Big Data Hit the Road

The First Year of Use of the SHRP 2 Safety Databases

JOHN NJORD AND KIRK STEUDLE

Njord is Executive Vice President, Tom Warne and Associates, Murray, Utah, and Chair of TRB's Safety Data Oversight Committee. He is the former Director of the Utah Department of Transportation. Steudle is Director, Michigan Department of Transportation, Lansing, a member of the Safety Data Oversight Committee, and past chair of the Oversight Committee for the second Strategic Highway Research Program (SHRP 2). Both Njord and Steudle have served as Chair of the TRB Executive Committee.

The second Strategic Highway Research Program (SHRP 2) conducted a naturalistic driving study (NDS) that was unprecedented in size and scope. The study collected data from more than 3,500 volunteer passenger-vehicle drivers, ages 16 to 98, during a three-year period, with most drivers participating for one to two years.

The study was conducted at sites in six states: Florida, Indiana, New York, North Carolina, Pennsylvania, and Washington. The two predominantly rural sites, in Indiana and Pennsylvania, covered about 10 counties each; the other four urban or mixed sites covered one to three counties each. The total study area encompassed more than 21,000 square miles.

Data collected included vehicle speed, acceleration, and braking; vehicle controls, when available; lane position; forward radar; and video views forward, to the rear, and on the driver's face and hands. The NDS data file contains approximately 35 million vehicle miles, 5.4 million trips, 2,705 near-crashes, 1,541 crashes, and more than 1 million hours of video. All together, these amount to 2 petabytes of data—"big data" by any definition.

The companion Roadway Information Database (RID) contains detailed roadway data collected on 12,538 centerline miles of highways in and around the study sites—approximately 200,000 highway miles of data from the highway inventories of the six study states, and additional data on crash histories, traffic and weather conditions, work zones, and ongoing safety campaigns in the study sites. The NDS and RID data can be linked, so that driving behavior and outcomes can be associated with the roadway environment.

PHOTO: VIRGINIA TECH TRANSPORTATION INSTITUTE

(*Above*:) A composite image assembles each of the four camera views of a participant in the Naturalistic Driving Study. Images from 5.4 million trips were part of the 2 petabytes of driving data gathered during the three-year study.

(*Below*:) Images preceding a crash. Video views of a driver's face and hands allow researchers to examine driver behavior; other views present the vehicle and roadway environment.

PHOTO: VTTI

Unparalleled Data

The central goal of the NDS was to produce unparalleled data from which to study the role of driver performance and behavior in traffic safety and the effects of the interaction between drivers and the roadway environment on the risk of crashes. Driver error is a contributing factor in more than 90 percent of all crashes.

Understanding the human side of driving is critical for making large-scale improvements in traffic safety. Improvements require an understanding of how the driver interacts with and adapts to the vehicle, traffic conditions, roadway characteristics, traffic control devices, and other environmental features. After-the-fact crash investigations can estimate these interactions only indirectly.

The NDS data record how drivers actually drive, what they are doing just before they crash or almost crash, and how they successfully avoid incidents the vast majority of the time. The NDS and RID data will serve for years in developing and evaluating safety countermeasures to prevent or reduce the severity of traffic crashes and injuries.

Assembling the Database

Collecting and assembling the data—a massive

Photo: VTTI

An installer equips an NDS car in Buffalo, New York, with instrumentation for recording acceleration, braking, speed, and other data.

undertaking—involved hundreds of people. In total, the SHRP 2 NDS collected 6,559,367 files.

A trip file usually encompassed a trip from 30 seconds after the ignition was turned on until the ignition was turned off. The data ingestion and database assembly processes aimed to preserve as much of the usable data as possible, even if a trip was short or if some of the data were missing.

In the end, approximately 1 percent of the trip files had to be excluded from the final database, mainly because video was missing or unusable. According to experts, the proportion of trip files lost was exceptionally low, given the extensive size of the project, the newly designed data acquisition system, and the lack of experience with large-scale naturalistic driving data collection at the six sites.

Only the participants who signed an informed consent agreement could be considered in the study, which required a manual review of each trip file to exclude data from drivers who had not consented. Data reductionists reviewed nearly 99 percent—or 6,483,997—of the trip files. The review assigned the correct participant identification number to each trip file, facilitating access by researchers to the drivers of interest for specific studies.

In total, the SHRP 2 NDS database included approximately 85 percent of the collected and usable trip files (see Table 1, page 5). The largest category of excluded trip files consisted of drivers who had not consented—these accounted for approximately 10.5 percent of the manually reviewed trip files.

Photo: VTTI

At driver assessment workstations, study participants answered demographic, health, and driving questionnaires and underwent vision tests and other assessments.

Structure and Guidance

While the SHRP 2 research program was under way and the data were being collected and assembled, an advisory committee explored the long-term stewardship of the safety data. The Long-Term Stewardship Committee—assembled in accordance with the Federal Advisory Committee Act—recommended a multiphase approach. During the first five-year period, or Phase 1, the safety data would not be moved, to build a more complete base of knowledge about the data demand and use and on the resources required to guide the remainder of implementation. Phase 1 is an experimental operational period for the safety data, during which a variety of research projects will be under way using the database.

A memorandum of understanding, signed on June 26, 2014, governs activities in Phase 1 by four partners: the Federal Highway Administration (FHWA), the American Association of State Highway and Transportation Officials (AASHTO), the National Highway Traffic Safety Administration (NHTSA), and the Transportation Research Board (TRB).

A cooperative agreement between FHWA and TRB provides the funding for Phase 1 activities by TRB staff, committees, and contractors. The objectives of the cooperative agreement are to promote conditions for making the SHRP 2 safety data available to qualified users and to gain experience and data to support decisions about the implementation and oversight of the data after Phase 1.

A Safety Data Oversight Committee (SDOC) provides policy guidance for Phase 1. Members include executives from state transportation agencies, representatives from the automotive industry, academics with expertise in big data and information technology, and traffic safety researchers; representatives from the four partners serve as ex officio members.

Expert Task Groups

Several expert task groups (ETGs) provide technical advice to the SDOC. An ETG established in mid-2014 has advised on statements of work for database management contracts, on a request for information released in January 2015, and on data sharing and privacy protection policies. This ETG included technical experts on information technology, databases, human factors research, transportation safety, statistics, and big data; the ETG disbanded with the conclusion of SHRP 2.

TRB has established two additional ETGs since March 2015—and may establish others—to address such issues as privacy protection, user community

development and outreach, information technology considerations, and sustainable business models for database operations after Phase 1. The SDOC and the ETGs include representatives from the health care industry, which has considerable expertise in analyzing large, complex data sets that involve privacy protection.

Phase 1 Issues

Phase 1 will address many issues associated with the SHRP 2 safety data. The cooperative agreement outlines some of the most important issues:

- Data usage, research interests, and the potential market for the data;
- Institutional structures and responsibilities;
- Costs, users' willingness to pay, funding sources, cost sharing, and user fee structures;
- Types of facilities, skills, management processes, and technologies for user access and for the protection of personally identifying information in the data;
- Performance measures for data marketing, delivery, dissemination, and access;
- Protection of personally identifiable information;
- The effectiveness of user tools and support;
- The types of information to be generated by the selected testing of options within the constraints of the Phase 1 program; and
- Options for implementation and oversight of the SHRP 2 Safety Data Program after Phase 1, including the pros and cons for each option.

Accessing Data and Metadata

The SHRP 2 NDS collected approximately 2 petabytes of data, which can be categorized as shown in Table 2 (page 6). Extensive work was needed to make the sheer volume of data more accessible and usable for researchers; the face of the usability efforts is the InSight website.[1]

InSight facilitates use of the vast data set by the transportation research community and other researchers. The website answers some research questions directly and provides the information necessary for planning ways to answer other questions that require more in-depth exploration of the SHRP 2 NDS. The website also includes thorough dictionaries of data and variables—for example, the SHRP 2 *Researcher Dictionary for Video Reduction Data*—to assist with interpretation.

Special care was taken to exclude any personally identifying information on the website, to minimize restrictions on access. Although the initial InSight website is complete and in operation, enhancements are expected. The website's query page is designed to assist researchers in interacting with the data on vehicles, drivers, trips, and events.

[1] https://insight.shrp2nds.us/.

TABLE 1 Number of SHRP 2 Trip Files in Each Driver Category

Driver Category	Number of Trip Files	Percent of Total Files
Consenting driver	5,512,900	85.02
Unknown (likely not consenting)	684,733	10.56
Trip before consent	39,936	0.62
No driver (e.g., car warming up in a driveway)	221,051	3.41
Data collection site technician	12,829	0.20
Multiple drivers	12,548	0.19
Total	6,483,997	100.00

Exploring InSight

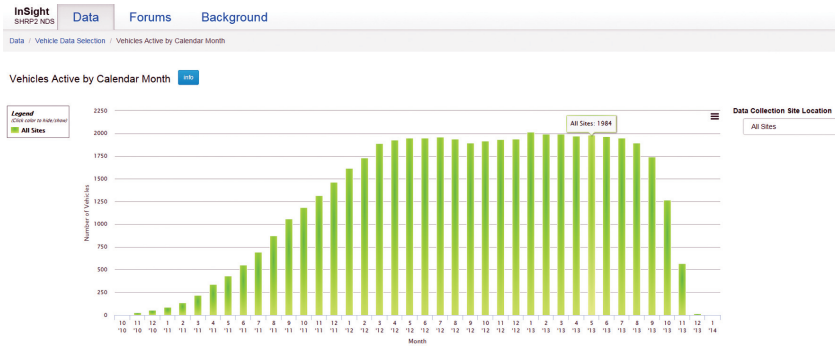
Expectations are that researchers will use InSight most frequently for work with data from the following sources:

- Detailed participant assessments, including demographic questionnaires, health and driving questionnaires, and vision tests, completed by more than 3,100 drivers.[2]
- Vehicle information—for example, safety and entertainment options—for all 3,358 vehicles.
- Summary variables for more than 5.5 million trip files—for example, the maximum speed reached, the maximum deceleration achieved, and the duration of each trip).
- Interactive heat maps detailing the roads driven—and the number of times—by drivers in the study at each of the six data collection sites. A SHRP 2 research report describes in detail how the heat maps were developed and how researchers can link the NDS and RID portions of the safety database (*1*).

[2] Detailed information and documentation on participating driver characteristics and assessments are available via InSight at https://insight.shrp2nds.us/data/category/drivers#/list.

PHOTO: VTTI

Researchers can take online training and earn certification in the use of the databases.



The vast NDS data set is accessible to researchers on the InSight website.

◆ Event data from the crashes, near-crashes, and baselines that were identified and selected. The Virginia Tech Transportation Institute has prepared a report on the crashes, near-crashes, and baselines in the NDS database (2).

Users can establish an account for basic access to InSight. To gain additional functionality and access to features, users can take online training, including the ethics of human subjects research; after passing a quiz and providing an electronic certificate, a user becomes a qualified researcher. The training typically takes one hour.

TABLE 2 Categories of Data Collected in the SHRP 2 Project

Participant Assessments	Demographic questionnaire Driving history Driving knowledge Medical conditions and medications Screening for attention deficit–hyperactivity disorder Risk perception Frequency of risky behavior Sensation-seeking behavior Sleep habits Results of visual, physical, and cognitive tests Exit interview
Vehicle Information	Make, model year, and body style Vehicle condition (tires, battery, etc.) Safety and entertainment systems
Continuous Data	Face, forward, rear, and instrument panel video Vehicle network data Accelerometers: gyroscopes, forward radar, GPS Additional sensor data
Trip Summary Data	Characterization of trip contents Start time and duration of trip Minimum, maximum, and mean sensor data Time and distance driven at various speeds, headways Vehicle systems usage
Event Data	Crashes, near-crashes, baselines 30-second events with classifications Postcrash interviews Other crash data
Cell Phone Records	Subset of participant drivers Call time and duration Call type (text, call, sending or receiving photos or videos, etc.)
Roadway Data	Matching trip GPS to roadway database Roadway classifications Other roadway data

InSight Users

Approximately 1,000 researchers have registered to use the InSight web portal. The majority of registered users—approximately 80 percent, as of May 2015—are from U.S. Internet domains, perhaps reflecting the location of the six collection sites. Nevertheless, the international contingent of InSight users is considerable, with approximately 10 percent of all users from European countries. The remaining 10 percent are largely from China, Canada, Japan, and Australia.

More than half of the InSight users come from academia—university faculty, staff, and graduate and undergraduate students. Another 20 percent are staff from federal, state, or local governments. Other sizable groups of InSight users are from Internet domains associated with nonprofit organizations, motor vehicle and subsystem manufacturers, and automobile insurance companies.

Going InDepth

A broader array of data is available beyond the InSight website. Through InDepth, qualified researchers with a standard data use license (DUL) can access a subset of SHRP 2 safety data to meet the needs of a research problem statement.

In most cases, once the DUL is in place, a research database is assembled and made available to the qualified researcher. The DUL specifies the purpose for which the data are to be used and the period of time for use. If the research involves personally identifying information, such as in-vehicle video, the DUL will specify the requirements for use in a secure data enclave that prevents copying.

Before the completion of the database, SHRP 2 undertook three pilot projects using the safety data. The undertaking was inherently difficult, akin to flying an airplane that is still being built; the research results proved valuable for several critical safety topics (see the sidebar, page 7).

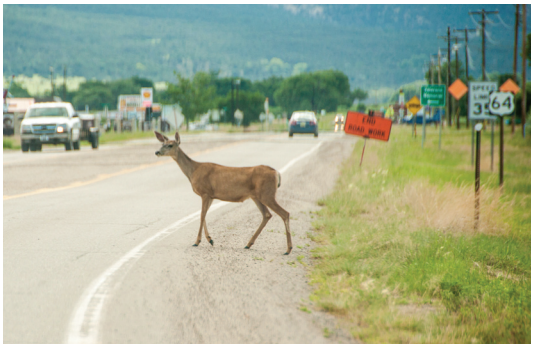
InDepth Users and Topics

Several organizations have completed research projects, have projects under way, or are arranging to use the SHRP 2 safety data via InDepth. Customers include organizations in North America and in Europe, from universities, private consulting firms, automotive original equipment manufacturers, state departments of transportation (DOTs), national laboratories, federal agencies such as FHWA and NHTSA,

public health organizations, nonprofit research institutions, and the automobile insurance industry.

FHWA and AASHTO are sponsoring almost a dozen implementation assistance projects, known as Concept to Countermeasure, applying the SHRP 2 safety data. These projects involve partnerships between state DOTs and research organizations such as universities (see the sidebar on page 8).

The diversity of the InDepth user base is encouraging, and the diversity of the research topics is impressive. Past, current, and pending research topics include driver behavior and safety on curves; offset left-turn lanes; lane departure warning systems; driver distraction and inattention; rural intersections; vehicle safety defects; speeding; animal–vehicle collisions; road rage; driver fatigue; crash risk by gender and age; seatbelt use; crash risk and driver health conditions; markings at pedestrian crossings; driver impairment risk and personality; autonomous vehicle safety systems; speed limits, roadway geometry, and driver behavior; closely spaced freeway interchange ramps; roadway departure; work zones; inclement weather, driver behavior, and traffic safety; and fuel economy and vehicle operating costs.



Data from the NDS are facilitating studies on animal–vehicle collisions, crash risk, seatbelt use, and more.

The last topic indicates the potential for using the database for research in nonsafety areas such as traffic operations, transportation planning, energy conservation, and environmental protection.

Early Uses of the Safety Data

Practical Findings from Pilot Projects

Even as the safety databases were being assembled, SHRP 2 undertook a series of research projects to verify the value of the naturalistic driving data and the roadway information data. The SHRP 2 Project S08 series analyzed the early NDS data to address high-priority topics, including safety on rural two-lane curves, driver inattention, and offset left-turn lanes.

In Part 1 of the S08 projects, four research teams worked on proofs of concept; three of the four teams moved forward to conduct full analyses in Part 2. Results from these projects could be used to design or refine cost-effective measures to reduce roadway departure crashes, warn inattentive drivers, and help state departments of transportation to design intersections that balance crash risk with construction and maintenance costs. The experience in analyzing then-incomplete SHRP 2 NDS and RID data helped establish efficient methods for identifying, extracting, and analyzing data that now benefit all users.

PHOTO: CHRIS FORD, FLICKR

Early projects mined NDS safety data to study horizontal curves on rural two-lane highways.

The three projects approved for full analyses in early 2013 were completed in 2014:

- Project S08A, led by SAFER at Chalmers University in Sweden, focused on the interaction between driver inattention—including distraction—and crash risk. This research is continuing under another funding source and has indicated that even short periods of inattention could be hazardous under certain driving conditions—for instance, in heavy traffic on multilane roadways.
- Project S08B, led by MRI Global of Kansas City, Missouri, considered the safety effect of offset left-turn bays, including positive, neutral, and negative offsets. The results indicate that positive offset left-turn lanes—which allow drivers a more unobstructed view of oncoming traffic—may have significant benefits for safety and for traffic flow.
- Project S08D, led by Iowa State University in Ames, focused on roadway departures on horizontal curves of rural two-lane highways. The research examined the ways that drivers negotiated the curves, and the findings indicated that correctly placed advance warning devices—such as raised pavement markings and chevrons—provide valuable advance information and warnings.

Additional information on the three S08 pilot projects, including the published research reports, is available at www.trb.org/Publications/PubsSHRP2ResearchReportsSafety.aspx.

Leading to Breakthroughs

Following are examples of the safety research under way or planned:

- A university in the Midwest is using time-series and GPS data, as well as forward video from crashes and near-crashes in the NDS database, to help a neighboring state DOT explore how crashes happen and are avoided in highway work zones.
- A major motor vehicle manufacturing company is developing a comprehensive database to examine driver distraction. Part of this project involves developing an algorithm to identify episodes of distracted driving from the SHRP 2 data. The company has indicated interest in making the resulting database available to other researchers.
- A nonprofit research organization is assisting a federal traffic safety regulatory agency with two major research projects examining speeding and the nonuse of safety belts. Each project is tapping into a variety of continuous data sources.

Efforts such as these should expand the literature on highway safety and lead to breakthroughs in making highways safer through an improved understanding of driver behavior. The first findings from the SHRP 2 safety data are beginning to appear in research journals.

References

1. McGlaughlin, S. B., and J. M. Hankey. *Naturalistic Driving Study: Linking the Study Data to the Roadway Information Database*. SHRP 2 Report SD-S31-RW-3, Transportation Research Board of the National Academies, Washington, D.C., 2015.
2. Hankey, J. M., M. A. Perez, and J. A. McClafferty. *Description of the SHRP 2 Naturalistic Database and the Crash, Near-Crash, and Baseline Data Sets*. Virginia Tech Transportation Institute, Blacksburg, 2015.

Concept to Countermeasure

FHWA and AASHTO Spearhead Use of SHRP 2 Safety Data

In August 2014, FHWA and AASHTO selected 10 states to participate in a proof-of-concept effort under the SHRP 2 Implementation Assistance Program (IAP), Concept to Countermeasure: Research to Deployment Using the SHRP 2 Safety Databases. The program has designated approximately $3 million in financial and technical assistance for research on 11 topics.

PHOTO: ANITA GOULD, FLICKR

Traffic slows to a crawl in a snowstorm near Toms River, Maine. Winter weather conditions and other research topics received early implementation assistance from FHWA and AASHTO.

IAP grant recipients are using the SHRP 2 safety data to conduct research on their topics and will pilot and promote any promising countermeasures identified. A long-term goal is the development of new countermeasures for national adoption. In partnership with researchers, state agencies are managing the research, will implement the findings, and will deliver the authorized results.

The effort is proceeding in three phases, to simplify the process and to reduce the risk and uncertainty for the participants. In the first, 9-month phase, participants used a reduced set of NDS and RID data to demonstrate that the research concept was viable and that an analysis with a larger data set would answer the question more definitively.

At the end of Phase 1, FHWA and the AASHTO Safety Task Force undertook a review of the work to determine whether the results are promising enough for the research to continue to Phase 2. The agencies selected for Phase 2 will have access to the full SHRP 2 safety data set and will negotiate a work plan, budget, and schedule.

If Phase 2 produces meaningful results likely to lead to an implementable countermeasure or a new behavioral strategy, FHWA would provide additional financial or technical support for Phase 3, which would address implementation. The implementation would not involve additional research but would include engineering or other support to update national manuals or policies or to develop strategies for incorporating the countermeasure and endorsing it for national adoption. Phase 3 also may include pilot-testing the safety countermeasure in the field, implementing public outreach, or other measures to improve highway safety.

The 30 applications from states for the IAP funds exceeded expectations; each of the 10 state DOTs selected received approximately $100,000 for each proposal. Research using the two safety databases started up in January 2015, and reports on findings were submitted in September from Florida, Iowa, Michigan, Minnesota, Nevada, New York, North Carolina, Utah, Washington, and Wyoming. Washington State DOT received two awards for separate research topics.

The topics researched include pedestrian–vehicle interactions, roadway departures, speeding, work zones, horizontal and vertical roadway curves, interchange ramps, adverse weather conditions, and roadway lighting.

Additional information about Concept to Countermeasure is available at www.fhwa.dot.gov/goshrp2/Solutions/Safety Topic/NDS/Concept_to_Countermeasure__Research_to_Deployment_Using_the_SHRP2_Safety_Data.