with the required format was named “Ramp_Clear” in the delivered dataset.

5.2.3 Intersection Data Processing

The intersection GIS layer from NDOT TSE includes the intersection properties required by Safety Analyst. However, the intersection layer was not generated based on the HPMS LSR, so offset exists between the intersections and the roads in the HPMS layer. When the HPMS is selected as the base road network and the standard LSR, the intersections need to be matched to the road crossings in the HPMS layer. The detailed steps are as follows:

1. Unsplitting HPMS road lines: use the Unsplit Line function of ArcMap to merge lines that have coincident endpoints and the same route mast IDs.
2. Create intersections from the HPMS roads: The output of Unsplit Line is input to the Intersect tool, as shown in the model snippet below. Intersect creates points where line segments touch.
 - When Intersect is run with one input feature class, the feature class is intersected with itself.
 - The Output type is POINT, meaning that the output features are points created where two streets intersect.
 - In this instance, the JoinAttributes is set to ONLY_FID, meaning that no attributes from the input features are carried to the output feature class.

3. Removing duplicates of generated intersections: The output of Intersect is a feature class with duplicate points at every intersection. Note that there are no points created for dead-ends, nor are there any pseudo-junctions. To remove the duplicates, keeping only one of the duplicate points, use the Delete Identical tool. For the Fields(s) parameter in Delete Identical, use the Shape field. This deletes all but one of the points that share the same x,y location.
4. Joining the intersections from the NDOT intersection layer to the intersections generated from the HPMS roads: The Spatial Join function was used. In the output feature, a matched record includes the intersection properties. Some of the generated intersections do not have matched intersection data, because the NDOT intersection dataset does not cover all HPMS road intersections. It also needs to be noted that some records in the NDOT intersection layer are not on HPMS roads, because HPMS roads do not cover all public roads in the states.
5. Remove the generated HPMS intersections without any matched intersection data.
6. The NDOT intersections not matching to the HPMS intersections were spatial-joined to the HPMS roads. As some public roads, especially minor roads, are not included in HPMS, the HPMS data may only include the major roads of an intersection. As the minor road is missed, no intersection is generated from HPMS, but an actual intersection exists at this location. In this step, the NDOT intersections with only major roads in HPMS were located along those major roads, so these intersections are included in the Safety Analyst database.
7. Merge the intersections extracted in Step 5 and 6.
8. The research team found that the Clark County intersection database generated by UNLV includes more intersections in Clark County than the NDOT data and HPMS intersections. Therefore, the extracted intersections were merged with the Clark County database to generate a state intersection database including 6847 intersections.
9. Determine the AADTs of intersection major roads and minor roads from the HPMS dataset. For the intersections with only major roads in HPMS, the AADT of minor roads were determined with the AADT estimation sheet provided by the NDOT Traffic Information.

The final GIS layer contains the mandatory data of Safety Analyst with the required format.

5.2.4 Crash Data Processing

The HPMS road network is used as the base geography reference of the Safety Analyst database, but the crash data provided by NDOT TSE uses different LRS. Crash data were processed with the ESRI GIS geoprocessing tools, so all the data were adjusted to the same LRS of the HPMS network. The crash data from NDOT TSE were extracted from NCATS and customized for the applications at NDOT TSE. The crash data from NDOT TSE is used as the base of the Safety Analyst crash dataset. However, Safety Analyst needs more crash elements than what the NDOT TSE crash data included. The research team retrieved all the additional data elements from the original crash records in the NCATS database by querying the unique crash ID. The NDOT crash data were geo-located on the NDOT safety route layer, which uses the different LRS from the HPMS network. It needs to be noted that NDOT TSE had planned to convert the current safety

data to the LRS of the HPMS network that would be the standard LRS of NDOT GIS data. Therefore, the locations of crashes were converted to the LRS of HPMS. The “Locate Feature Along Routes” function of ESRI GIS was used to match crash data with the HPMS layer.

The NDOT TSE crash data uses the coordinates to describe crash locations. To include the crash data into the Safety Analyst database. Safety Analyst requires crash location in the LRS format that describes crash locations with route ID and the distance from the beginning of the road. As the NDOT TSE crash data were geolocated based on the NDOT TSE safety route network, this processing step first matched the crash records to the NDOT TSE safety routes for valid roadway ID. Then the crashes were matched to the HPMS road network with the information of roadway ID and the location to determine the LRS milepost values. The major steps of processing crash data are listed as follows.

- 1) Identify roadway ID (ROUTE_MAST) of each crash
 - a. Load the Safety Route road layer and the NDOT crash layer into ArcMap.
 - b. Open the Spatial Join tool through the menu path of “Geoprocessing—ArcToolbox—Analysis Tools—Spatial Join”. This tool Joins attributes from one feature to another based on the spatial relationship. The target features and the joined attributes from the join features are written to the output feature class.
 - c. Select the NDOT crash layer as Target Features and the Safety Route layer as Join Features, choose the Output Feature Class (Output Path).
 - d. A new layer will be created and added into the ArcMap. The new layer includes all the crash data elements and the roadway ID (ROUTE_MAST) information for LRS.
- 2) Identify LRS milepost information of each crash

The ArcMap linear referencing tool of Locate Feature Along Routes (When you locate point features along routes, you are determining the route and measure information where your point data intersects your route data.). Open the Locate Feature along Routes tool through the following menu path: “Geoprocessing—ArcToolbox—Linear Referencing Tools--Locate Feature Along Routes” Use the crash data assigned with roadway ID as the Input Features and use the HPMS road network layer as the Input Route Features. Be sure to check “Keep only the closest route location (optional).”

The final GIS layer containing the mandatory data of Safety Analyst with the required format.

5.2.5 CSV File Import

After the data processing, the GIS database with different GIS layers were generated. The GIS layers were then converted to csv files that can be imported into the Safety Analyst software.

- 1) For the GIS layer, open the attribute table, use the “Table Option”—“Export”, create a name with the suffix “.csv” for the export file, select “Text File” for “Save as type”.
- 2) For the roadsegment layer, open the file “AltRoadSegment.csv” in the Template folder (I created a folder under the Task 5 Design and Develop the Comprehensive Database for

Safety Analyst). Copy the content in the first step into the corresponding columns in the Template file. Check the required format of the dataset and make any necessary adjustment to meet the requirement.

- 3) The other datasets can use the similar method to obtain the csv files from the GIS layers.

For CSV file import, each CSV file can contain data representing only one of the primary data elements described in the following sections. The first row (line) of the file is a single field that specifies the name of the data element (i.e., the type of data) contained in the file. The second row contains a comma-separated list of the names of the items for the data contained in the file. Those names must match the names provided in the item descriptions for each of the elements described in the following sections, or must match the name of an item defined using the Deployment Attribute editor of the Safety Analyst Administration Tool. Each subsequent row contains a comma-separated list of the values for the data items specified in the second row of the file. The order of the values must correspond to the order of the items in the second row, and a value (or empty value) must be specified for each item.

All names and values can be optionally delimited by double quotes ("). When using quotes, empty values are specified as ("). If not using quotes, empty values are specified by consecutive commas (i.e., the value is not specified, but its separating comma is included).

5.3 DATASET VALIDATION

Data validation is performed during the import processing, post-processing, and calibration phases of the Safety Analyst data management workflow. Because data quality is validated during the import and post-processing phases of the workflow, the validation during calibration focuses more on having sufficient numbers of sites and associated crashes for the generation of calibration factors and crash distributions.

5.3.1 Import Validation

Minimal validation is performed during data import. In general, data are validated to a sufficient level to be able to store them in the Safety Analyst database (e.g., proper identification and location information). Detailed validation is performed during the post-processing phase of data management, and is described in a subsequent section of this chapter.

Import Errors

A data error prevents a dataset from changing states, i.e., a data error encountered during import will prevent the dataset from transitioning from the *created* to the *imported* state. Once 50 errors are encountered during import, the import process is automatically terminated.

There are only a few types of data errors that are considered during import processing. These errors prevent the data records from being inserted into the dataset:

- 1) Missing Agency ID in a primary data element
- 2) Missing Leg Id in an intersection leg data element
- 3) Missing Leg Type in an intersection leg data element
- 4) Missing Crash Date in a crash data element

- 5) Missing or invalid location specification in a primary data element

Import data errors must be addressed by supplying the missing data items or by removing the offending primary data elements.

Import Warnings

A data warning reflects a problem or potential problem in a data element. A dataset with data warnings can be processed, but problems identified by the warnings *may* prevent some data elements from being valid. Invalid data elements will not be able to be analyzed.

The possible warnings that may be encountered during import processing are listed in the following:

- 1) Invalid (unrecognized) data item for a data element (data item is ignored)
- 2) Agency ID is not unique (more than one data element with the same ID). Subsequent elements with the same ID will not be inserted into the dataset.
- 3) Invalid traffic year specification: year prior to 1970 (traffic specification is ignored)
- 4) Invalid traffic AADT specification: AADT value less than 1 (traffic specification is ignored)
- 5) Leg traffic specified for a non-existent leg (traffic specification is ignored)

5.3.2 Post-Processing Validation

Comprehensive data validation is performed during the post-processing phase of the data management process. The data validation is described in the following sections.

Post-Processing Errors

A data error during post processing prevents the dataset from transitioning from the *imported* to the post processed state. Once 50 errors are encountered during post processing, the process is automatically terminated.

The following errors may be encountered during post processing:

- 1) Non-unique Agency ID (this should have been detected during import)
- 2) Two intersections with the same location
- 3) Two or more roadway segments that overlap in location

Post processing data errors must be addressed by supplying missing data items and correcting invalid data items, or by removing the offending primary data elements.

Post-Processing Warnings

A data warning reflects a problem or potential problem in a data element. A dataset with data warnings can continue to be processed, but problems identified by the warnings *may* prevent some data elements from being valid. Invalid data elements will not be able to be analyzed.

Much of the validation that occurs during post processing is unique to the type of data element undergoing validation. All warning messages include both the agency-specified

site/crash identifier and the internally-assigned Safety Analyst primary key identifier. If a problem is severe enough to render the site invalid for analysis, the warning message will indicate **invalid**.

5.4 DATA ACCURACY OF THE STATEWIDE SAFETY ANALYST DATABASE

UNR CATER randomly selected data records in the generated Safety Analyst database, and manually checked whether the properties of roads, intersections, ramps, traffic and crashes are correct by using the image data in Google Earth and Google Streetview. Table 5-10 shows the results of data accuracy validation.

Table 5-10 Data Validation of Statewide Database

Roadsegment Dataset	Checked element	Checked number	Correct number	Accuracy Rate
	OperationWay	51	51	100%
	d1numThruLane	51	49	96.8%
	d2numThruLane	51	51	100%
	medianType	51	47	92.2%
Intersection Dataset	majorRoadName	50	48	96%
	intersectionType1	50	50	100%
	trafficControl1	50	49	98%
Ramp Dataset	rampType	50	50	100%
	numOfLanes	50	49	98%

The results show that all the checked datasets have the accuracy rate higher than 90%.

6 SUMMARY

The completed Safety Analyst database contains four datasets: road segment dataset, intersection dataset, ramp dataset and crash dataset. Each dataset includes its CSV files and corresponding GIS layer. The CSV files in each dataset are shown in Table 6-1.

Table 6-1. CSV Files in the Nevada Safety Analyst Database

Road Segment Dataset	Intersection Dataset	Ramp Dataset	Crash Dataset
AltRoadwaySegement.csv	AltIntersection.csv	AltRamp.csv	AltAccident.csv
AltSegmentTraffic.csv	AltMajorRoadTraffic.csv	AltRampTraffic.csv	
	AltMinRoadTraffic.csv		

The road segment dataset contains 5,152 road segments, which has a total length of 10849.37 miles. The dataset includes 12-years AADT records from 2004 to 2015. The intersection dataset contains 6,847 intersections, including 749 intersections from the Clark County Safety Analyst database. The intersection type distribution is shown in Figure 6-1.

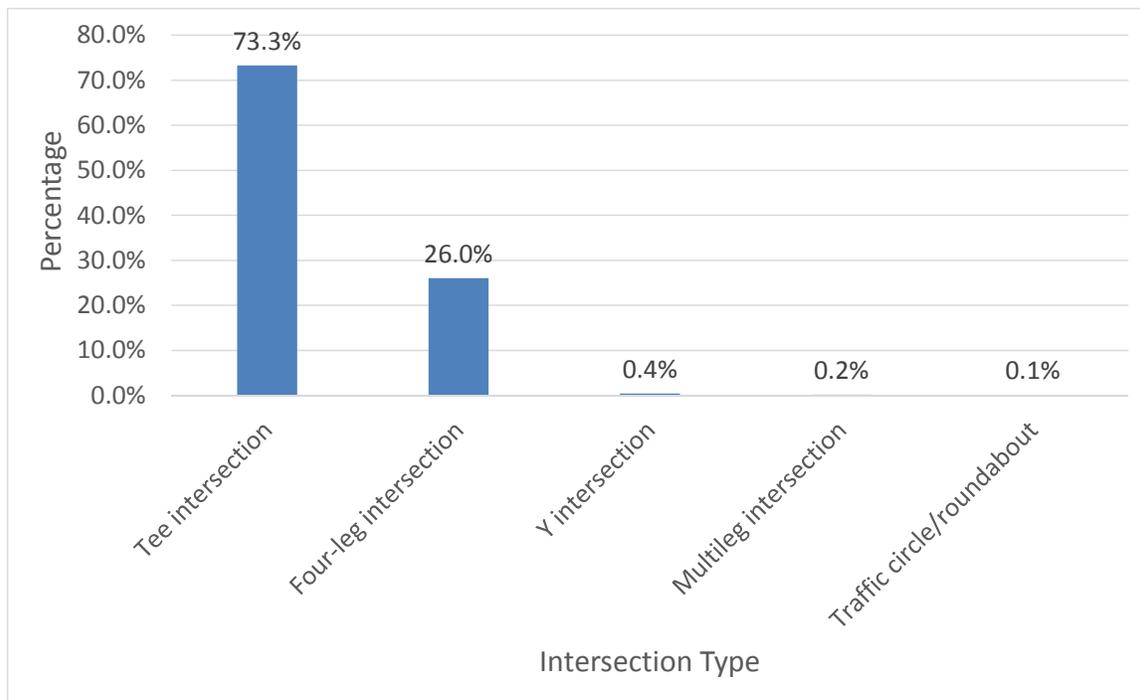


Figure 6-1. Intersection type distribution chart of the Nevada Safety Analyst database

The division of traffic control types is summarized in Table 6-2. The dataset includes 11-years AADT data of the main roads, and default AADT values of the minor roads.

Table 6-2. Numbers of Intersections in Different Control Types

Traffic Control Type	Total Number	Percentage
Stop sign control	4264	62.3%
No control	1431	20.9%
Signalized control	1112	16.2%
Yield sign	28	0.4%
Roundabout	6	0.1%
flasher	5	0.1%
Other	1	Less than 0.1%

There are 1,178 ramps in the ramp dataset, with a total length of 353.56 miles. The ramp dataset includes 571 off-ramps, 561 on-ramps, and 46 freeway-to-freeway ramps. Among the 1,178 ramps, 786 of them are diamond ramps, and 218 of them are partial clover leaf loop ramps. The AADT data are also from 2004 to 2015. The crash dataset contains 423,126 crashes in 10 years from 2006 to 2015. 249,198 crashes were identified as road-segment-related, 32,419 crashes have been designated as crashes on ramps and 141,509 crashes were intersection-related. Figure 6-2 shows the distribution of crashes by years.

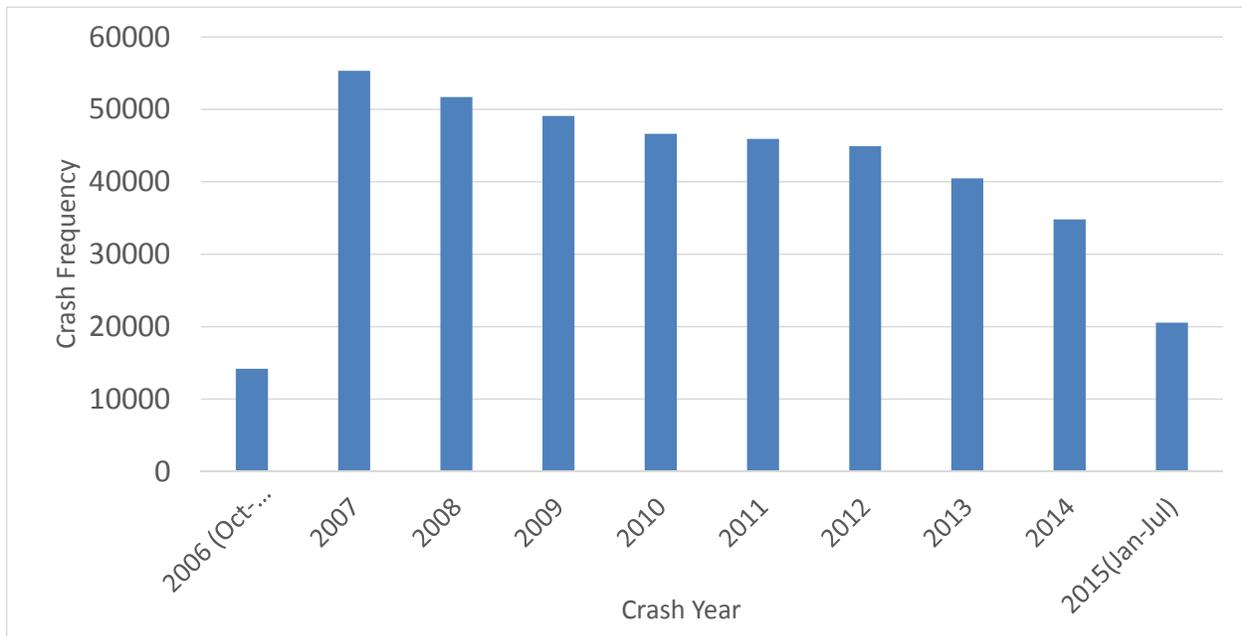


Figure 6-2. Crash distribution in different years

The crash severity distribution is shown in Figure 6-3. About half of those crashes are Property-Damage-Only and 0.6% of crashes are fatal crashes.

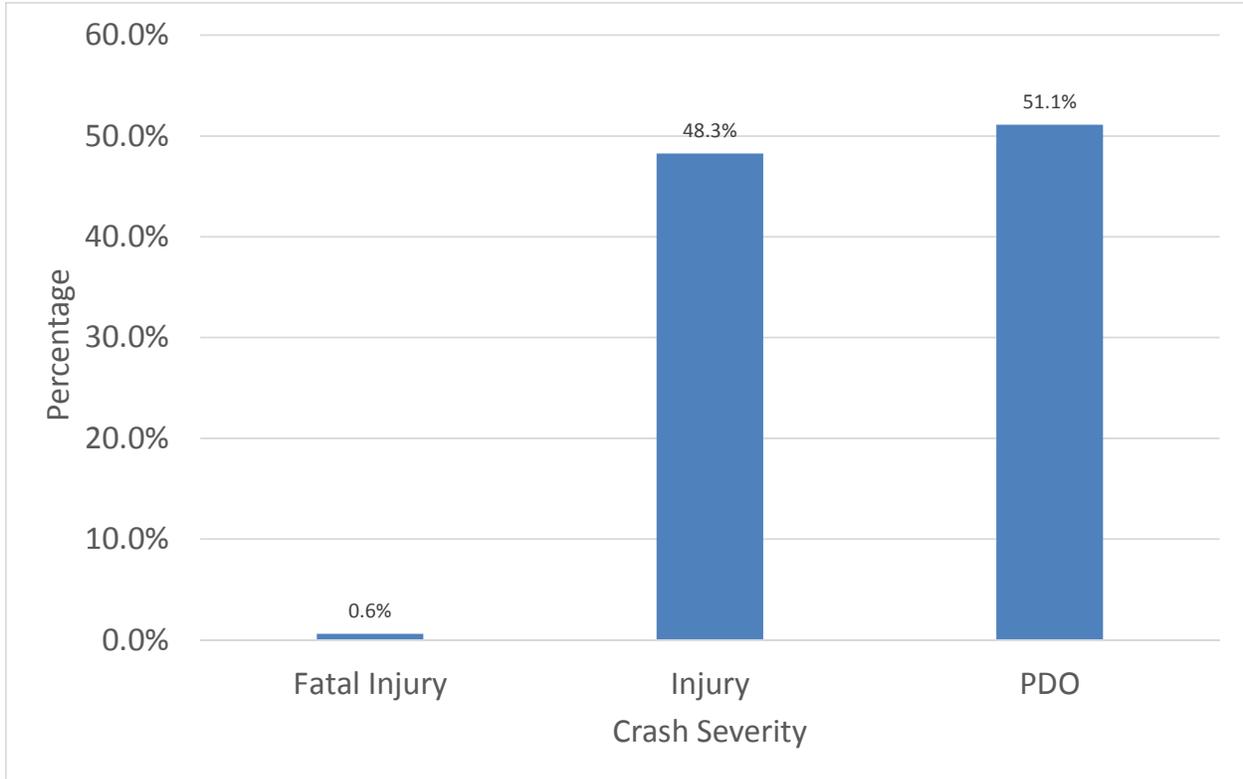


Figure 6-3. Distribution of crash severity

The crash type distribution is illustrated in Figure 6-4. Rear-end and angle crashes are the two major crash type.

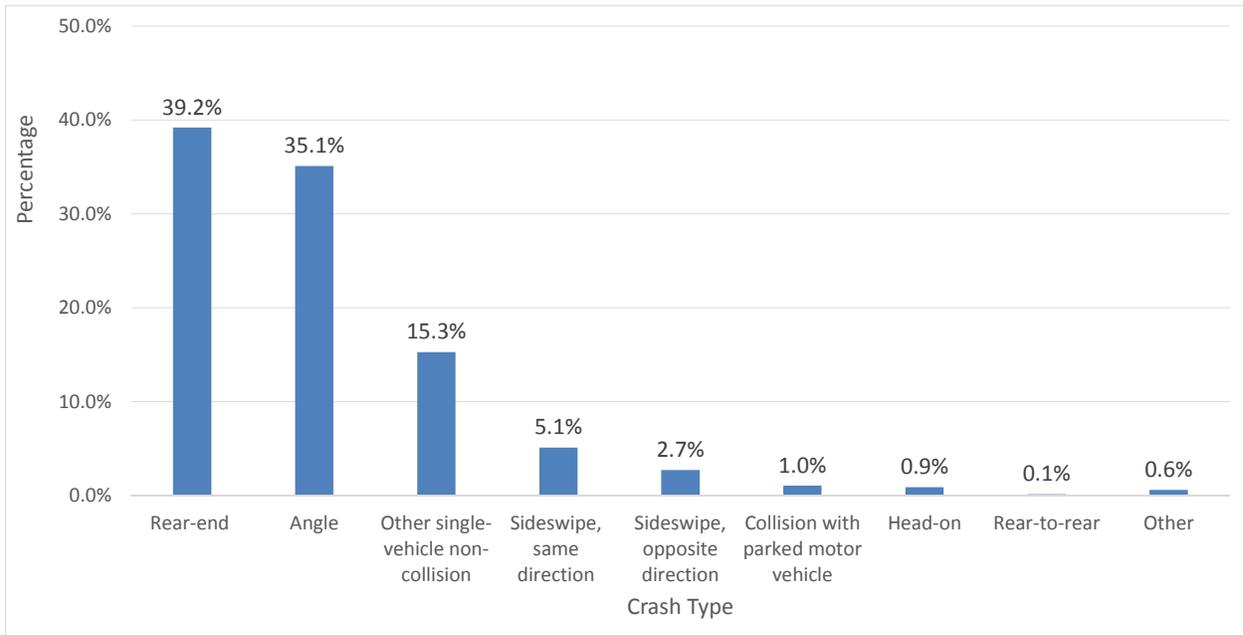


Figure 6-4. Distribution of crash types

Among those crashes, 7,425 crashes are work zone related crashes, 5 crashes involved school buses, and 27,399 crashes are alcohol related.

The GIS database includes 7 GIS layers:

- 1) Roadsegment_Clear
- 2) Intersection
- 3) Ramp_Clear
- 4) Crash_Clear
- 5) Roadsegment_Ramp_Clear
- 6) Roadsegment_Ramp_HPMS
- 7) Crash_Records

The top 4 layers are the corresponding GIS layers of SafetyAnalyst Database. The layers can be linked with the SafetyAnalyst Database through the unique attribute “agencyID”. For example, the “Roadsegment_Clear” layer has a column named “agencyID”, the “AltRoadSegment.csv” also has a column named “agencyID”, which can be used to link the roads in the “Roadsegment_Clear” layer.

The “Roadsegment_Ramp_Clear” is one layer that combined the road segments and ramps. Since there are some issues in the milepost information in the HPMS layer, not all roads in the HPMS layer are covered in the “Roadsegment_Clear” and “Ramp_Clear” layers. The “Roadsegment_Ramp_HPMS” layers contains all the HPMS roads. The column named “Located” represents whether the road segment is covered in the road segment/ramp layer. “1” means road segment is covered in road segment/ramp layer and “0” means road segment is not covered in road segment/ramp layer.

The “Crash_Records” contains all crash records. The column named “Located” represents whether this crash record is correctly assigned into the roads. “1” means the crash record is assigned to the road network and “0” means the crash record is not assigned to the road network.

This research project generated the Nevada statewide Safety Analyst database that integrates the state data of roads, intersections, ramps, traffic, and crashes. The database was formatted by following Safety Analyst requirements so that it can be directly loaded into the Safety Analyst software. UNR CATER has used the database to perform different Network Screening to identify the road segments and intersections with high crash frequency or high crash rates. Multiple other network screening can be conducted with this database, which will improve existing traffic safety management at NDOT TSE. The GIS database can be a supplemental to the Safety Analyst database, to extend the data applications to more scenarios. With the same information as the Safety Analyst database, the GIS layers allow different geo-analysis in GIS software and can be used for various data visualization.

It needs to be noted that the completed database is based on the HPMS road network and with additional local street/intersection information. It has not covered all public roads in Nevada. When more data are available in the future, this database can be extended with the same

Development of a Nevada Statewide Database for Safety Analyst Software

documented data processing procedure. New crash data and AADT values will be added to the database to update the crash and traffic information.

REFERENCES

1. SafetyAnalyst. McLean, VA: U.S. Dept. of Transportation, Federal Highway Administration; 2006.
2. Interactive Highway Safety Design Model: IHSDM. Washington, D.C.: U.S. Dept. of Transportation, Federal Highway Administration; 2000.
3. Highway safety manual. Washington, D.C.: American Association of State Highway and Transportation Officials; 2010.
4. Federal Highway Administration. *Safety Analyst*. <http://www.safetyanalyst.org/>.
5. Harwood, D. W., D. J. Torbic, K. R. Richard, and M. M. Meyer. *SafetyAnalyst: Software Tools for Safety Management of Specific Highway Sites*. NTIS Accession No. PB2010-111996. FHWA Publication No. FHWA-HRT-10-063. U.S. Department of Transportation. Federal Highway Administration, Washington, DC, July 2010.
6. *Safety Analyst User Manual*. ITT Corporation, Colorado Springs, CO. 2011.
7. Federal Highway Administration. *Highway Safety Manual Case Study 2: Implementing a New Roadway Safety Management Process with SafetyAnalyst in Ohio*. http://safety.fhwa.dot.gov/hsm/casestudies/oh_cstd.pdf.
8. Khan, G. S. Parker, X. Qin. *SafetyAnalyst Evaluation and System Requirements Assessment*. Wisconsin Department of Transportation. March 5, 2009.
9. Lu, J., Gan, A., Kirolos, H., Alluri, P., and Liu, K. *Comparing Locally-Calibrated and SafetyAnalyst-Default Safety Performance Functions for Florida's Urban Freeways*. Proc., Transportation Research Board 91st Annual Meeting, Washington D.C., 12-4730. 2011.
10. Gan, A., K. Haleem, P. Alluri, D. Saha. *Standardization of Crash Analysis in Florida*. Contract No. BDK80 977-10. March, 2012.
11. Srinivasan, R., C. Lyon, B. Persaud, C. Martell, and J. Baek. *Methods for Identifying High Collision Concentration Locations (HCCL) for Potential Safety Improvements – Phase II: Evaluation of Alternative Methods for Identifying HCCL*. CFS Number 2078A DRI. California Department of Transportation, Sacramento, California. January 21, 2011.
12. Dixon, K. K., Meng, L., I.V. Schalkwyk and C. Monsere. *Highway Safety Manual (1st Edition) Part B Oregon Data Assessment*. Oregon Department of Transportation. Traffic Roadway Section. 2011.
13. Ogle, J., and Alluri, P. *Getting Started with SafetyAnalyst: Georgia's Experience*. Proc., Transportation Research Board 90th Annual Meeting, Washington, D.C., 11-3691. 2010.

14. Garber, N. J. and G. Rivera. *Safety Performance Functions for Intersections on Highways Maintained by the Virginia Department of Transportation*. Virginia Department of Transportation. 2010.
15. Kelly, P. C. *Traffic Safety Analysis of Utah County Highways using Safety Analyst*. Master Thesis. Brigham Young University. 2013.
16. Khan, G., K. R. Santiago-Chaparro, M. Chiturri, and D. A. Noyce. *Development of Data Collection and Integration Framework for Road Inventory Data*. In Transportation Research Record: Journal of the Transportation Research Board, No. 2160, Transportation Research Board of the National Academics, Washington, D.C., 2010, pp. 29-39.
17. Alluri, P. and J. Ogle. *Road Safety Analysis: States' Current Practices and Their Future Direction*. In 91st Annual Meeting Transportation Research Board, CD-ROM Transportation Research Board of the National Academies, Washington, D.C., 2011.
18. *Data Integration Primer*. Federal Highway Administration. U.S. Department of Transportation. 2001.
19. Pendyala, R. M., and Oak Ridge National Laboratory. *Development of GIS-Based Conflation Tools for Data Integration and Matching: Final Report*. Florida Department of Transportation. Presented at 87th Annual Meeting of the Transportation Research Board, Washington, D.C., 2008.
20. O'Paci, P., R. Dubois, N. Armentrout, and S. Bower. *Maine's Approach to Data Warehousing for State Departments of Transportation*. In Transportation Research Record: Journal of the Transportation Research Board, No. 1719, Transportation Research Board of the National Academics, Washington, D.C., 2000, pp. 227-232.
21. Hall, J. P., R. Robinson, and M. A. Paulis. *Enterprisewide Spatial Data Integration of Legacy Systems for Asset Management – The Case of the Illinois Department of Transportation*. In Transportation Research Record: Journal of the Transportation Research Board, No. 1917, Transportation Research Board of the National Academics, Washington, D.C., 2005, pp. 11-17.
22. Izadpanah, P., S. Nichol and A. Hadayeghi. *Configuration of SafetyAnalyst Software for Efficient and Effective Safety Management*. Presented at the Session of the Conference of Transportation Association of Canada. 2012.



Nevada Department of Transportation
Rudy Malfabon, P.E. Director
Ken Chambers, Research Division Chief
(775) 888-7220
kchambers@dot.nv.gov
1263 South Stewart Street
Carson City, Nevada 89712