SHRP 2 Safety Data, Update 6

This is the sixth in a series of updates about Phase 1 of the implementation of the SHRP 2 Safety Data by TRB in collaboration with its partners, Federal Highway Administration (FHWA), the American Association of State Highway and Transportation Officials (AASHTO), and the National Highway Traffic Safety Administration (NHTSA). As was noted in the first of these updates, "Phase 1" is governed by a Safety Data Oversight Committee (SDOC). "Phase 1" is an experimental program designed to make the safety data as widely available as possible to qualified researchers and to gather information needed to lay the foundation for use of the data beyond the first four to five years.

This edition of the update focuses on some things that users and potential users of the data need to know about human subjects data in general and the Naturalistic Driving Study data as a specific example of human subjects data.

What Protections Were Promised to Participants in the SHRP2 Naturalistic Data Study?

Unlike many databases typically used in transportation safety research, the SHRP 2 safety databases were developed using human research subjects. Studying human subjects—as opposed to using more traditional sources such as police report-based crash records—provides a much better opportunity to understand how drivers interact with their vehicles, other motorists, roadways, and environmental factors including weather.

As human subjects, all 3000-plus SHRP2 Naturalistic Driving Study participants were promised confidentiality in exchange for participation in the study and recording of their driving data, including various video streams inside and outside the vehicle. The operators of the NDS database have an ethical obligation to take that promise of participant confidentiality seriously. To protect the confidentiality of participants, data that contains any personally identifying information (PII) is not made available outside of a secure data enclave (SDE). Qualified researchers who wish to view and analyze PII must meet eligibility criteria and agree to the requirements of a data use license (DUL).

What Is PII?

The SHRP 2 NDS data include some items that could be used to identify the human subjects who participated in the NDS. The most commonly used PII data elements include driver face video; full trip GPS traces which can be used to identify a person's home, work, and school locations; and unaltered forward video of a crash. There are other less commonly used yet potentially identifying data elements. PII data elements must be viewed in a secure data enclave to protect the privacy of study participants.

What is a Secure Data Enclave?

A secure data enclave (SDE) is a physically isolated environment (which can co-located with the data or remote) in which carefully controlled access to restricted data can be provided to qualified researchers. A physical SDE is a room or set of rooms which meets a set of security rules so that restricted data cannot be taken out of the room or copied. A remote (or virtual) SDE accomplishes this same level of security via electronic and physical means. There is currently one physical Secure Data Enclave available in which PII from the Naturalistic Driving Study can be used. That SDE is at the Virginia Tech Transportation Institute (VTTI) in Blacksburg, Virginia. However, the Transportation Research Board and VTTI are now working with the Turner-Fairbank Highway Research Laboratory outside Washington, DC to develop and pilot test a remote secure data enclave. That remote SDE, if successful, could be the model for other remote secure enclaves in the future, subject to availability of funding.

Use of any secure data enclave is subject to certain restrictions and scheduling constraints.   The cost of secure data enclave use may be recovered from data licensees.

What is a Certificate of Confidentiality?

Research subjects who participated in the SHRP2 NDS are protected by a Certificate of Confidentiality issued by the Department of Health and Human Services (HHS) in accordance with the provisions of section 301(d) of the Public Health Service Act (42 U.S. C., 241(d)). All data licensees are considered to be a cooperating agency under the terms of the Confidentiality Certificate; as such, all Data Licensees, Principal Investigators, and Research Staff are required to protect the privacy of the individuals who are the subjects of SHRP2 NDS by withholding their identifying characteristics from all persons who are not signatories to a data use license. Identifying characteristics" are considered to include those data defined as PII under the terms of the data use license.

What is an IRB?

An institutional review board (IRB), also known as an independent ethics committee (IEC), ethical review board (ERB) or research ethics board (REB), is a committee that has been formally designated to approve, monitor, and review research involving humans. In the case of the NDS, VTTI manages the review and approval process for new and amended data use licenses under authority granted by the Virginia Tech IRB and the Transportation Research Board. The National Academies of Sciences (NAS) IRB is also involved; this IRB does not review individual data use licenses, but does periodically review the process used to approve licenses to make sure that human subjects are protected. Finally, each data licensee needs to gain approval from the IRB (or equivalent in the case of non-US licensees) that governs human subject research for their institution.

What Are Original Data and De-Identified Summary Data?

The SHRP 2 NDS data include both original data (collected directly from the participant at enrollment, about the vehicle at installation, from each trip as the driver conducted normal daily driving, and about crashes via post-crash interviews) and de-identified summary data. The research protocol that was negotiated between the Transportation Research Board and the Institutional Review Boards states that

any original data must be destroyed by any licensing agency at 30-40 years after data collection ended (depending on the type of data) and thus these data must be tracked. (Data use licenses being issued now typically call for the return or destruction of original data at two years past project completion, which is usually long enough to complete a research project and generate publications.) This sort of limitation on the life of original data is very common in situations where human subjects data are being used for research, for instance in medicine.

De-identified data have been processed to prevent participants' identities from being disclosed. Summary data have been transformed in a significant way, for instance being presented as an average or standard deviation instead of original data. De-identified summary data from the NDS may be kept indefinitely. Most projects will request a combination of original and de-identified summary data. For these reasons, projects have a default licensee data retention length of two years past project completion. De-identified summary data (which cannot be reverted to original data by transformation) produced by the licensee as part of their work processes on this DUL may be retained and used by the licensee without licensing agency or data steward constraints.

Do Any Similar Restrictions Apply to the Roadway Information Database?

The Roadway Information Database (RID) that is available to complement the NDS contains no personally-identifying information. No research protocol was approved by an IRB for the RID since one was not required by Federal laws. Therefore the RID data can used with far fewer limitations than the NDS data and they do not have to be destroyed after 30 or 40 years. A simple terms of use agreement is required, mainly to help measure the use of the RID and to help prevent version control issues that might accompany unlimited copying of the data.