SHRP2

STRATEGIC HIGHWAY RESEARCH PROGRAM

Accelerating solutions for highway safety, renewal, reliability, and capacity

SHRP2 Data Access

VTTI Second International Symposium on Naturalistic Driving Research

Joanne Harbluk - Data Access ETG

31 August 2010

TRANSPORTATION RESEARCH BOARD
OF THE NATIONAL ACADEMIES

Overview: Data Access Issues

1. Review of data to be collected:
    - source, characteristics
2. Need for participant protection:
    - Ethics, IRB, privacy
3. Data Security: requirements; approaches
4. Data Access: requirements; process
5. Questions and comments

SHRP2 Data Access

The Goal:

Maximize availability & use of SHRP2 NDS data

- Protect the rights and privacy of study volunteers: Ethics, IRB, privacy concerns
- Data Security
- Data access for qualified researchers

1. Data Summary

Sources of data for research purposes:

1. **Driver assessment tests/questionnaires** for primary drivers
2. **Instrumented vehicles:** continuous video; vehicle sensor/parametric data; GPS data
3. **Crash investigations** for selected crashes
4. **Roadway/roadside characteristics** from mobile data collection vans and external sources

800 TB of video + 100 TB of vehicle sensor data + roadway data = ~1 petabyte of data

Data

6 Data Types:

1. Contact information
2. Auxiliary study information
3. Driver questionnaire/test data
4. Vehicle data
5. Driving data
6. Additional crash data

Data

6 Data Types:

1. Contact information 2. Auxiliary study information	**Securely stored separately**
3. Driver questionnaire/test data 4. Vehicle data 5. Driving data 6. Additional crash data	**Securely stored in SNDS Data base**

2. Participant Protection

- **Human Subjects Research:**
  - **DHHS policy for the Protection of Human Subjects (45 CFR 46)**
- **Institutional Review Boards (Multiple):**
  - 6 site contractor IRBs
  - Virginia Tech IRB
  - NAS IRB
- **NIH Certificate of Confidentiality**
- **Elements of consent process:**
  - Participant Consent Forms
  - Web site
  - Video presentation

Sensitive Data that Require Protection

Personally Identifying Data & Information

- Contact Information
- Auxiliary study information

Certain Driving Data Elements

- Driver Face Video
- Complete GPS records of trips
- Audio Recordings
- Crash Data & Video; fatal/severe; driver, struck ped/cyclist

Post-Crash Investigations

* Combinations of data that by themselves would not reveal identity

3. Data Security

When Data Might be "Seen"

Vehicle data:

- Hard drives swapped ~ 4 months
- SHRP 2 NDS data activities are mostly automated
- Therefore, most data "not seen" for several months after collection; "unseen" for many months after initial storage, i.e., no real-time knowledge of driver behavior

Crashes:

- System provides automatic notification of collisions via a small data packet.
- Does crash meet criteria for detailed investigation?
- If so, data may be viewed within days, depending on when DAS/hard drive is retrieved from the vehicle

Chain of Custody for Vehicle Data

- Data are encrypted when collected on the DAS
- Site contractor swaps the hard drive after 4 months; data remain encrypted.
- Encrypted data will be loaded onto a dedicated server at data facility, then uploaded to a secure server (ASAP) at VTTI or locked in secure facility. Data remain encrypted throughout.
- @VTTI: data is made readable, data is checked, driver ID is processed; re-encrypted and stored on secure server
- SHRP 2 NDS data base housed at VTTI for the duration of SHRP 2

Key Elements of Secure Data Base Management

DATA SECURITY PERSPECTIVE

- **Data base security:** Security measures that the data steward has in place to protect the data from unapproved use or distribution.
- **Control of access:** Mechanisms that act as a gate for data access within the data steward's organization.
- **Data access security:** Security measures in place to protect the data while it is being used.
- **Costs:** Associated with data base operations, maintenance, and security.

SHRP 2 NDS Physical/Electronic Data Security

- **Isolated server—no connection to VT system**
- **Restricted entry to building and to rooms containing server and computers; project-contingent access**
- **Viewing computers have no connection to internet and university system; no writeable drives or USB ports**
- **Computers and databases are password protected**
- **Recording devices (camera phones), laptops not permitted to be carried in**

4. Data Access

Key Elements of Secure Data Base Management

DATA USERS' PERSPECTIVE

- **Proposal review:** Proposals from researchers reviewed by funding agency; the data steward or a committee chosen by the data steward.
- **Data use requirements:** Conditions a researcher must agree to in order to access the data.
- **User training:** Training requirements for researchers who are approved to access the data.
- **Output control:** Controls imposed on data used by researchers in publications, presentations, or other forums as needed to protect the privacy of the participants.
- **Costs:** Access to the data will involve costs.

Data Access During SHRP 2

- **S08 Naturalistic Driving Study Analysis contractors are likely to be the first users**
  - **data sharing agreements consistent with approved consent form/protocol protections.**

- **Approach centers on role-based security for access to types or sets of data.**
  - **Non-identifying data**: steps that lead to open access
  - **De-identified data**: some restrictions based on probability of using data combinations to identify
  - **Identifying data**: most restrictive, data enclave/secure remote access
- **Need for IRB reviews: VT, researcher's institution, NAS**
- **S08 proposal reviews by S08 Analysis of the Naturalistic Driving Study Data ETG; others?**

Data Sharing Challenges

Major challenges both short-term and long-term regarding use, access, and sharing of identifiable information:

- For those not used to the requirements of human subjects research and the significance of identifiable information
- Established privacy procedures must be followed; contrast with current social milieu wrt privacy, personal boundaries
- Identifying external datasets with which the data could be merged
- For far-future users of the identifiable data: potential loss of institutional memory; legacy decisions
- Reaching out to the vast potential user population

Thank you!

joanne.harbluk@tc.gc.ca

Walter Diewald
SHRP 2 Senior Program Officer
Transportation Research Board
(202) 334-3260
wdiewald@nas.edu
http://www.trb.org/shrp2/