Safety IAP Issues Resolution Workshop

Pam Hutton, AASHTO SHRP2 Implementation Manager

David Plazak, TRB Associate Director for Safety Data

2016 TRB Safety Data Oversight Committee
May 10-11, 2016, Woods Hole, MA

U.S. Department of Transportation
Federal Highway Administration

AMERICAN ASSOCIATION
of STATE HIGHWAY and
TRANSPORTATION OFFICIALS
AASHTO

TRANSPORTATION RESEARCH BOARD
OF THE NATIONAL ACADEMIES

Presentation Agenda

- Meeting Summary
- Goals
- Key Issues
- Highlights of Workshop Discussion
- Action Items and Recommended Next Steps
- Potential Future Marketing Options for the NDS/RID
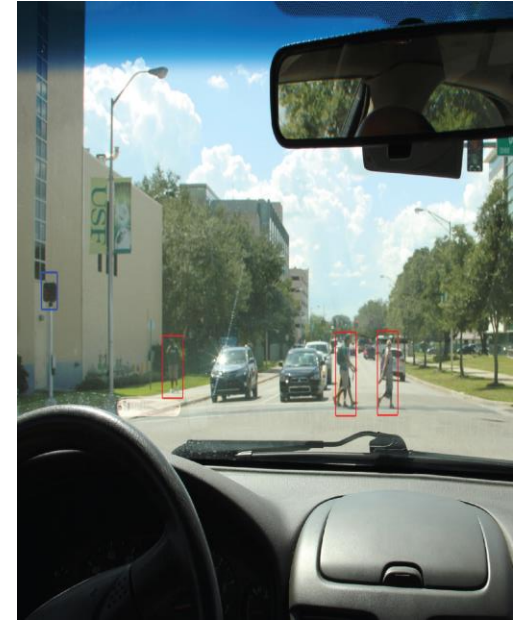
Issues Resolution Workshop

- Recommended by SDOC
- Opportunity for NDS/RID **Users** to have full discussions with NDS/RID **Providers** (VTII/IA State)
- 33 in person attendees, 3 call-ins
  - IAP Researchers
  - State Representatives
  - TRB Expert Task Group Members
  - SHRP2 Safety Task Force Members
 - Contractors
  - TRB, FHWA and AASHTO

Workshop Goals

- Receive input from **users** of NDS and RID databases

- Received input from **providers** about processes necessary to complete data collection requests

- Discuss ways to streamline requests and/or improve customer service after requests are initiated

- Arrive at "actionable resolutions" to improve the process for everyone moving forward

- Build stronger communication links between users and providers

Key Issues

- Process of Data Acquisition – Timing, Status, Cost, Contracting

- Enhancements to the NDS/RID – data quality

- Complex Structure of the Database and Implications for Users

- Personally Identifying Information (PII) – Constraints and Implications

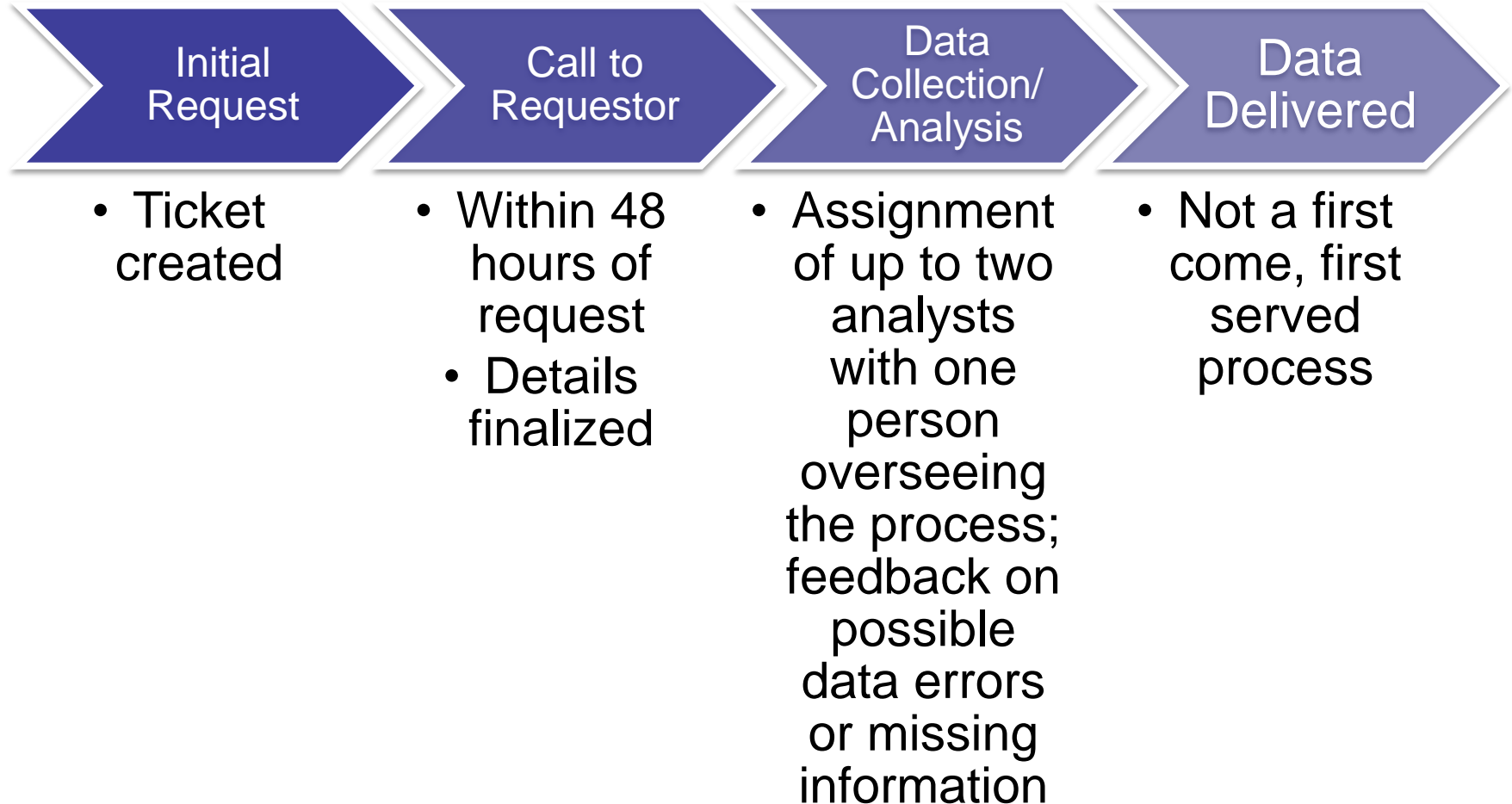- Modifications to
  Data User Licenses
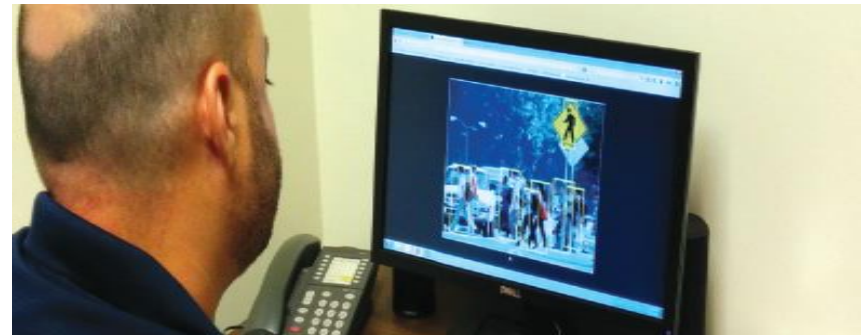
Workshop Agenda Overview

Time	Description
8:00 – 8:15 AM	Welcome and Introductions
8:15 – 8:30 AM	Workshop Overview
8:30 – 9:00 AM	Presentation of Efforts to Date to Addressing Known Concerns
9:00 – 10:15 AM	Discussion of Topics Pending
10:15 – 10:30 AM	Break
10:30 – 11:45 AM	Discussion of Topics Pending (cont.)
11:45 – 1:00 PM	Lunch
1:00 – 3:30 PM	PII and Parking Lot Topics
3:30 – 3:45 PM	Break
3:45 – 4:30 PM	Marketing of Data
4:30 – 5:00 PM	Wrap Up

Discussion Items

Efforts Underway to Improve the Process

Initial Request

- Ticket created

Call to Requestor

- Within 48 hours of request
- Details finalized

Data Collection/ Analysis

- Assignment of up to two analysts with one person overseeing the process; feedback on possible data errors or missing information

Data Delivered

- Not a first come, first served process

Typical Costs for Data
(from Exemplar Document)

Categories	Typical Groups	Example Areas of Interest	Level of Effort	Typical Timeline	Range of Resources
1: InSight-Only	Driver Behavior Risk Prevention Age-Related Driver Impairment & Medical Conditions	Driver Interactions and Traits	Low < 100 hours of Data Analyst time	< 1 Month	$500 - $750 Mean: $575 SD: $91
2: InSight-Expanded	Safety System Development Machine Based Learning	Modeling	Varies between low, moderate, and high based on complexity	Range: 1 month for low effort Over 2 months for high effort	$15,000 - $50,000 Mean: $27,361 SD: $15,754
3: Particular Location or Characteristic	Driver Behavior and Factors Roadway Infrastructure Vehicle & External Environment	Diverse (e.g., Distraction, Speeding, Seatbelt Use, Work Zones, Roadway Lighting)	Varies between low, moderate, and high based on complexity	Range: 1 month for low effort Over 4 months for high effort	$1,100 - $90,000 Mean: $24,510 SD: $26,695
4: Aggregate Data	Statistical Distributions Dataset Joins	Risk	Moderate to High	4 months	$45,000 - $275,000 Mean: $149,802 SD: $116,120

9

Battelle Effort and Analysis

- Battelle Study Overview

- Re-identification Risk Assessment – public use data set options

- Connection with remote enclave discussion – risks, costs, specifications, locations

- Connection to Data Review and Quality Analysis – speed data, video, terminology

Personally Identifying Information - User Perspective

- Biggest Challenge for Users was PII

- How to address circumstances under which the location of crashes may be usable by teams in their research, but not released publically?

  - Location could be made available in secure enclaves

  - Battelle looking into possibilities. Will report to SDOC in the future.

  - Commitment to NDS participants is biggest challenge (legal liability – serious consequences)

- Users need to clearly understand the criteria that are used to exclude vehicle traces from InDepth datasets that researchers receive.

Personally Identifying Information - Provider Perspective

- **Participant protection from public release of PII**

- Re-identification Risk Options – Removing 2/3 of variables doesn't improve this risk.

  - More categories allow for more unique cases which make cases less unique. Take 10 levels of a variable and chose only 3 (more nuanced approach).

  - Adding near misses with crashes – could make individual identification more difficult and be useful information at same time.

- Consider other categories of events that also have implications for PII – such as ticket data.

  - It is going to be a process to determine real risks and future risks. While trying to avoid show stoppers contractors have been conservative. There is no such thing as a "risk free situation."

- Biggest future risk is computer scientists who develop new algorithms to re-identify information using other public info (assessor's records, Google Earth, etc.) – worst case scenario could be stalkers, or those intent on looking for ID holes.

Options for More Access to PII - "Light Bulb" Moment

- All data is available at the secure enclaves. STAC will open at Turner Fairbank this summer
- Other options under consideration:
  - A secure enclave in the Midwest and/or West Coast
  - **Virtual enclave -** Rent space (a seat) on VTTI network to retrieve this information
- Longer-term:
  - **Individual enclaves - isolated, small, limited amount of PII released to a very limited group of people/agency.**
  - This type of approach has worked with other similar datasets
  - May need a pilot location
  - Would not be available for current IAP-related research projects

Next Steps

Workshop Recommendations

InSight web page:

- Provide extensive FAQs with tips on how to effectively navigate through the process:
    - Managing the request process
    - Potential hurdles and time delays
    - Typical time to receive data and costs
- Use the training data set as an example for cost of data retrieval and how changes affect those costs
- Clarify requests for large data amounts (10K trips or more) and what this entails

InSight Data Access Website
SHRP2 Naturalistic Driving Study
Learn More About SHRP2 NDS...
TRB TRANSPORTATION RESEARCH BOARD OF THE NATIONAL ACADEMIES
SHRP2 STRATEGIC HIGHWAY RESEARCH PROGRAM
VirginiaTech TRANSPORTATION INSTITUTE

Workshop Recommendations

- Enhance access to previously developed datasets
  - Encourage users to agree to share on Data Use License form when they have completed their work.
  - Make available a catalogue of data sets from researchers for others to reuse or build upon (such as work zone, safer data set)
  - Provide contact information for the datasets
- Explore enhanced access to data
  - **Individual enclaves and virtual enclaves**
  - Locate remote enclaves in the Midwest and West Coast

Workshop Recommendations

- Improve the interface between states, contractors and IRB's – through FAQs and other communications

  - Tracking lessons learned - questions researchers should ask

  - Providing info schedules and time frames,

  - Info on funding and contracting, how to work with lawyers

- Modify language to align it with current highway design terminology (Glossary or modification to legends).

- Develop a hierarchy list from users on what fields of information are practical and useful to them.

Marketing Discussion Items

Market Research Questions

1. *What do these data allow us to do that is new and different?*

2. *What are some key advantages and disadvantages of using these data?*

3. *What should the "Elevator Speech" about the data include?*

The answers are in TAB 3 of your binder.

Questions?