Report 15-UT-037

NSTSCE

National Surface Transportation Safety Center for Excellence

Roadway Epochs

Documentation and User Manual

Robert McCall • Shane McLaughlin

Submitted: August 24, 2015

Technology
Impairment
Mobility
Infrastructure

Housed at the Virginia Tech Transportation Institute
3500 Transportation Research Plaza • Blacksburg, Virginia 24061

ACKNOWLEDGMENTS

The authors of this report would like to acknowledge the support of the stakeholders of the National Surface Transportation Safety Center for Excellence (NSTSCE): Tom Dingus from the Virginia Tech Transportation Institute, John Capp from General Motors Corporation, Lincoln Cobb from the Federal Highway Administration, Chris Hayes from Travelers Insurance, Martin Walker from the Federal Motor Carrier Safety Administration, and Cathy McGhee from the Virginia Department of Transportation and the Virginia Center for Transportation Innovation and Research.

The NSTSCE stakeholders have jointly funded this research for the purpose of developing and disseminating advanced transportation safety techniques and innovations.

ABSTRACT

This report describes the contents of the Roadway Epochs Database, which is a collection of tables providing indices for users to access vehicle data from the 100-Car and Second Strategic Highway Research Program (SHRP 2) Naturalistic Driving Studies based on the characteristics of the roadways traveled by study vehicles. This document is designed to give an overview of the organization of the tables as well as a brief introduction to the available vehicle and infrastructure data. The document opens with a section outlining intended uses and general information regarding the Roadway Epochs Database. The database contains 35 tables containing data from five state Departments of Transportation (DOTs), Nokia/Navteq, the SHRP 2 Roadway Information Database (RID), and the Federal Highway Administration (FHWA) Office of Highway Performance Monitoring. Epochs of naturalistic data are tied to information such as the number of lanes, average annual daily traffic counts for heavy trucks and light vehicles, International Roughness Index scores, vehicle miles traveled, functional class, and identified features such as transition zones, frontage roads, bridges, and tunnels. The document closes with a few sample queries and a brief introduction to Structured Query Language (SQL).

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND SYMBOLS

AADT	annual average daily traffic
AAWDT	average annual weekday traffic
DAS	data acquisition system
DOT	Department of Transportation
FHWA	Federal Highway Administration
HOV	high-occupancy vehicle
HPMS	Highway Performance and Monitoring System
IRB	Institutional Review Board
IRI	International Roughness Index
NDS	naturalistic driving study
NOAA	National Oceanic and Atmospheric Administration
RID	Roadway Information Database
SHRP 2	Second Strategic Highway Research Program
SQL	Structured Query Language
VMT	vehicle miles traveled
VTTI	Virginia Tech Transportation Institute

CHAPTER 1. WHAT IS THE ROADWAY EPOCHS DATABASE?

A systems approach to understanding surface transportation safety integrates considerations of the driver, the vehicle, and the roadway. The naturalistic data sets housed by the Virginia Tech Transportation Institute (VTTI) are rich resources for understanding the interactions of these three subsystems. The use of naturalistic data began primarily in the domains of driver and vehicle research. More recently, those with a background in roadway infrastructure have been turning to naturalistic data sets for answers to research questions. In light of this, there is an opportunity to increase the accessibility of naturalistic data to researchers in roadway infrastructure.

The Roadway Epochs Database expands access to the available naturalistic data sets by introducing roadway- and infrastructure-centric indexing. In order to do this, several road-related data sets were combined with naturalistic driving data from two studies. The resulting Roadway Epochs Database provides a series of variables that can be used to locate sensor and camera data in the driving databases for each of the included collections.

The Roadway Epochs Database contains pointers to driving epochs that have already been associated with kinematic data in naturalistic data sets. In other words, this database provides a shortcut for researchers to locate driving data that correspond to roadway features of interest. Using these pointers, researchers can rapidly locate driving data, but the driving data from the naturalistic studies are not duplicated in the Roadway Epochs Database. This approach was taken to not overburden researchers with data, to minimize redundancies in data storage, and to simplify database management by not requiring version control of multiple copies. A conceptual illustration of the database is provided in Figure 1.

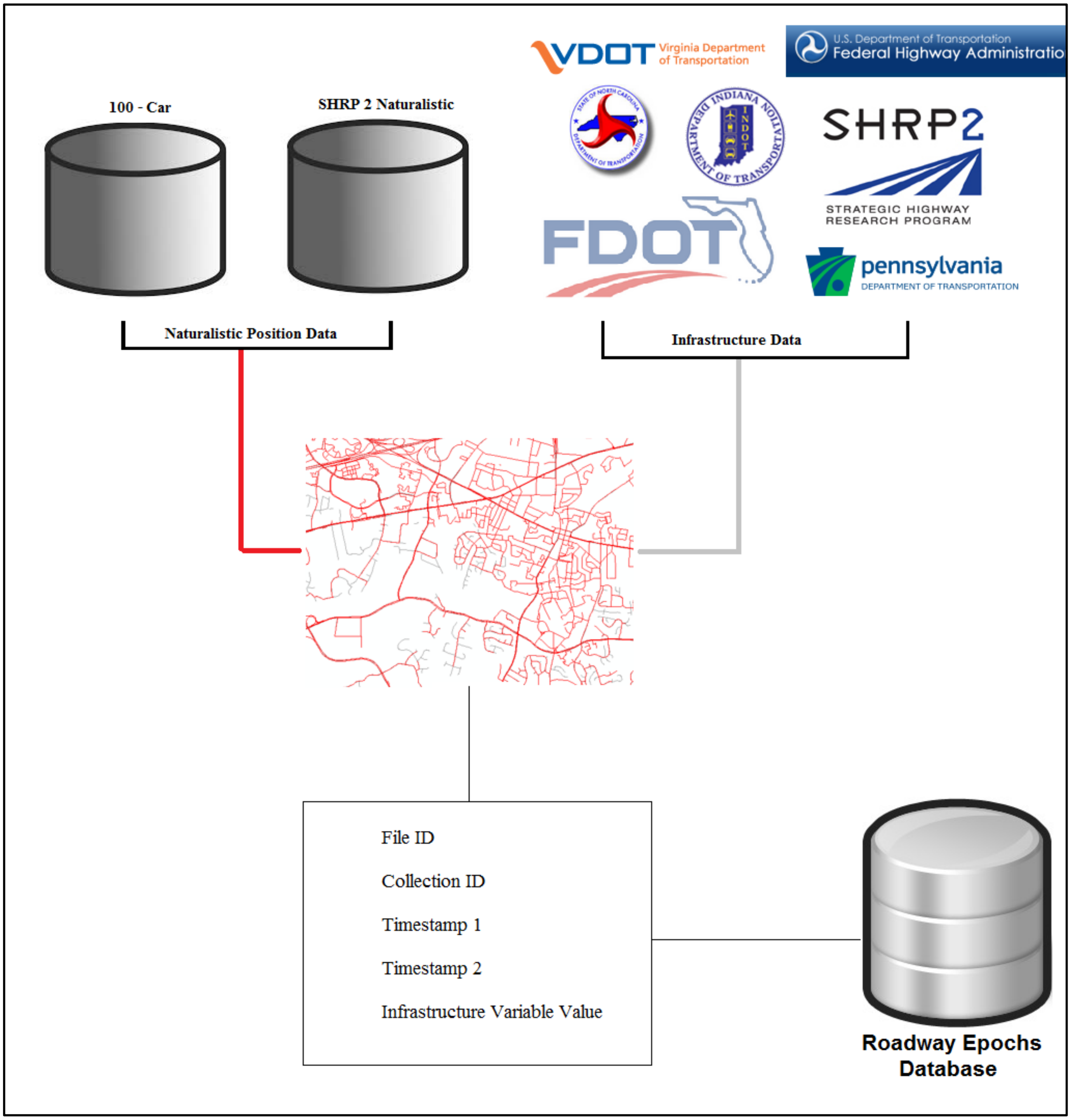

Figure 1. Chart. Conceptual outline of the Roadway Epochs Database.

The Roadway Epochs Database currently includes naturalistic data from two studies: the 100-Car Naturalistic Driving Study (NDS) and a sample of the naturalistic driving portion of the Second Strategic Highway Research Program (SHRP 2). The 100-Car NDS data were collected in 2003 and 2004 as a first-of-its-kind, large-scale naturalistic driving study. One hundred vehicles were outfitted with data acquisition units recording an array of sensor and video data. The equipped vehicles were located in northern Virginia, Maryland, and the Washington, DC, metro area. Over the course of data collection, more than 2,000,000 miles and 43,000 hours of data were amassed from 245 primary and secondary drivers (Neale, Dingus, Klauer, Sudweeks, & Goodman, 2005).

The SHRP 2 project is the largest naturalistic driving study ever performed. It included 3,147 drivers in six states and resulted in over 31 million miles of driving data that took place over 5,000,000 key-on-to-key-off trips. The Roadway Epochs Database provides data from 6,000 of those trips, representing 1,000 randomly sampled files from each of the six installation locations. For more in-depth information about the SHRP 2 data, the reader is encouraged to visit the SHRP 2 naturalistic data website (https://insight.shrp2nds.us).

The roadway data indexed in the Roadway Epochs Database consist of a combination of attributes from multiple sources, including the Federal Highway Administration (FHWA) Highway Performance and Monitoring System (HPMS), multiple state Departments of Transportation (DOTs), digital maps, and the Roadway Information Database (RID) currently being developed as a part of the SHRP 2 program. The data also include weather data from the National Oceanic and Atmospheric Administration (NOAA), such as temperature and precipitation values for the date and time a trip occurred. For a full list of available data, see Table 1.

In general, two types of questions are answerable using the Roadway Epochs Database. The first of those is a question of scope. Questions of scope simply query the database to find the total number of epochs that meet particular criteria. For instance, a researcher may be interested in identifying how many epochs occur on a tollway with a particular traffic count. The second type of question may wish to combine information from the Roadway Epochs Database with naturalistic data from either of the two included studies. For instance, a researcher might query an infrastructure-related table in the Roadway Epochs Database to locate epochs in an infrastructure category of interest (e.g., specific Roughness Index) and pair that with vehicle variables such as braking or vehicle speeds.

Driver information related to each of the epochs in the Roadway Epochs Database is stored in a separate table organized according to the file in which it was collected (file_id), rather than individual epochs. More than one epoch can be pulled from each file. The driver information table includes information such as gender, age range, and vehicle type. Driver information entries are discussed in more detail on page 14. Outside of the file table, each record (i.e., row) in the database table indicates an epoch of driving.

COMMON COLUMNS FOR EACH TABLE

To facilitate understanding, every table in the database shares a common set of variables, as described below.

file_id	The file ID is a unique number identifying the trip file within the original study collection. One trip file is typically one key-on-to-key-off ignition cycle, but sometimes trips are divided into multiple files. Currently, files cannot be rejoined into a single trip, but that is not a concern for the Roadway Epochs Database, which provides data in much shorter windows.
study	The collection ID indicates which study contains the epoch in question. The value may be an "h" for 100-Car or an "s" for SHRP 2.

timestamp_1	Timestamp 1 is the point in the trip where the epoch of interest was entered in either milliseconds or video frames depending on the collection (video frames for 100-Car, milliseconds for SHRP 2).
timestamp_2	Timestamp 2 is the timestamp corresponding to the end of an epoch in the file.
install_location	Install location indicates the location where the data acquisition system (DAS) was originally fitted to the vehicle and corresponds to the postal abbreviation of the state (i.e., VA., PA, WA, NY, FL, NC, IN).

CREATING THE ROADWAY EPOCHS DATABASE

Data Selection

The Roadway Epochs Database was created by taking input from civil engineers and roadway designers around VTTI, finding data sets that addressed that input, and then translating those data sets into usable epochs in the database. Input was gained via interviews with three researchers specializing in signage and lighting, two researchers specializing in pavement, and one graduate student with a background in civil engineering. The interviews were intended to identify the types of data that civil engineers and infrastructure researchers considered valuable for use in research. Requests from prospective users ranged from income brackets for particular areas traversed (roadway planners) to roughness indices for pavement (pavement specialists). The commonly mentioned content gleaned from those interviews guided the search for DOT and FHWA HPMS data, as well as which digital map data should be included.

The variable set represented in the Roadway Epochs Database represents a compromise between what was available and what was feasible. The database is intended as an initial iteration for researchers to gain familiarity with both content and structure. It can easily be expanded in both the number of trips and available infrastructure tables, depending on user needs. The following sections provide a brief description of the method used to join the infrastructure and naturalistic trip data, along with suggestions for further reading regarding the process.

Infrastructure variables included in the final product originate from four sources: HPMS data from the FHWA, the Roadway Information Database (RID; collected from state DOTs and aggregated by FHWA), state DOTs, and digital maps. RID variables include on-ramps, roundabouts, tollways, tunnels, and bridges. FHWA HPMS data include variables such as federally reported annual average daily traffic (AADT) and road ownership. Digital map data include information such as the number of lanes, speed limit ranges, and paved versus unpaved segments. State DOT data vary by state, but most include some traffic count variables. In addition to the four sources of infrastructure data, the NOAA provided a source for precipitation data, which was processed into a table listing the trips in the Roadway Epochs Database that were in the general area and timeframe of precipitation. While the naturalistic data were collected over several years spanning from 2003 to 2012 depending on the study, data represented in the state data tables, unless otherwise indicated, are from the year 2012.

Joining Infrastructure Data to Trips

Epochs within the database are defined by the time a driver spends traversing a section of roadway with the specific infrastructure properties of interest. Multiple infrastructure data sources were associated with trip data via digital maps that characterize the road network as a set of links. The four infrastructure data sources either already had these links identified or were associated with the digital map links as a first step in this work. Vehicle position data (latitude and longitude) over the course of the trips were matched to the same digital map links through a matching algorithm developed at VTTI. The matching algorithm provides advantages over standard spatial join techniques. First, through awareness of the underling road network connectivity, the matching algorithm significantly reduces matching errors typical of classical spatial join techniques that rely on the proximity of points to roads. Second, the matching algorithm is able to process large data sets much more quickly than classical spatial join techniques. The matching algorithm identifies each digital map link that was traversed in a trip as well as the timestamps at the beginning and end of the traversal. At the time of this document, all of the SHRP 2 and 100-Car trips have been matched to digital map links. These data are stored with the naturalistic collections. For a full review of the matching process please see *Naturalistic Driving Study: Linking the Study Data to the Roadway Information Database* (McLaughlin & Hankey, 2015).

Naturalistic driving data were joined to the infrastructure data in one of two ways. Two of the infrastructure sources (the RID and Nokia/Navteq digital maps) were already associated with links that were the same as those used to describe the driving data. In those cases, a relational join across databases allowed epochs of trips to be associated with roadway links using their digital map link identifiers. FHWA and state DOT data were downloaded from their respective sources as shapefiles to be interpreted in ArcMaps version 10.0.

The federal and state DOT shapefiles are made up of lines describing sections of roads with a constant attribute value (e.g., the same value of AADT). A section of road based on a constant attribute value may be longer or shorter, or even slightly a different shape, than the representation of the same section of road in the RID and Nokia/Navteq data. In other words, the two approaches to mapping divide up roads along their lengths differently. Inclusion of federal and state DOT information in the Roadway Epochs Database required that the state and federal links be translated to the common link system used by the VTTI matching algorithm, the RID, and the Nokia/Navteq digital map data. Each of the DOT and FHWA HPMS data sets were joined with digital map data via intersecting spatial joins and subsequently associated with trip information through relational joins.

Figure 2 provides an illustration of how the data appeared prior to the spatial join. Road centerline data are provided by the Virginia DOT shapefile. The lines in red represent roads that participants traveled, as identified by the VTTI matching algorithm output and translated into spatial objects. The lines in gray represent roads where no naturalistic driving occurred. The table resulting from the spatial join can be queried for all trips passing over the resulting links. The following section outlines the procedure used for joining the data sources.

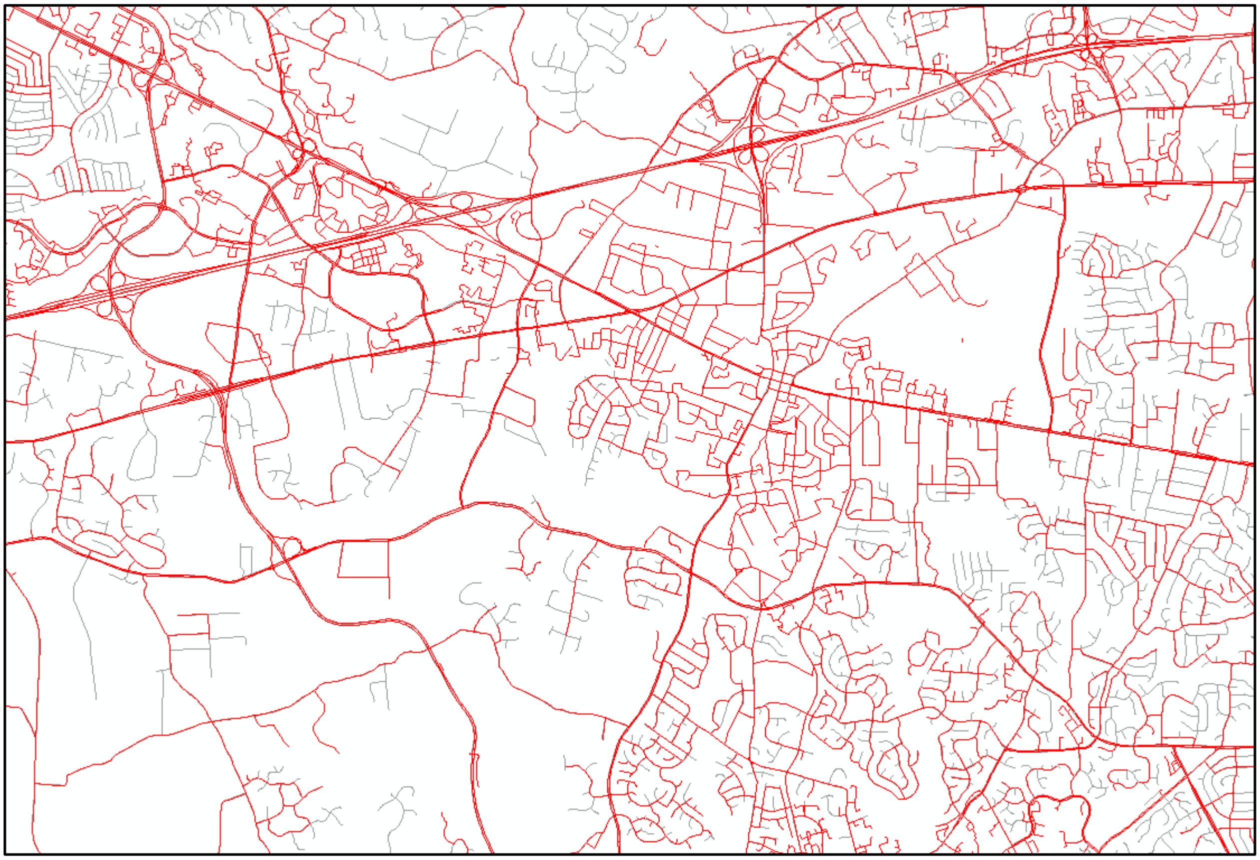

Figure 2. Map. Trips (in red) overlaid on VDOT infrastructure data (gray) prior to joining the two data sources.

State DOT and FHWA HPMS shapefiles were imported into ArcMap version 10. The data source shapefile layers were then overlaid one atop the other with matching projections. Where the spatial objects based on the matching algorithm output and infrastructure data matched, an intersecting join was performed. The spatial join resulted in the DOT and FHWA data taking on the unique identifiers of the links used in the RID, Nokia/Navteq, and the VTTI matching algorithm data.

At this point, due to the differences in the length of objects between the sources, the length attribute from the federal and state sources no longer reflected the length of the new entity as illustrated in Figure 3. Figure 3 displays a single hypothetical epoch with multiple combined data sources. The epoch is defined as the time between the two timestamps defined by Nokia/Navteq digital map data (top line) and is shared with the SHRP 2 RID data (red). State DOT and HPMS data (orange and brown) may overlay a portion of the epoch or extend beyond an individual epoch. Overlaying the links output from the matching algorithm with the digital infrastructure data from other sources has one other potential source for error. It is possible that cases exist where the state DOT, HPMS, or digital map links use a different shape. For instance, a curve in Virginia DOT data may use five points to describe the roadway geometry while RID and Nokia/Navteq data may use three. In that instance, the incongruency of the two geometries would lead to a failure for the two segments to intersect even though they are the same stretch of roadway. In the Roadway Epochs Database, driving epochs will not be identified where the two map sources fail to intersect rather than attempting approximations to eliminate those errors.

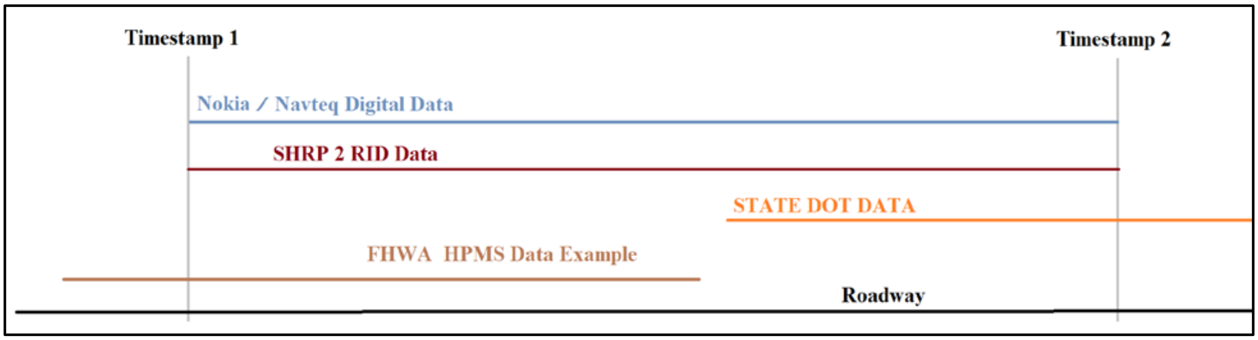

Figure 3. Chart. Combining data sets into a single epoch (time between timestamp 1 and timestamp 2) based on Navteq/Nokia links and joined across multiple data sets.

CHAPTER 2. DATABASE CONTENTS – NATURALISTIC COLLECTIONS

The two naturalistic driving studies currently included in the Roadway Epochs Database provide researchers multiple options in regard to trip and sensor data. This section will outline a small portion of available sensor data for the two collections (i.e., SHRP 2 and 100-Car). For the full set of variables for each collection, please refer to the appropriate study's primary documentation and literature (100-Car: Dingus et al., 2006; SHRP 2: https://insight.shrp2nds.us). The sensor data discussed in this chapter are held separately from the Roadway Epochs Database and reside with their naturalistic driving collections at VTTI. Access to sensor data from either the 100-Car or SHRP 2 study will require approval from the Institutional Review Board (IRB) for work with human subjects. Access to SHRP 2 data will additionally require a data use license (https://insight.shrp2nds.us). Required IRB forms may be found at http://www.irb.vt.edu. Please check with your local IRB regarding their rules of data access; requirements beyond those implemented by VTTI may apply. Additionally, access to personally identifiable information such as face cameras will require additional stipulations.

100-CAR VEHICLE AND SENSOR DATA

Accelerometer variables include lateral and longitudinal accelerations recorded at a standard rate of 10 Hz. Note: 100-Car acceleration and gyroscopic measurements do not follow the SAE J670 standard.

> *Lateral Accelerations* (labeled as *x*-axis): Acceleration values range from −6g to 6g. Lateral movements from the right of the sensor are recorded as positive.
>
> *Longitudinal Accelerations* (labeled as *y*-axis): Acceleration values range from −6g to 6g. Forward motions are recorded as positive.

GPS variables include speed, heading, latitude, and longitude, and were recorded once per second.

> *Latitude*: The data conform to WGS84 with 0 to 90 degrees north cataloged as positive values, and 0 to 90 degrees south as negative.
>
> *Longitude*: The data conform to WGS84 with 0 to 180 degrees east indicated by positive values, and 0 to 180 degrees west indicated by negative values.
>
> *Speed*: GPS speed is reported in km/h with an accuracy of ±1 km/h and a resolution of 0.02 km/h.
>
> *Heading*: Heading is reported in degrees ranging from 0 to 359 and represents the compass heading of the vehicle taken every second with an accuracy of ±1.5 degrees.

A gyroscope located in the DAS recorded yaw rates at 10 Hz.

> *Yaw Rate*: Yaw rate is recorded angular velocity around the vertical (*z*-axis). It was recorded at 10 Hz in deg/s.

Timestamp: Timestamps for 100-Car are reported as video frames since the beginning of the trip.

Cameras: There were five camera views monitoring the driver's face (blurred in this report) and driver's side view of the road, the forward road, the rear road view, the passenger side road view, and an over-the-shoulder view for the driver's hands and surrounding areas. Camera views are shown in Figure 4.

Figure 4. Photo. A compressed video image from the 100-Car Study data. The driver's face (upper left quadrant) has been blurred to protect the driver's identity. The lower right quadrant is split between the left-side view (top) and the rear view (bottom). Definition of camera views, image, and caption from Dingus et al., 2006.

SHRP 2 SAMPLE

SHRP 2 sensor data conform to the SAE J670 standard for vehicle dynamics. Lateral movement is indicated on the *y*-axis, longitudinal movement is indicated on the *x*-axis, and vertical movement is indicated on the *z*-axis. A graphical representation of the coordinate system is provided in Figure 5. Definitions of sensor variables conform to the SHRP 2 Time Series Data Dictionary. A few of the driving-related variables from the SHRP 2 collection that are available in the Roadway Epochs Database are presented in the following section. The reader is encouraged to see the SHRP 2 data dictionary on the Insight Data Access website (https://insight.shrp2nds.us).

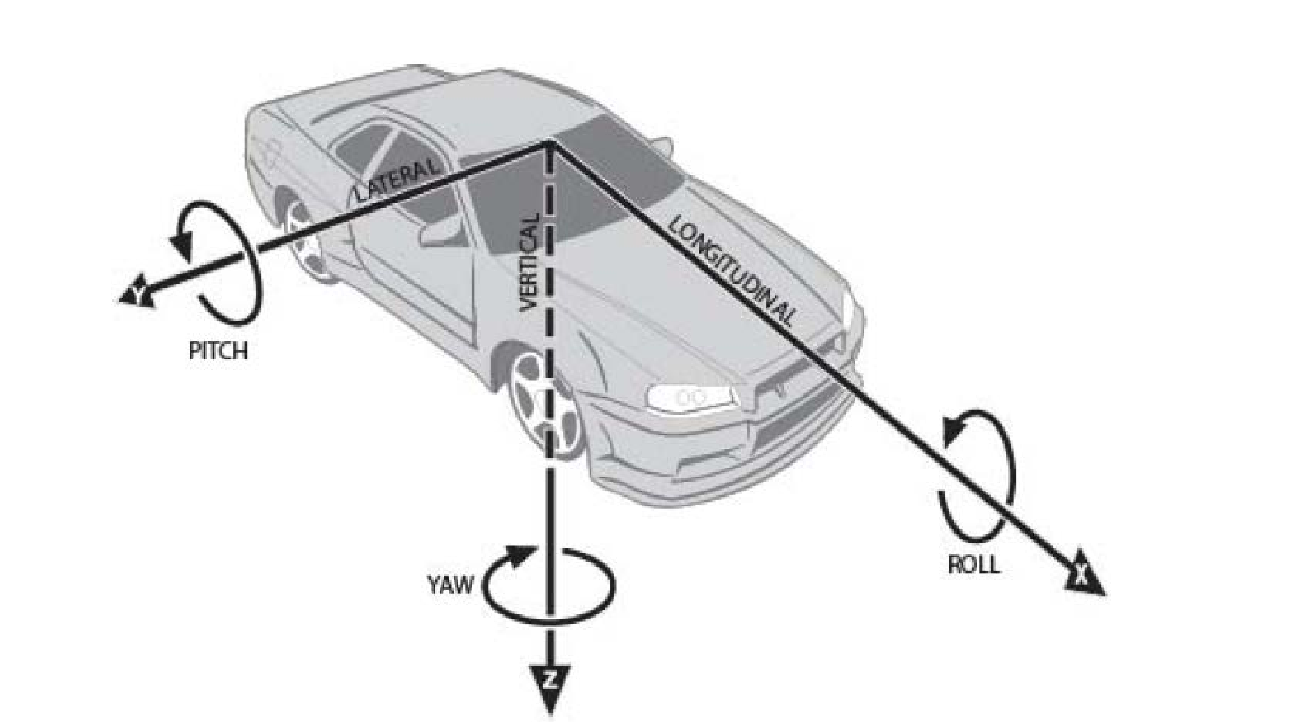

Figure 5. Diagram. SAE standard axes (SAE J670).

Sensor Data

GPS variables include speed, heading, latitude, and longitude, and were recorded once per second from a portion of the DAS mounted on the windshield near the rearview mirror.

> *Latitude*: The data conform to WGS84 with 0 to 90 degrees north cataloged as positive values, and 0 to 90 degrees south as negative.
>
> *Longitude*: The data conform to WGS84 with 0 to 180 degrees east indicated by positive values, and 0 to 180 degrees west indicated by negative values.
>
> *Speed*: GPS speed is reported in km/h with an accuracy of ±1 km/h and a resolution of 0.02 km/h.
>
> *Heading*: Heading is reported in degrees ranging from 0 to 359 and represents the compass heading of the vehicle taken every second with an accuracy of ±1.5 degrees.

Accelerometer variables include lateral and longitudinal accelerations recorded at a standard rate of 10 Hz.

> *Lateral Accelerations* (labeled *y*-axis): Acceleration values range from −6g to 6g. Lateral movements from the right of the sensor are recorded as positive.

Longitudinal Accelerations (labeled *x*-axis): Acceleration values range from −6g to 6g. Forward motions are recorded as positive.

A gyroscope located in the DAS recorded yaw rates at 10 Hz.

Yaw Rate: Yaw rate is recorded angular velocity around the vertical (*z*-axis). It was recorded at 10 Hz in deg/s.

A front-mounted radar unit provided target information for up to eight targets, recorded at 12 Hz for the *x*- and *y*-axes.

Target Identification: An integer value that increases with each new identified target, beginning with 1. The value is used to differentiate one radar target from others.

Range Forward (*x*-axis): Range information on forward targets measured on the *x*-axis in meters. Range values span between 1.5 and 240 meters with a resolution of 1.8 meters and a 0.5-meter accuracy up to 10 meters.

Range Lateral (*y*-axis): Range information to targets measured on the *y*-axis. Laterally to the left of the radar is recorded as a negative value, and to the right is positive. Range values span between 1.5 and 240 meters with a resolution of 1.8 meters and 0.5-meter accuracy up to 10 meters.

Range Rate Forward (*x*-axis): Range rate to forward radar targets on the *x*-axis. Values are coded positive for separating vehicles and negative for closing vehicles. Values are between −250 km/h and 250 km/h with an accuracy of 0.25 km/h and a resolution of 0.2 m/s.

Range Rate Lateral (*y*-axis): Range rate to forward targets measured on the *y*-axis. Positive values indicate laterally separating vehicles, and negative values laterally closing vehicles. The data may span from −250 to 250 km/h with a resolution of 0.2 m/s and accuracy to 0.25 km/h.

Vehicle Data

Turn Signal Indicator: Indicates the state of illumination of the turn signals. Data are coded as 0 = off, 1 = left, and 2 = right. The data were recorded at 1 Hz.

Accelerator Pedal Position: Gives the position of the accelerator pedal collected from the vehicle network and normalized using manufacturer specifications.

Brake Pedal Activation: Indicates an on or off press of the brake pedal. Data were collected at 2 Hz and are coded as 0 = off and 1 = on.

Steering Wheel Position: Gives the position and direction of the steering wheel from neutral position. Data were recorded in degrees at its native resolution with values ranging from −720 to

720 degrees. Clockwise rotations are recorded with positive values and counterclockwise as negative.

Network Speed: Indicates the vehicle speed collected from the vehicle network. Values range from −80 km/h to 240 km/h and were collected at native resolution.

Seat Belt Driver: Indicates whether or not the seat belt is being used by the driver; 0 = engaged, 1 = not engaged.

Timestamp: Millisecond timestamps for every recorded value.

Cameras: Four continuously recording videos were captured that included views of the forward roadway, the driver's face, center stack and pedal, and the rear view. There was also a periodic snap shot of the cabin that was blurred to allow things like the number of passengers to be identified. The views are shown in Figure 6.

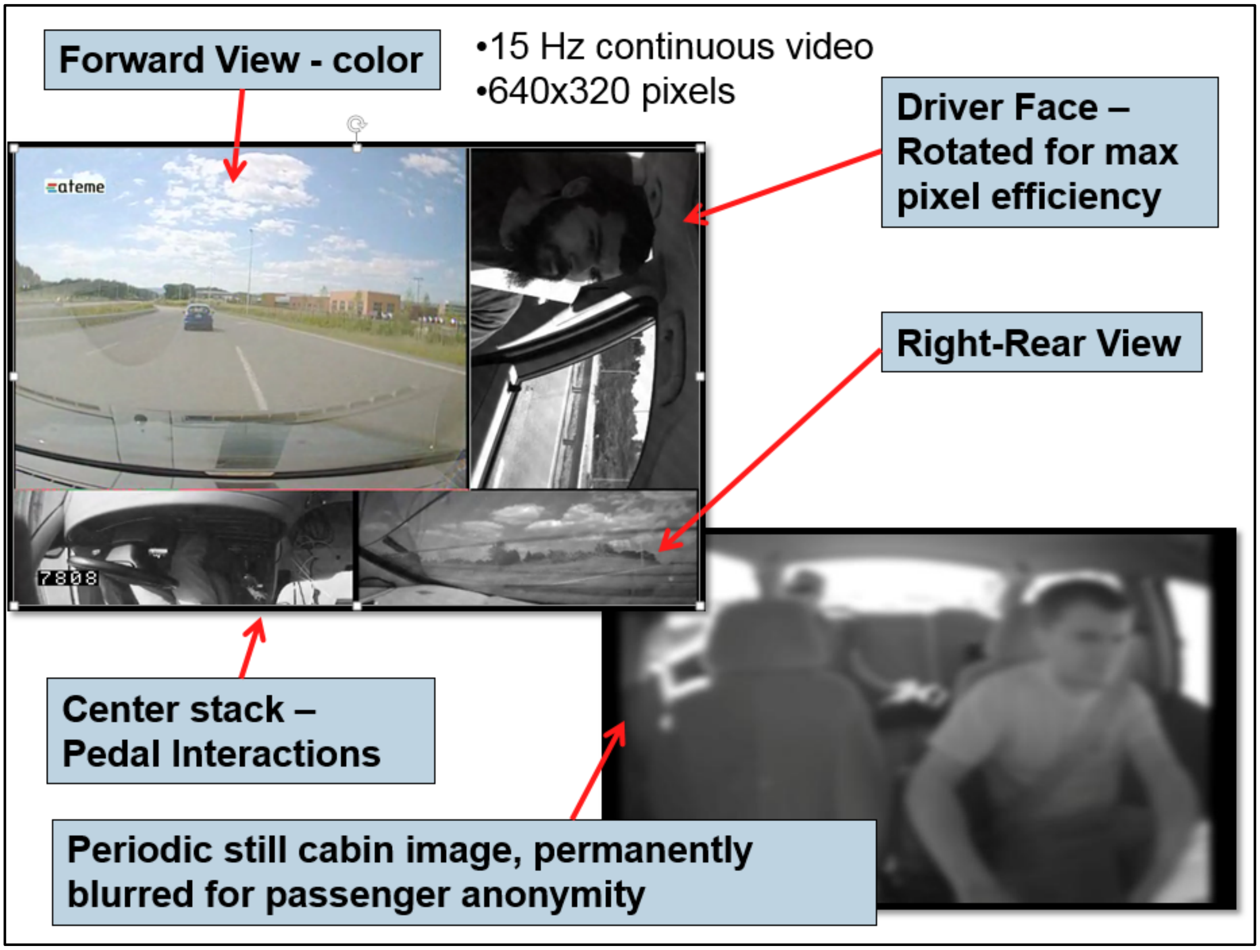

Figure 6. Photos. SHRP 2 camera images from Perez, 2014.

File Table

The file table (Figure 7) provides vehicle and trip information for every epoch in the database. The information available varies by study, but at minimum a vehicle class and participant ID will be present. Other columns in the table include the year, make, and model of a vehicle where available, as well as the local date and time for when the trip occurred. Entries vary for 100-Car and SHRP 2 trips.

File_id	The identifier within each study for an individual trip.
Participant_id	Unique identifier for each participant in the study.
Local_date_time	The local date and time for the beginning of a trip.
Vehicle Make	The manufacturer of the participant vehicle.
Vehicle Model	The model of the participant vehicle.
Vehicle Class	The type of vehicle driven in the study.
Age	The age of the participant in years.
Gender	The gender of the participant, male or female.

	file_id	vtti_participant_id	local_date_time	vehicle_year	vehicle_make	vehicle_model	vehicle_class
1	9245810	170	2011-09-16 09:27:57.537	2005	Toyota	Prius	CAR
2	37665662	5080926	2012-05-19 13:48:51.027	2004	Lexus	RX	SUV_CROSSOVER
3	12058470	489070	2011-10-15 21:38:42.527	2006	Hyundai	Sonata	CAR
4	18442666	1490	2011-11-07 14:03:35.683	2010	Toyota	Corolla	CAR
5	25413851	502931	2012-06-08 17:35:59.683	2005	Nissan	Sentra	CAR
6	72050357	5080940	2013-04-26 10:32:45.643	2006	Dodge	Caravan	VAN_MINIVAN
7	27155324	497011	2012-06-22 21:29:06.187	2008	Ford	Focus	CAR
8	70055944	502201	2013-01-10 12:44:07.430	2011	Kia	Soul	SUV_CROSSOVER

Figure 7. Screenshot. Sample of file-specific table.

CHAPTER 3. DATABASE CONTENTS – INFRASTRUCTURE COLLECTIONS

The infrastructure-related data available in the Roadway Epochs Database vary by location of the trips and the source of the data. Four high-level categories are available: general digital map data, roadway attributes from federal sources, roadway attributes from the SHRP 2 RID, and attributes accumulated from state DOTs. Digital map and federal data are available for trips that occurred in any of the contiguous 48 U.S. states. RID information is available for the six SHRP 2 sites (Washington, Pennsylvania, North Carolina, New York, Indiana, and Florida). State DOT data are available for four of the SHRP 2 states and Virginia.

This section is organized by data source and follows the structure presented in Table 1. Each infrastructure table is accompanied by a brief description as well as a screen capture of the table to familiarize the reader with the data present. Table 1 provides the reader with a general overview of the infrastructure tables present in the Roadway Epochs Database. An X indicates the presence of the variable for that particular study. It is important to note that if a vehicle from 100-Car were to enter one of the SHRP 2 states or vice versa, that trip would have all the variables available to epochs from the study for which the location was originally included. That is, the infrastructure variables of interest are available to any trip which traversed those links, regardless of collection.

The RID is composed of roadway data collected as part of SHRP 2 and is identified by the prefix "epochs_rid_" followed by a data label. Federal tables are identified by the prefix "epochs_federal_" followed by the variable ID (AADT, access_control, HOV, etc.). Epochs related to Navteq/Nokia digital maps are identified with the prefix "epochs_dmapping_" and the variable ID.

The following sections describe the data that have been included in the Roadway Epochs Database. The tables are organized according to the predominant naturalistic data set in each table. No tables are exclusive to either SHRP 2 or 100-Car. The collection containing the epoch of interest is listed along with the file ID for each entry. While the SHRP 2 RID was collected predominantly for the SHRP 2 effort, there are many 100-Car traversals made by participants leaving the installation area (northern Virginia, Washington D.C., and southern Maryland).

Table 1. Infrastructure tables by study in Roadway Epochs Database.

	100-Car	**SHRP 2**
Nokia/Navteq Attributes		
dbo.epochs_dmapping_frontage	X	X
dbo.epochs_dmapping_functional_class	X	X
dbo.epochs_dmapping_lane_categories	X	X
dbo.epochs_dmapping_paved	X	X
dbo.epochs_dmapping_ramps	X	X
dbo.epochs_dmapping_speed_limits	X	X
dbo.epochs_dmapping_transition_areas	X	X
FHWA HPMS		
dbo.epochs_federal_aadt	X	X
dbo.epochs_federal_access_control	X	X

	100-Car	SHRP 2
dbo.epochs_federal_hov	X	X
dbo.epochs_federal_international_roughness_index	X	X
dbo.epochs_federal_ownership	X	X
dbo.epochs_federal_route_sign	X	X
dbo.epochs_federal_through_lanes	X	X
dbo.epochs_federal_functional_system	X	X
SHRP 2 Roadway Information Database		
dbo.epochs_rid_bridge_segments	X	X
dbo.epochs_rid_functional_class	X	X
dbo.epochs_rid_ramp_segments	X	X
dbo.epochs_rid_roundabouts	X	X
dbo.epochs_rid_tollway_segments	X	X
dbo.epochs_rid_tunnel_segments	X	X
State Departments of Transportation		
dbo.epochs_state_florida_2013_AADT		X
dbo.epochs_state_florida_2013_traffic_percentages		X
dbo.epochs_state_virginia_average_annual_weekday_traffic	X	
dbo.epochs_state_virginia_traffic_counts	X	
dbo.epochs_state_virginia_traffic_percentages	X	
dbo.epochs_state_virginia_k_factor	X	
dbo.epochs_state_pennsylvania_aadt		X
dbo.epochs_state_pennsylvania_daily_truck_vmt		X
dbo.epochs_state_pennsylvania_daily_vmt		X
dbo.epochs_state_pennsylvania_percent_truck		X
dbo.epochs_state_pennsylvania_traffic_count		X
dbo.epochs_state_pennsylvania_truck_AADT		X
dbo.epochs_north_carolina_percent_traffic_type		X
Other		
dbo.epochs_precipitation		X

NOKIA/NAVTEQ ATTRIBUTES

The digital map tables utilize the underlying linking structure for the VTTI matching algorithm as explained in Chapter 1. The Roadway Epochs Database and the Nokia/Navteq data set share a common underlying link architecture. Every link traversal represented in the Roadway Epochs Database is present in the digital mapping tables, for a total of roughly 757,600 traversals. Digital mapping tables include the identification of frontage roads, a functional class system, the number of total lanes on a roadway, paved and unpaved links, identification of on- and off-ramps, speed limits, and identified roadway transition areas.

Frontage

The frontage table (Figure 8) lists epochs that occurred on a frontage road. Epochs that were not matched to frontage links are not represented in this table.

	file_id	timestamp_1	timestamp_2	study	location
1	58975	20270	20280	h	va
2	110385	8493	8553	h	va
3	77038	50751	50822	h	va
4	175018	15204	15274	h	va
5	175348	11606	11667	h	va
6	35533	7166	7596	h	va
7	55388	3352	3362	h	va
8	202007	8148	8188	h	va

Figure 8. Screenshot. Sample of digital map frontage road table.

Functional Class

The functional class variable represented in the digital map tables (Figure 9) is not the same functional class designation that is given in the federal tables. It is a unique functional class designation, ranked from 1 to 5, that loosely correlates to residential streets on one end (5) and major throughways, such as a controlled access interstate, on the other (1).

	file_id	timestamp_1	timestamp_2	study	location	functional_class
1	123064	23781	23892	h	va	1
2	156246	33387	33417	h	va	1
3	91981	5120	5260	h	va	2
4	116444	13951	14072	h	va	2
5	186704	13	25	h	va	2
6	106683	5523	5542	h	va	2
7	171745	2535	2556	h	va	4
8	182894	690	740	h	va	2

Figure 9. Screenshot. Sample of digital map functional class table.

Lane Categories

The lane categories table (Figure 10) indicates the number of total lanes present on a link. The lane number applies to total lanes present in each direction on a roadway.

	file_id	timestamp_1	timestamp_2	study	location	lane_grouping_per_direction
1	71871	11369	11379	h	va	2-3
2	77054	4381	4431	h	va	2-3
3	70278	4252	4291	h	va	1
4	186159	13051	13071	h	va	2-3
5	83309	11548	11948	h	va	1
6	160711	4952	5053	h	va	2-3
7	168465	6160	6220	h	va	2-3
8	56274	6579	6609	h	va	3+

Figure 10. Screenshot. Sample of digital map lane categories table.

Paved

The paved table (Figure 11) indicates whether or not an epoch occurred on a paved roadway. Paved epochs will have a value of "Y" and nonpaved epochs have a value of "N" in the paved column. There are roughly 1,100 epochs of unpaved roadway.

	file_id	timestamp_1	timestamp_2	study	location	paved
1	76230248	481897	495897	s	ny	Y
2	168108	6005	6065	h	va	Y
3	10752	11254	11275	h	va	Y
4	191476	10388	10617	h	va	Y
5	106890	2341	2700	h	va	Y
6	164730	10104	10223	h	va	Y
7	88736	2487	2677	h	va	Y
8	97574	25249	25649	h	va	Y

Figure 11. Screenshot. Sample of digital map paved table.

Ramps

The ramps table (Figure 12) indicates epochs that occurred on an on-ramp or an off-ramp. Only those epochs that contained a ramp are included in the table.

	file_id	timestamp_1	timestamp_2	study	location
1	113485	4175	4185	h	va
2	42871	15489	15490	h	va
3	177983	4138	4147	h	va
4	45689	9976	9986	h	va
5	201999	21781	21782	h	va
6	155245	3296	3326	h	va
7	73691	11713	11714	h	va
8	43714	14706	14716	h	va

Figure 12. Screenshot. Sample of digital map ramps table.

Speed Limit

The speed limit table (Figure 13) provides speed limit data for each epoch. Speed limit ranges are as they appear in the Nokia/Navteq data set. The speed_limit_source_txt column indicates the data source identified for these epochs in the Nokia/Navteq digital map database. Speed_limit_source_txt entries of "posted" indicate a posted speed limit. Speed_limit_source_txt entries of "derived" appear for derived speed limits. Derived speed limits are those that have been estimated based on a number of factors such as average traffic flow and roads with similar characteristics. Roughly 10% (81,000) of epochs list a derived speed limit.

	file_id	timestamp_1	timestamp_2	study	location	speed_limit	speed_limit_source_txt
1	62648	7327	7338	h	va	55-64 mph	posted
2	43536	9141	9251	h	va	55-64 mph	posted
3	155156	3203	3333	h	va	31-40 mph	posted
4	14296	3612	3703	h	va	55-64 mph	posted
5	89117	10636	10696	h	va	55-64 mph	posted
6	183017	3833	3868	h	va	31-40 mph	posted
7	146651	49180	49260	h	va	55-64 mph	posted
8	69846	2455	2485	h	va	31-40 mph	posted

Figure 13. Screenshot. Sample of digital map speed limit table.

Transition Areas

Roughly 1,778 transition epochs are identified in the Roadway Epochs Database (Figure 14). A transition area is one in which the number of traffic lanes increases or decreases.

	file_id	timestamp_1	timestamp_2	study	location
1	85517	15250	15370	h	va
2	180106	21087	22107	h	va
3	155324	15066	15176	h	va
4	133496	10340	10390	h	va
5	119622	13549	13609	h	va
6	162964	17586	18256	h	va
7	122627	7913	8042	h	va
8	23793	15637	15737	h	va

Figure 14. Screenshot. Sample of digital map transition area table.

FHWA HPMS TABLES

FHWA HPMS tables consist of Highway Performance Monitoring System data for federally controlled roadways in the SHRP 2 and 100-Car epochs. Tables provide information such as International Roughness Index (IRI) values, functional class information, access control, and federal AADT numbers. FHWA HPMS tables in the Roadway Epochs Database contain roughly 59,000 epochs each. The information in the FHWA HPMS tables was collected by state DOTs and reported to the federal government. It was then aggregated into FHWA HPMS shapefiles containing each of the contiguous 48 states. Covered roadways in the FHWA HPMS data set are illustrated below in Figure 15.

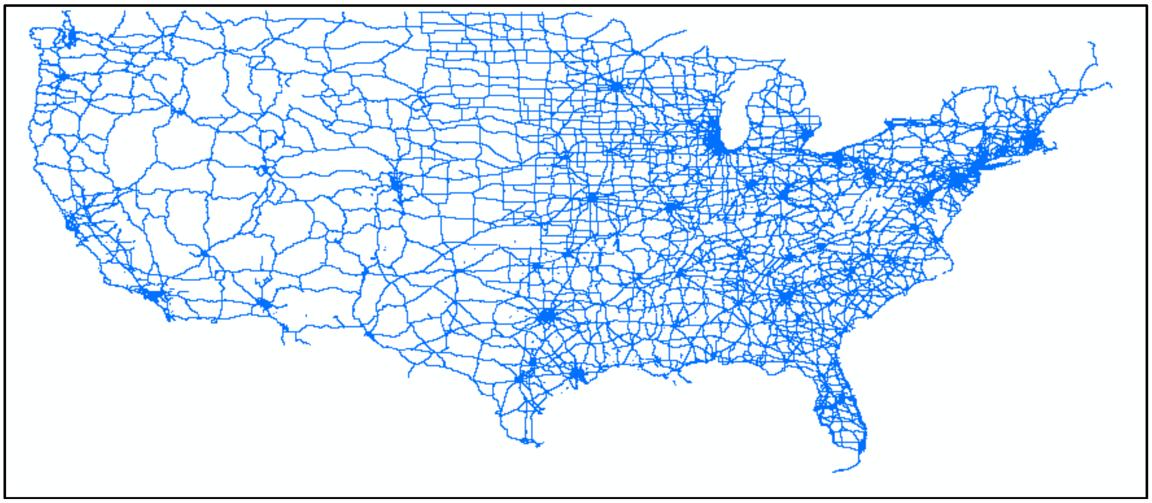

Figure 15. Map. FHWA HPMS Coverage.

AADT

The AADT table (Figure 16) provides measures of average annual daily traffic as reported in the HPMS provided by the FHWA.

	file_id	timestamp_1	timestamp_2	study	location	aadt_vn
1	54880	26076	26526	h	va	11645
2	60743	8592	8632	h	va	16530
3	79043	2317	2567	h	va	23400
4	96870	17900	18030	h	va	118000
5	32146	5648	5717	h	va	24148
6	173918	4023	4363	h	va	178559
7	65822847	744126	754126	s	ny	128400
8	104170	9280	9400	h	va	51549

Figure 16. Screenshot. Sample of federal AADT table.

Access Control

The access control table (Figure 17) identifies the degree of control over access to a location. Access control is categorized using a three-level scale. The table lists the epoch of interest along with the numerical code provided by the FHWA HPMS and a text translation of that value.

	file_id	timestamp_1	timestamp_2	study	location	ACCESS_CON	access_control_txt
1	75825689	119961	121961	s	wa	2	partial
2	18267637	207610	212610	s	wa	2	partial
3	44720527	110795	121796	s	wa	2	partial
4	41284182	746020	761020	s	wa	2	partial
5	49837132	676034	710035	s	wa	2	partial
6	46756571	874078	881079	s	wa	2	partial
7	64085908	991909	992910	s	wa	2	partial
8	26899108	3508819	3512819	s	wa	1	full

Figure 17. Screenshot. Sample of federal access control table.

HOV Lanes

The federal HOV table (Figure 18) contains traversals of high-occupancy vehicle (HOV) lanes identified by the HPMS data. The table displays whether the lanes are full-time or part-time HOV lanes using an integer value provided by the original shapefile and a text translation.

	file_id	timestamp_1	timestamp_2	study	location	hov_lanes	hov_lane_type
1	151731	4588	4619	h	va	2	normal_lane
2	102221	19246	19356	h	va	2	normal_lane
3	13475	17569	17579	h	va	2	normal_lane
4	47748	8793	9704	h	va	2	normal_lane
5	30464	22965	23095	h	va	2	normal_lane
6	134825	14490	14560	h	va	2	normal_lane
7	114358	27184	27285	h	va	2	normal_lane
8	13229	6275	6355	h	va	2	normal_lane

Figure 18. Screenshot. Sample of federal HOV table.

International Roughness Index (IRI)

The IRI table (Figure 19) provides the federally reported IRI index to the nearest inch per mile from the FHWA HPMS shapefile and contains approximately 300,000 epochs. It is important to note that the links used to form the epochs listed in the database are longer than the data collection area for a particular IRI measurement. A roadway epoch may extend for a significant portion of a mile while the IRI sample for that epoch would be no longer than 0.10 miles, as per federal standards. The IRI measurement, however, would have occurred someplace on the link used to form the epoch of interest.

	file_id	timestamp_1	timestamp_2	study	location	IRI
1	72857	21449	21840	h	va	160
2	139551	2604	2674	h	va	165
3	177888	3115	3175	h	va	85
4	99414	34168	34297	h	va	70
5	47608	25061	25171	h	va	48
6	133972	1816	1996	h	va	85
7	134151	34516	34632	h	va	145
8	148801	10929	11019	h	va	105

Figure 19. Screenshot. Sample of federal IRI table.

Ownership

Ownership tables identify the entity that has legal ownership of a particular segment of roadway. The table (Figure 20) contains the epochs of interest, the integer value represented in the federal data, and a text translation of that data.

	file_id	timestamp_1	timestamp_2	study	location	ownership	ownership_txt
1	106156	176	486	h	va	1	state_highway_agency
2	170419	7443	7453	h	va	1	state_highway_agency
3	181828	8752	8852	h	va	1	state_highway_agency
4	54121	5428	5458	h	va	1	state_highway_agency
5	116519	10273	11934	h	va	1	state_highway_agency
6	99511	22920	22990	h	va	1	state_highway_agency
7	8804	2244	2314	h	va	1	state_highway_agency
8	59078	13291	13301	h	va	1	state_highway_agency

Figure 20. Screenshot. Sample of federal ownership table.

Route Sign

The routes sign table (Figure 21) indicates whether an epoch on the federal tables is signed as a U.S. highway, a state highway, or an interstate highway. The table consists of both the numerical designation given by the FHWA HPMS data and the text translation.

	file_id	timestamp_1	timestamp_2	study	location	route_sign	txt_description
1	106156	176	486	h	va	3	US
2	181828	8752	8852	h	va	2	interstate
3	54121	5428	5458	h	va	2	interstate
4	116519	10273	11934	h	va	2	interstate
5	99511	22920	22990	h	va	4	state
6	8804	2244	2314	h	va	4	state
7	59078	13291	13301	h	va	3	US
8	149283	22116	23472	h	va	2	interstate

Figure 21. Screenshot. Sample of federal route sign table.

Through Lanes

The federal through lanes table (Figure 22) indicates the number of lanes designated for through-traffic for individual epochs. Turn lanes are not included in the number of through lanes in each epoch.

	file_id	timestamp_1	timestamp_2	study	location	through_lanes
1	57641	4981	5110	h	va	4
2	76781	27306	27336	h	va	4
3	150793	3164	3445	h	va	4
4	8575	95742	96093	h	va	6
5	102691268	421651	430651	s	ny	4
6	77056	24707	24748	h	va	6
7	125475	3494	3604	h	va	4
8	148152	2479	2549	h	va	4

Figure 22. Screenshot. Sample of federal through lanes table.

Functional System

The functional system table (Figure 23) identifies the federal functional class designations of roadways by both numerical value and the textual translation of that value. It is important to note that the federal functional class and the digital map functional class are two separate systems that do not necessarily correlate.

	file_id	timestamp_1	timestamp_2	study	location	functional_system_value	functional_system_txt
1	112503	27861	28523	h	va	1	interstate
2	115277	1740	1830	h	va	1	interstate
3	29336	4959	5978	h	va	1	interstate
4	123188	7685	7864	h	va	7	local
5	174498	41691	41941	h	va	1	interstate
6	19864	5442	5572	h	va	1	interstate
7	65213	17990	18011	h	va	3	principal_arterial_other
8	186017	2479	2609	h	va	3	principal_arterial_other

Figure 23. Screenshot. Sample of federal functional system table.

RID

The naturalistic portion of the data is in the process of being supplemented by the RID, which consists of 25,000 centerline miles of roadway data collected as a portion of the SHRP 2 effort and combined with state inventories in each of the SHRP 2 states. At the time of this writing, the RID effort is currently still in progress. The RID contains variables describing a range of roadway-related attributes for a subsample of roads in the six states representing the six locations for the naturalistic driving study (Florida, North Carolina, Indiana, Pennsylvania, New York, and Washington). The road network surrounding the central area of data collection has been retained in its entirety in the RID, whereas more distant roads are not necessarily included. The patches of dense road network visible in Figure 24 indicate where the full road network is included. For a full data dictionary of the RID, please see http://www.ctre.iastate.edu/shrp2-rid/rid.cfm. The RID Metadata page is updated as more information is added to the RID. The following tables examine the number of times that the 6,000 randomly selected SHRP 2 NDS and all of the 100-Car vehicles traversed any of the spatially defined RID segments. If a section of roadway present in the RID was traversed, an entry was made in the tables below for that section. Areas of the RID that were not traversed within the narrow scope of the Roadway Epochs project were not included.

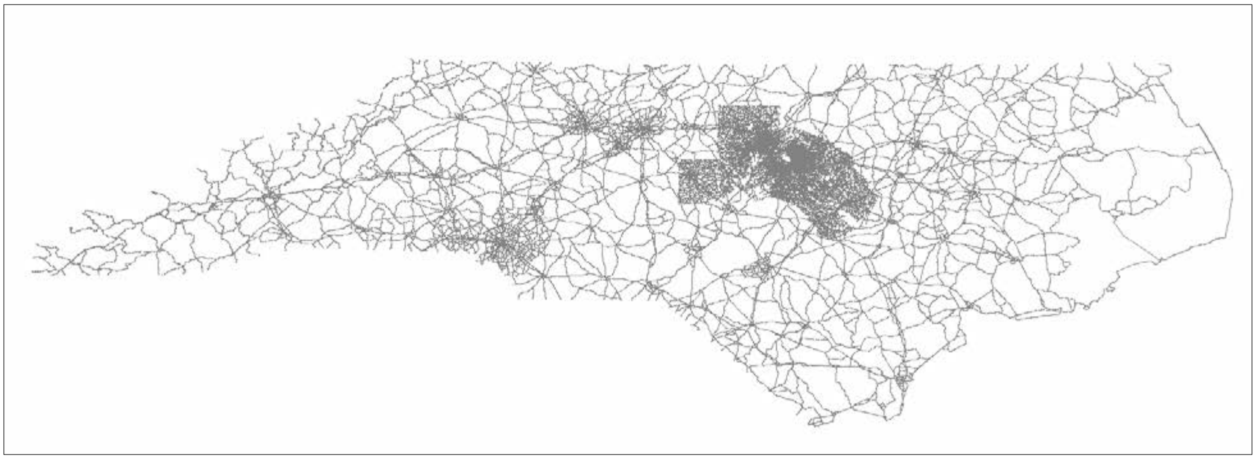

Figure 24. Map. RID representation of North Carolina.

Bridge Segments

The bridge segments table (Figure 25) indicates epochs that occurred on a bridge in the SHRP 2 states (North Carolina, New York, Pennsylvania, Washington, Florida, and Indiana). The timestamps represent an approximation of the actual time spent on the bridge. There are no entries for non-bridge segments in the bridge table. There are 5,923 bridge traversals currently in the database.

	FILE_ID	timestamp_1	timestamp_2	study	location
1	28168832	4645523	4646523	s	nc
2	19209776	1428196	1442196	s	nc
3	5039	9833	9862	h	va
4	149020	1005	1756	h	va
5	186001	2115	2915	h	va
6	145807	47498	47507	h	va
7	172859	23242	23542	h	va
8	4585	10530	10580	h	va

Figure 25. Screenshot. Sample of RID bridge segments table.

Functional Class

Functional class in the RID data is distinct from the functional class provided in the FHWA HPMS tables. The values range from 1 through 5, with a 1 indicating a higher traffic volume road and a 5 lower traffic volume. Approximately 350,000 epochs are contained in the RID functional class table (Figure 26).

	FILE_ID	timestamp_1	timestamp_2	study	location	functional_class
1	25889585	67444	75444	s	nc	5
2	25889585	75444	80444	s	nc	5
3	25889585	80444	81444	s	nc	5
4	25889585	81444	82444	s	nc	5
5	25889585	82444	86444	s	nc	5
6	25889585	86444	95444	s	nc	5
7	25889585	95444	96445	s	nc	5
8	25889585	96445	105445	s	nc	5

Figure 26. Screenshot. Sample of RID functional class table.

Ramp

On- and off-ramps were identified as a part of the RID portion of SHRP 2. There are 9,393 on-ramp and off-ramp epochs in the RID data of the Roadway Epochs Database. Figure 27 provides a screenshot of the RID ramps table.

	FILE_ID	timestamp_1	timestamp_2	study	location
1	27731422	1765982	1788983	s	nc
2	96875	394	555	h	va
3	28037562	258451	265451	s	nc
4	21874	40146	40416	h	va
5	170983	14759	14789	h	va
6	170983	14609	14749	h	va
7	19135682	1463706	1470706	s	nc
8	24804548	532808	534808	s	nc

Figure 27. Screenshot. Sample of RID ramps table.

Roundabout

There are 567 roundabout traversals currently in the Roadway Epochs Database. Roundabouts were identified through the RID as a part of SHRP 2. Figure 28 provides a screenshot of the RID roundabout table.

	FILE_ID	timestamp_1	timestamp_2	study	location
1	55422634	423407	424407	s	nc
2	171986	68913	68923	h	va
3	45519159	1591100	1594100	s	nc
4	28500618	107285	108285	s	nc
5	34658061	2175891	2176891	s	nc
6	171986	68902	68913	h	va
7	171987	6863	6873	h	va
8	25837044	965688	968688	s	nc

Figure 28. Screenshot. Sample of RID roundabout table.

Tollway Segments

The tollway epochs table (Figure 29) indicates time spent on a tollway in the RID states. As shown below, a single trip may have multiple consecutive epochs on the same stretch of tollway due to the linking process. To see the traversals in the appropriate order, order the query by timestamp. There are 12,621 tollway epochs in the database.

	FILE_ID	timestamp_1	timestamp_2	study	location
1	32833863	1229471	1232472	s	nc
2	32833863	1058469	1062469	s	nc
3	32833863	1210471	1214471	s	nc
4	32833863	1140470	1160470	s	nc
5	32833863	1241472	1250472	s	nc
6	32833863	1062469	1095469	s	nc
7	32833863	1160470	1164470	s	nc
8	32833863	1164470	1210471	s	nc

Figure 29. Screenshot. Sample of RID tollway segments table.

Tunnel

The tunnel segments table (Figure 30) indicates epochs that occurred in a tunnel in the SHRP 2 states (North Carolina, New York, Pennsylvania, Washington, Florida, and Indiana). The timestamps represent an approximation of the actual time spent in the tunnel. There are no entries for non-tunnel segments in the tunnel table. There are 373 tunnel epochs in the database.

	FILE_ID	timestamp_1	timestamp_2	study	location
1	97135	32155	32245	h	va
2	97135	32145	32155	h	va
3	143698	46652	47512	h	va
4	203535	54795	54805	h	va
5	163781	10774	10783	h	va
6	10508160	631051	692052	s	pa
7	10508160	618051	619051	s	pa
8	10508160	602050	603050	s	pa

Figure 30. Screenshot. Sample of RID tunnel segments table.

STATE DATA

The state data infrastructure tables are composed of data from the states of Virginia, Florida, Indiana, North Carolina, New York, Pennsylvania, and Washington. These states were selected for the Roadway Epochs Database based on the installation locations and home territory of participants in the 100-Car and SHRP 2 studies. Tables are organized by state to allow users a convenient means of assessing the available data from each DOT to better structure their research queries. Each state table begins with the prefix "epochs_state_" then the state name and a data label.

Florida

The Florida data set was collected in the year 2013 and contains two tables. Those two tables include measures of AADT (Figure 31) and traffic percentages (Figure 32).

	FILE_ID	timestamp_1	timestamp_2	study	location	aadt	addt_flag	aadt_flag_text
1	54605523	812559	825559	s	fl	24500	F	computed_from_factor_category
2	114140	49599	50039	h	va	43500	C	ttms
3	60633782	1198490	1209490	s	fl	38500	F	computed_from_factor_category
4	102303770	805994	807994	s	fl	123000	C	ttms
5	24783935	863806	868806	s	fl	5600	C	ttms
6	60633904	207710	234711	s	fl	2100	C	ttms
7	7742627	2288155	2296155	s	fl	18800	S	last_survey_and_prior_year_factor
8	32301361	135319	154319	s	fl	16500	S	last_survey_and_prior_year_factor

Figure 31. Screenshot. Sample of Florida AADT table.

The traffic percentages table for Florida contains the percentage of commercial truck traffic over a 24-hour period. The table contains just under 13,000 epochs of data, predominantly consisting of SHRP 2 files.

	FILE_ID	timestamp_1	timestamp_2	study	location	proportion_of_trucks_over_24_hours
1	81418628	1268037	1280037	s	fl	3.8
2	67542978	2044744	2054744	s	fl	11.1
3	75649613	795485	805486	s	fl	7.2
4	42769932	303035	374036	s	fl	7.2
5	1441055	189414	190414	s	fl	15.6
6	78253	24748	26268	h	va	8.8
7	19562811	607854	612854	s	fl	2.7
8	142293	42412	42492	h	va	2.8

Figure 32. Screenshot. Sample of Florida percent truck traffic table.

Virginia

Data provided by the Virginia DOT resulted in four tables with a total of 665,615 epochs. The majority of Virginia epochs are from the 100-Car Naturalistic Driving Study, with a handful from SHRP 2. Take care to note the "study" column for any data being requested. Definitions as provided by the Virginia DOT for quality indicators are provided in a column with a "txt" label. The text for each epoch shows the meaning of the character value in the data column (i.e., quality indicator "F" is factored short-term traffic count data). Definition columns may be identified by the suffix "collection_text."

Average Annual Weekday Traffic

Average annual weekday traffic (AAWDT) is the estimate of typical traffic between Monday and Thursday over one year. The average annual weekday traffic table (Figure 33) gives an integer value for the AAWDT for select epochs in the state of Virginia. The table includes the value of AAWDT, the quality indicator (AAWDT_QUAL column), and the text translation of the single-character code (aawdt_collection_text column).

	FILE_ID	timestamp_1	timestamp_2	study	location	AAWDT	AAWDT_QUAL	aawdt_collection_text
1	52459	182	203	h	va	17000	F	factored_short_term_traffic_count_data
2	192244	5628	5848	h	va	14000	F	factored_short_term_traffic_count_data
3	196833	2127	2147	h	va	65000	F	factored_short_term_traffic_count_data
4	170068	21723	21743	h	va	12000	F	factored_short_term_traffic_count_data
5	75432	8108	8209	h	va	58000	G	factored_short_term_traffic_count_data_w_growth_element
6	130152	5640	5810	h	va	34000	G	factored_short_term_traffic_count_data_w_growth_element
7	96195	39531	39681	h	va	35000	F	factored_short_term_traffic_count_data
8	205071	8527	8537	h	va	38000	F	factored_short_term_traffic_count_data

Figure 33. Screenshot. Sample of Virginia AAWDT table.

Traffic Counts

The traffic counts table (Figure 34) contains average daily traffic counts for epochs in Virginia, along with a quality column and the text translation of the single-character quality indicator.

	FILE_ID	timestamp_1	timestamp_2	study	location	ADT	quality_of_aadt	adt_collection_text
1	111925	7266	7686	h	va	4500	F	factored_short_term_traffic_count_data
2	161441	5759	5768	h	va	7400	G	factored_short_term_traffic_count_data_w_growth_element
3	136851	4823	4853	h	va	40000	N	aadt_of_similar_neighboring_traffic_link
4	60031	2575	2666	h	va	38000	F	factored_short_term_traffic_count_data
5	43128	1776	1996	h	va	13000	F	factored_short_term_traffic_count_data
6	177856	36876	36926	h	va	121000	A	average_of_complete_continuous_count_data
7	177986	4489	4750	h	va	19000	F	factored_short_term_traffic_count_data
8	146084	37809	37961	h	va	7000	G	factored_short_term_traffic_count_data_w_growth_element

Figure 34. Screenshot. Sample of Virginia AADT table.

Traffic Percentages

The traffic percentages table (Figure 35) provides the percentage of measured traffic that consists of bus, truck, and two- or four-wheel vehicles as a float value.

	FILE_ID	timestamp_1	timestamp_2	study	location	percent_two_and_four_tire	percent_bus_traffic	percent_truck_traffic
1	65077	5193	5433	h	va	99.24500000	0.20100000	0.25200000
2	157042	1225	1275	h	va	97.96000000	0.57200000	1.23900000
3	56197	28885	28895	h	va	88.67700000	0.78600000	0.86900000
4	44527645	3015914	3016914	s	nc	88.67700000	0.78600000	0.86900000
5	30467	11085	11215	h	va	95.02700000	0.70400000	0.86300000
6	7262	16174	16283	h	va	99.20700000	0.22500000	0.47200000
7	109329	64387	64507	h	va	99.20700000	0.22500000	0.47200000
8	96906	22651	22842	h	va	91.09100000	0.59200000	0.66100000

Figure 35. Screenshot. Sample of Virginia traffic percentages table.

K-factor

The K-factor table (Figure 36) contains K-factors for each traversed epoch, along with a K-factor quality indicator and the text translation of the single-character value. The text in the k_factor_quality_txt column indicates the measure the K-factor was based on, such as a neighboring link, the highest hour collected in a 24-hour period, or the 30th-highest hour observed over an extended period of time. Data values where no quality indicator is given leave the k_factor_quality_txt field null.

	FILE_ID	timestamp_1	timestamp_2	study	location	K_FACTOR	K_FACTOR_Q	k_factor_quality_txt
1	191359	6787	6897	h	va	0.10690000	F	factor_based_onHighest_hour_collected_in_a_48_hou...
2	80466	81984	82184	h	va	0.08860000	F	factor_based_onHighest_hour_collected_in_a_48_hou...
3	104347	73775	73916	h	va	0.13250000	A	factor_based_on_30th_highest_hour_observed_of_>2...
4	116543	35170	35339	h	va	0.09610000	F	factor_based_onHighest_hour_collected_in_a_48_hou...
5	58302	7270	7820	h	va	0.09910000	F	factor_based_onHighest_hour_collected_in_a_48_hou...
6	153064	24274	24824	h	va	0.09910000	F	factor_based_onHighest_hour_collected_in_a_48_hou...
7	2108	30938	31388	h	va	0.17770000	F	factor_based_onHighest_hour_collected_in_a_48_hou...
8	97076	2274	2774	h	va	0.17770000	F	factor_based_onHighest_hour_collected_in_a_48_hou...

Figure 36. Screenshot. Sample of Virginia K-factor table.

Pennsylvania

Six tables in the Roadway Epochs Database provide Pennsylvania DOT data. The majority of trips that provided input to the Pennsylvania tables are from the SHRP 2 collection with a few from 100-Car. Overall, there are 110,221 epochs represented in the Pennsylvania tables.

AADT

The AADT table (Figure 37) contains AADT values from the data collection year 2012. The AADT value is for all traffic types. Truck traffic AADT are provided in a separate table in the database.

	file_id	timestamp_1	timestamp_2	study	location	aadt
1	83290852	40950	41950	s	pa	11300
2	68133	44382	44402	h	va	11517
3	146647	10374	10584	h	va	5062
4	130115	53180	53549	h	va	34683
5	177898	2231	2301	h	va	18028
6	70680975	487462	493462	s	pa	15049
7	54341714	508219	512219	s	pa	6706
8	63510724	1498806	1500180	s	pa	17283

Figure 37. Screenshot. Sample of Pennsylvania AADT table.

Daily Truck VMT

The daily truck VMT table (Figure 38) provides an estimation of the average daily vehicle miles traveled (VMT) by commercial trucks on roadway segments that were traversed by study vehicles.

	file_id	timestamp_1	timestamp_2	study	location	daily_truck_vmt
1	103329	30317	30348	h	va	72
2	204038	34832	36413	h	va	20288
3	105363072	750064	758064	s	pa	52
4	21679	20578	20818	h	va	323
5	21658	8831	8941	h	va	149
6	324729	7195	7205	h	va	163
7	76579214	2136875	2141876	s	pa	55
8	156241	113352	113382	h	va	38175

Figure 38. Screenshot. Sample of Pennsylvania daily truck VMT table.

Daily VMT

The daily VMT table (Figure 39) for Pennsylvania in the Roadway Epochs Database contains values for the estimated daily total vehicle miles traveled for a roadway segment in the database as an integer value.

	file_id	timestamp_1	timestamp_2	study	location	daily_vmt
1	110736	30007	30017	h	va	88931
2	23005	8153	8163	h	va	101761
3	154256	26334	26424	h	va	14355
4	45031	10599	10619	h	va	8963
5	156247	57714	57964	h	va	1441
6	108290653	424668	425668	s	pa	4272
7	18817	4482	4543	h	va	395
8	19128	9756	9766	h	va	57822

Figure 39. Screenshot. Sample of Pennsylvania VMT table.

Percent Truck

The percent truck table for the state of Pennsylvania (Figure 40) represents the percentage of traffic consisting of commercial trucks on roadways traversed by study vehicles. Note that this value is given as an integer, which differs from the percent truck value in the Virginia tables, where the value is given in decimal form.

	file_id	timestamp_1	timestamp_2	study	location	TRK_PCT
1	110736	30007	30017	h	va	31
2	23005	8153	8163	h	va	11
3	154256	26334	26424	h	va	10
4	45031	10599	10619	h	va	5
5	156247	57714	57964	h	va	7
6	108290653	424668	425668	s	pa	2
7	18817	4482	4543	h	va	4
8	19128	9756	9766	h	va	31

Figure 40. Screenshot. Sample of Pennsylvania percent truck table.

Traffic Count

The traffic count table (Figure 41) provides observed daily traffic count information for roadways in the Roadway Epochs Database.

	file_id	timestamp_1	timestamp_2	study	location	weekday_traffic_count
1	62779298	1476480	1479480	s	pa	536
2	189146	1591	1630	h	va	609
3	58451	2614	2764	h	va	852
4	204903	8791	8991	h	va	160
5	70656	33350	33360	h	va	102
6	60525460	198749	208749	s	pa	49
7	63855424	466437	473437	s	pa	283
8	204049	22001	22091	h	va	1007

Figure 41. Screenshot. Sample of Pennsylvania weekday traffic count table.

Truck AADT

The truck AADT table (Figure 42) gives the average annual daily traffic counts for commercial trucks on roadway segments in Pennsylvania that are represented in the Roadway Epochs Database.

	file_id	timestamp_1	timestamp_2	study	location	truck_aadt
1	110736	30007	30017	h	va	9506
2	23005	8153	8163	h	va	2241
3	154256	26334	26424	h	va	906
4	45031	10599	10619	h	va	386
5	156247	57714	57964	h	va	604
6	108290653	424668	425668	s	pa	70
7	18817	4482	4543	h	va	64
8	19128	9756	9766	h	va	2528

Figure 42. Screenshot. Sample of Pennsylvania truck AADT table.

North Carolina

The database includes two tables with data from the North Carolina DOT. North Carolina data are from 2013 rather than 2012. The tables provided include an overall AADT value and a breakdown for commercial trucks.

Percent Traffic Type

The percent traffic type table (Figure 43) in the Roadway Epochs Database contains both the percentage of traffic and traffic count of commercial trucks in North Carolina on roadway segments in the Roadway Epochs Database. The Source field identifies the source of the data provided (as indicated by the North Carolina DOT).

	file_id	study	location	timestamp_1	timestamp_2	Source	single_unit_truck_percentage	multi_unit_truck_percentage	single_unit_truck	multi_unit_truck
1	32324363	s	nc	1216351	1231351	HPMS	0.02455	0.0148	1840	1110
2	149014	h	va	1094	1393	HPMS	0.07011	0.02039	880	260
3	21887	h	va	36691	36851	HPMS	0.03416	0.19626	690	3970
4	71772	h	va	8754	8794	NON HPMS	0	0	0	0
5	104977	h	va	19322	19472	HPMS	0	0	0	0
6	5208	h	va	27160	27540	HPMS	0.03069	0.11772	2300	8800
7	186001	h	va	11226	11276	HPMS	0.07011	0.02039	1020	300
8	5028	h	va	352	382	HPMS	0	0	0	0

Figure 43. Screenshot. Sample of North Carolina percent traffic type table.

Indiana

State of Indiana data originate from the Indiana DOT. Information contained in the state of Indiana tables predominantly comes from the SHRP 2 vehicles. The table (Figure 44) contains 3,752 rows. State data include the number of lanes represented on the links according to Indiana DOT information. Further information about roadways in Indiana and other states is provided through the federal and digital mapping data.

	file_id	study	location	timestamp_1	timestamp_2	lanes_on_state_link	SOURCE
1	1516445	s	in	2198044	2213044	4	INDOT_COVCNTCMPL
2	65961680	s	in	802139	809139	2	INDOT_COVCNTCMPL
3	6525982	s	in	154752	163752	2	INDOT_COVCNTCMPL
4	107266118	s	in	163591	174592	2	RID_2000
5	1495492	s	in	3094548	3119548	2	RID_2000
6	1516445	s	in	11070145	11132146	2	INDOT_COVCNTCMPL
7	82160338	s	in	7872557	7967558	4	INDOT_Interstate_maps
8	82605981	s	in	12869878	12908878	2	INDOT_COVCNTCMPL

Figure 44. Screenshot. Sample of Indiana number of lanes table.

PRECIPITATION TABLE

The precipitation table (Figure 45) lists SHRP 2 epochs that occurred during periods of precipitation as reported by the NOAA. Latitude and longitude at the beginning of each trip in the Roadway Epochs Database, along with the start time and date, were compared to NOAA historic data in order to find those trips that contained precipitation. Since the entirety of the trip was either said to contain precipitation or not, there are no timestamps in this table. Variables include the temperature in degrees Celsius, precipitation over the last hour in millimeters ("-1" indicates trace amounts), the reported wind speed in meters per second, and the distance to the reporting weather station in kilometers. For a complete review of the weather data methodology, see McCall, McLaughlin, Williams, and Buche (2014).

	file_id	study	location	temp_c	precip	wind_mps	distance_weather_station_km
1	38515689	s	nc	19.4	2	1.5	85.2628470923839
2	55213111	s	nc	21.1	1	3.6	14.1141009742533
3	97883215	s	nc	25.6	-1	0	13.2856734936848
4	11196269	s	nc	11	-1	2.6	25.5464692574797
5	18103650	s	nc	12.2	-1	4.1	22.1616755069426
6	51746479	s	nc	6.7	-1	2.6	25.6760785025502
7	12172505	s	nc	12.2	2	7.2	17.5640721482897
8	46007652	s	nc	17.8	-1	2.6	13.7498447139441

Figure 45. Screenshot. Sample of precipitation table.

CHAPTER 4. SQL POINTERS

This section first introduces basic components of a Structured Query Language (SQL) query and then provides a series of sample queries which increase in complexity as you read through the section. Sample queries are indented and presented in the Calibri typeface, with SQL keywords capitalized, thus:

```
SELECT file_id, study, timestamp_1, timestamp_2, IRI
```

Copying the sample query verbatim will provide the user with the desired outcome.

COMPONENTS OF A BASIC QUERY

The components of a basic query are listed in Table 2.

Table 2. Basic components of SQL queries.

Statement	Purpose
SELECT	Provides a list of columns that the user wants to select from the table of interest.
FROM	Identifies the table that the SELECT statement is using.
WHERE	Lists conditions or restrictions for the data to be included in the query results. An example would be “WHERE data < 5,” which would return results where the values in the column “data” are less than 5. It is similar to an IF in other languages.
ORDER BY	Tells the database how to list the desired results. ORDER BY takes a column name followed by either ASC for a list of ascending values (lowest to highest) or DESC for a list of descending values (highest to lowest).

OPERANDS

SQL allows for simple comparisons to place constraints on fetched results. Some of the more common operands are shown in Table 3.

Table 3. SQL operands.

Operand	Comparison
<	Returns data which are less than the value specified.
>	Returns data which are greater than the value specified.
=	Returns data equal to the value specified.
!=	Returns data **not** equal to the value specified.
BETWEEN	Returns data between two values.
OR	Inclusive or. Returns C if condition A or B is met.
AND	Exclusive and. Returns C if and only if A and B are met.
‘ ‘	Used to indicate a character in the comparison with one apostrophe on either side of the character (‘3+’).

SAMPLE QUERIES

Example #1 – Results Meeting a Set of Conditions

Problem: A researcher wants to find roadway segments with an IRI between *X* and *Y* to test the relationship between *z*-axis accelerations and Roughness Index scores. The accelerometer values would be requested from VTTI’s naturalistic data collections based on the file_id and timestamps located through the Roadway Epochs Database. In order to request sensor data from VTTI, four pieces of information are needed: the study the data were collected for, the file_id containing the epoch of interest, and the starting and ending timestamps for the epoch. To acquire those four pieces of information, a query will be launched in the Roadway Epochs Database for epochs matching our IRI requirements.

Query:

```
SELECT file_id, study, timestamp_1, timestamp_2, IRI
FROM epochs_federal_international_roughness_index
WHERE IRI BETWEEN 86 AND 187
```

Explanation: The SELECT statement indicates column names for the data pull. The FROM statement indicates the table of interest. WHERE statements indicate conditions that must be met in order for the data to be returned. WHERE statements can contain logical conditions such as greater than, less than, or between for any two values. Data that match those criteria (where logical comparisons equal true) will be returned.

Example #2 – Using JOIN and AS

The next query involves joining two tables into one results set. To accomplish this, three things should be added to the structure.

1. A JOIN clause identifies the table to be fetched with the original results set. We are joining on both file_id and a timestamp to guarantee that the returned results set falls within the time periods of interest for each file_id.

2. An AS statement allows us to refer to tables by using an alias. In this case, the original is referred to as “iri_table” and the table we are joining to is referred to as “lanes.”

3. A table name appended to variables appearing in both tables allows us to specify the table and variable of interest. In the example below, this is accomplished by fetching “lanes.file_id” rather than simply “file_id.” This is done to indicate that we are interested in the file IDs present in “iri_table.”

Problem: A researcher wants to perform a similar analysis to the one above, but in this case the number of lanes is also of interest. In order to combine the two tables, a JOIN function will be required. See below for an example of how to use the above query with a join. The query below may be amended to join any two or more tables.

Query:

```
SELECT lanes.file_id, lanes.study, lanes.timestamp_1, lanes.timestamp_2, iri,
lane_grouping_per_direction
FROM epochs_federal_international_roughness_index AS iri_table
JOIN epochs_dmapping_lane_categories AS lanes
ON lanes.file_id = iri_table.file_id
AND lanes.timestamp_1 = iri_table.timestamp_1
WHERE IRI BETWEEN 86 AND 187
AND lane_grouping_per_direction = '3+'
ORDER BY file_id, timestamp_1
```

Explanation: The SELECT statement has been amended to indicate which table we wish to use for each of the variables in the select. This only has to be done when the variable is repeated in both tables being joined. The second difference between this query and the one preceding it is the addition of the term AS. AS simply allows us to save typing time by giving the table an alias. Third, we tell the query with what table we wish to join our original table via the JOIN command, which functions as a second FROM. The JOIN serves as the second table with data present in the SELECT. In the example above, the JOIN provides the query with lane_grouping_per_direction. We also limit our query to only those results which match with values on the joined table “lanes” by indicating that we only want values of file_id present on lanes. The ON statement tells the query what conditions need to be met prior to the joining. In this case, the timestamp_1 and file_id values must be identical. The WHERE statement functions the same as in Example #1. ORDER BY informs the database to order the results in this case first by file ID then by timestamp.

Example #3 – Questions of Scale and Counts

Problem: A researcher needs to know how many epochs occurred on roadway segments with particular IRIs. That researcher wants to construct a bar graph of the counts for each IRI value in the table. In order to construct the bar graph, the researcher needs to know what those counts look like. The following query will produce each value of IRI with a tally of how many times it is represented. For instance, an IRI of 187 is represented 1,332 times in the database.

Query:

```
SELECT IRI, count (iri) AS nIRI
FROM  epochs_federal_international_roughness_index
WHERE IRI BETWEEN 86 AND 187
GROUP BY IRI
ORDER BY IRI DESC
```

Explanation: The SELECT statement works the same as in the previous examples. The inclusion of the COUNT () command informs the database that you wish to have a count (frequency) returned rather than the data itself. In this example, the AS statement informs the database that the resulting count will have the column name nIRI. The WHERE and BETWEEN statements are unchanged. GROUP BY indicates that the counts being returned should be grouped according to the unique value of the IRI. Another example would be to use file_id in the GROUP BY rather than IRI, which would return counts according to file ID.

CHAPTER 5. CONCLUSIONS

The Roadway Epochs Database has achieved the goals proposed. It has increased accessibility to VTTI's naturalistic data collections by adding gateways to the sensor data for infrastructure-related research questions. The SQL tutorial and examples included in this report provide users who are unfamiliar with database organization a few basic techniques to query the data. Epochs of naturalistic sensor data can now be queried by speed limit, pavement condition, the number of lanes, or traffic counts. The entirety of the 100-Car data set has been processed using the VTTI linking algorithm. Moreover, the number of naturalistic collections in the Roadway Epochs Database is easily expandable. This is particularly true for SHRP 2 NDS data, where only 1,000 trips from each of the six installation locations (total 6,000) were utilized.

The tables included in the Roadway Epochs database include traversals of roadway segments as classified by infrastructure data sources regardless of naturalistic data collection. That is, even when data come from the SHRP 2 RID, if a 100-Car participant drove across that segment of roadway, the appropriate epoch's RID table will contain the relevant roadway information. The same is true for digital mapping data, FHWA HPMS sourced data, and data from the individual state DOTs. To handle situations such as these, each table includes information regarding which of the two naturalistic collections contains the recorded vehicle sensor values.

Future work could include the inclusion of more data sets, both naturalistic and infrastructure-related. Naturalistic data sets, such as those including commercial motor vehicle drivers, older drivers, and younger drivers, may give Roadway Epoch Database users more demographic-specific knowledge regarding infrastructure variables. Those collections would be particularly valuable in the context of added tables that indicate construction zones or the dates, times, and contents of variable message signs. Increasing the number of state DOT sources would increase the value of the data set to those pursuing research on regions other than Virginia and the SHRP 2 locations.

REFERENCES

Dingus, D. A., Klauer, S. G., Neale, V. L, Peterson, A., Lee, S. E., Sudweeks, J., Perez, M. A., Hankey, J., Ramsey, D. A., Gupta, S., Buche, C., Doerzaph, Z. R, Jermeland, J., & Knipling, R. R. (2006). *The 100-Car naturalistic driving study, Phase II - Results of the 100-car field experiment* (DOT HS 810 593). Washington, DC: National Highway Traffic Safety Administration.

McCall, R., McLaughlin, S., Williams, S., & Buche., T. (2014, September). An analysis of temperature and precipitation data in a naturalistic motorcycle study. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 58,* 2117-2121. doi:10.1177/1541931214581445

McLaughlin, S., & Hankey, J. (2015). *Naturalistic Driving Study: Linking the study data to the Roadway Information Database* (Report No. S2-S31-RW-3). Wshington, DC: Transportation Research Board.

Perez, M. (2014, September 10). *Second Strategic Highway Research Program (SHRP 2) safety databases training: Course 1 overview of the SHRP 2 naturalistic driving data*. Presentation. Virginia Tech Transportation Institute, Virginia Tech Research Center, Arlington, VA.

Neale, V., Dingus, T. A., Klauer, S. G., Sudweeks, J., & Goodman, M. (2005). *An overview of the 100-car Naturalistic Study and findings* (Paper No. 05-0400). Washington, DC: National Highway Traffic Safety Administration.