Florida Traffic Safety Portal Statewide Crash Shapefile Creation:

- List of Acronyms
- Introduction
- CAR Database Extractions
- GIS Datasets
- Geo-Locating Crashes
- Linear Referencing System
  - On System
  - Off System
- Parsing

List of Acronyms

**CAR** – Crash Analysis and Reporting System – A database and query system used to perform crash data analysis.

**DHSMV** – Florida Department of Highway Safety and Motor Vehicles – The state agency responsible for collecting crash reports from law enforcement agencies.

**FDOT** – Florida Department of Transportation – The state agency responsible for improving the safety of state and local roads.

**GIS** – Geographic Information System – A system used to display and analyze location based information.

**LRS** – Linear Referencing System – A measurement system used to locate events along a linear feature like a road or a stream.

**RCI** – Roadway Characteristics Inventory – An inventory database and query system used by the FDOT to document the dimensions, types, conditions, features and characteristics of a roadway, such as guardrail type or roadway width.

**SHS** – State Highway System – The system of roads owned or maintained by the FDOT.

**SHSP** – Strategic Highway Safety Plan – The plan the FDOT uses to direct safety efforts and funding to provide safer roads for the motoring public.

**SSO** – State Safety Office – The office within the FDOT responsible for providing crash data and directing safety efforts within the state.
Introduction

Motor vehicle crashes reported by a law enforcement officer on a Florida Traffic Crash Report long form are collected and maintained by the DHSMV. To obtain more accurate information about the crashes which occur on Florida roads and streets, the FDOT’s SSO Crash Report Analysts review each crash report and assign an appropriate location coordinates based on the reporting officer’s description of the crash. This white paper explains the process of creating the GIS datasets representing locations of individual crashes. More information about the collection and processing of the crash report data from the law enforcement officer to the SSO can be found in the white paper titled “Crash Data Management.”

CAR Database Extractions

To assist in the crash report review process, the crash information is uploaded to the FDOT’s CAR database. Once a crash has been reviewed and accepted, the information is stored in the CAR database and can be downloaded. A query is used to extract crashes from the CAR System into tabular format for a calendar year(s) or specific portion of a calendar year.

GIS Datasets

Crash data are collected and reported with three distinct focuses: Crash, Vehicle and Occupant. The Crash dataset documents one entry per crash and focuses on information about the location of the crash event and the conditions at the time of crash like weather or lighting. The Vehicle dataset documents one entry per vehicle and focuses on information about the vehicle(s) involved in a crash like vehicle type and sequence of events that the vehicle went through during the crash event. The Occupant dataset documents one entry per person involved in a crash and provides information about each, like their age, injury level, their actions that contributed to the crash event and whether they were an occupant of a vehicle or were a non-motorist. GIS data are provided at these levels with some summaries provided at the crash data level.

Crash level GIS file names are denoted by the system type and four digit year (e.g., “On2012” and “Off2012”). Crash level data and Vehicle level data can be joined together by using the field “KEYFIELD1”, which is the four digit year and nine digit crash report number as assigned by the FDOT SSO based on the DHSMV crash report number. The Crash level data include information about when and where a crash occurred, the roadway characteristics of On system roads (roadways actively maintained by the FDOT), and a variety of “flags”, or yes/no fields, which correspond to the 2012 SHSP Emphasis Areas or to other areas of traditional concern.

The Vehicle dataset repeats the basic information about where the crash occurred and gives more detailed information about the vehicle type, the causes of the crash and the sequence of events which the vehicle went through during the crash. Vehicle level data file names are denoted by the letter “V”, the system type and four digit
year (e.g., “Von2012” or “Voff2012”). The Occupant level data can be joined to both the Crash and Vehicle level data by using “KEYFIELD1” and “KEYFIELD2” respectively. KEYFIELD1 is defined above. “KEYFIELD2” is the four digit year concatenated with nine digit FDOT crash report number and the two-digit vehicle sequence number from the crash report. Vehicle level data contains a record for each vehicle involved in a crash event. The Vehicle level data will not contain any entries that correspond to non-motorists who were not occupants of a vehicle. For example, if there were two vehicles involved in a crash event, there would be one record in the crash level and two records in the vehicle level datasets.

The Occupant dataset also contains a field, “KEYFIELD3”, which can be used to distinguish between records for each person involved in a crash and can be used as a unique identifier. Using the previous example of a crash with two vehicles, a driver and one occupant in each vehicle, there would one record at the Crash level for the event, two records at the Vehicle level with information about each vehicle, and four records at the Occupant level, one for the driver of each vehicle and one for each occupant. The records are distinguished by creating a unique identifying number in “KEYFIELD3” by joining the crash year, FDOT crash record number, vehicle sequence number and person sequence, creating a 17-character string. Occupant level data files names are denoted by the letter “O”, the system type and four digit year (e.g., “Oon2012” or “Ooff2012”).

**Geo-Locating Crashes**

After the crash information has been downloaded in tabular format, each record must be located to create the GIS datasets. This process, known as geo-locating, is different for crashes located on roads under the jurisdiction of state agencies like the FDOT and state-chartered expressway authorities. These roads which are also known as the SHS or “On system” roads, store crash location information in different form than crashes located on publically accessible roads, (i.e., local roads and streets that are not On System), which are known as “Off system roads”. Off system roads are public roadways that are not maintained by the FDOT.

**Linear Referencing System**

The FDOT uses an LRS to locate crash events, features, and roads within the SHS. An LRS is a method of locating events along a line (a roadway) using the distance from a specific beginning point. For SHS event data, using the FDOT LRS ensures that any event occurring along the roadway is always located on the roadway itself and not at a random point in space.

The SSO maintains a separate LRS for Off system streets for its own reporting and analysis purposes. The Off system LRS is based on a commercial GIS product from HERE, Inc. (previously called Navteq Streets). Each GIS segment in the data which does not correspond to the FDOT LRS is coded with an LRS identifier entitled
“ROADWAY”, and a calculated beginning and ending milepoint, so an LRS GIS route system can be built. Crash locations are then located along the Off system LRS, and roadway identifier and mile-point are assigned to them.

**On System**

On system crashes are located using the “ROADWAYID” and “LOCMILEPT” fields in each level of data. These crashes are located on the FDOT’s RCI basemap dataset using LRS coordinates. The RCI LRS uses a measured distance from the origin of the segment. The RCI basemap dataset includes lines representing each roadway within the inventory. It is a representation of the roadway system, which is comprised of individual features (intersections, bridges, auxiliary lanes, medians, et al.) positioned along the roadway segment. When a Crash Report Analyst reviews a crash report, they determine the location of the crash in terms of a specific roadway at a distance from a documented feature along the roadway. Using the roadway ID and milepoint data, the crashes are placed at the location determined by the Crash Report Analyst.

**Off System**

Off System crashes are located using a different process and system than that which is used for On system crashes. Instead of using a specific roadway ID and milepoint, Off system crashes are stored using their geographic coordinates. To ensure crashes are located along an existing roadway, the Crash Report Analysts locate a crash using a commercially available roadway dataset. This dataset is comprised of nodes and segments representing a roadway. Unlike the RCI basemap however, the commercially available dataset does not document the same range roadway features as the FDOT RCI. After crashes are located using the commercial dataset, the location coordinates (latitude, longitude, map dataset line segment, closest dataset nodes and offset distances) are extracted from a tabular database into a GIS dataset. The tabular crash data are then joined to the GIS dataset of crash locations by matching a common field comprised of the crash year and crash report number.

**Final Shapefile Generation**

The data from each process (On system and Off system) are exported into GIS datasets representing the Crash, Vehicle and Occupant level data. Using the coordinates extracted, the crashes are projected against a geographic coordinate system to finish the processing of the shapefile. Finally, the shapefile is transformed into a State Plane projection.

**Parsing**

After creating GIS datasets representing On and Off system crashes for the entire state of Florida, new Crash, Vehicle, and Occupant datasets are created from the statewide files for specific geographic areas and types of crashes. These datasets represent specific subsets of crashes corresponding to the 2012 SHSP Emphasis Areas, as well as FDOT Districts and Counties. The 2012 SHSP places emphasis on
identifying the location of crashes involving aggressive driving, at-risk drivers (aging road users and teen drivers), distracted driving, impaired driving, intersection crashes, lane departure crashes, and crashes involving a vulnerable road user (bicyclist, pedestrian or a motorcyclist) in an effort to identify high risk areas. The crash files supporting the 2012 SHSP Emphasis Areas can be sorted by State, FDOT District and County levels for use in mitigating areas with elevated crash rates.