THE CENTER FOR
OPEN DATA ENTERPRISE

Improving Safety Data:

A Roundtable with the

U.S. Department of Transportation

TABLE OF CONTENTS

THE CENTER FOR
OPEN DATA ENTERPRISE

FOREWORD

Over the past half century, the U.S. Department of Transportation (DOT) has pursued a vision of a safe, dependable transportation system. DOT uses its safety programs and regulations as effectively as possible to reduce crashes, fatalities and injuries for all users of the transportation system. Underpinning our safety programs and regulations is a wealth of data, much of it made publicly available. That data informs our decision-making and supports our objective of the safest possible outcomes for all users of the transportation system.

The amount of data being generated by our safety programs is increasing and changing – and we recognize the need to continually improve the quality of the data we collect and share. We also understand that our data users – internal and external – are looking for DOT to leverage technology to more effectively share its data.

To help DOT better understand how we can improve our surface transportation safety data, we worked with the Center for Open Data Enterprise to hold an Open Data Roundtable. We believe that activities like the Open Data Roundtable - in combination with our other outreach activities- allow us to hear directly from data users inside DOT, other government partners, stakeholders, and our private sector partners. The feedback we receive through these channels helps DOT prioritize its open data activities to better serve all our of data users.

At DOT we are working relentlessly to lead the world in building a 21st century infrastructure designed to give our businesses and our people a competitive economic advantage and a better quality of life. In doing so, we need to harness data and technology to build a safer, more efficient and more effective transportation system.

Responding to the feedback we received from our data users will help us provide the best quality data in the most effective way possible, advancing the safety of the transportation system across the Nation. We look forward to working together to implement the findings and recommendations from this report.

Richard McKinney
Chief Information Officer, U.S. Department of Transportation

INTRODUCTION

On March 24, 2015, the U.S. Department of Transportation and the Center for Open Data Enterprise co-hosted an Open Data Roundtable in Washington, DC, bringing together data users with DOT data providers. The goal of this Roundtable was to discuss how to better leverage DOT data with a particular focus on surface transportation safety. This report combines observations, feedback and suggestions on all the topics discussed at the event, including those that apply to safety data specifically, to DOT data more broadly, or to the U.S. federal data system as a whole.

BACKGROUND

Open data from government – free data, accessible online, that anyone can use and reuse without restrictions– is being recognized as a major public resource. Under the federal Open Data Policy, government agencies are releasing more and more valuable datasets, fueling collaboration across the public and private sectors. The DOT Roundtable was one of a series of Open Data Roundtables conducted by the nonprofit Center for Open Data Enterprise.

This Roundtable was held at a time when DOT has been taking a number of steps to improve and expand its open data offerings. The day began with presentations on new DOT initiatives from officials at the National Highway Traffic Safety Administration (NHTSA), the Federal Motor Carrier Safety Administration (FMCSA), and the Federal Highway Administration (FHWA). The initiatives they reported on included:

- Data modernization projects to improve crash data with better sampling and state data gathering through new tools used at the crash scene, and increased focus on severe crashes.
- New APIs for recall data and consumer complaint data on automobiles and the NHTSA investigations database.

Improvements in the FHWA's nationwide database of public roads, highways, and bridges and new data releases being planned for the Travel Monitoring and Analysis System (TMAS).

- New ways of making key data available to meet researchers' specific needs, such as the recently updated Long Term Pavement Performance Program website and the Intelligent Transportation Systems Research Data Exchange.
- Data collection from the Naturalistic Driving Study, which has recorded the actions of volunteer drivers of privately owned vehicles.
- New ways of gathering data through smartphone apps to help regulate large trucks and buses.

Beyond these individual data initiatives, DOT has recognized the importance of combining datasets that bring different perspectives to serious safety problems. One of the most important issues flagged at the Roundtable was the need to combine crash data with hospital data in order to understand the long-term impacts of vehicle crashes and how different kinds of safety equipment can mitigate injury. The DOT Safety Council has committed to growing its Safety Data Action Team to include personnel with health and medical expertise. The Safety Data Action team will coordinate with the U.S. Department of Health and Human Services and the Health Data Consortium to address this issue raised at the Roundtable, as well as work with other government agencies on DOT safety priorities.

FORMAT OF REPORT

This report is being released as a public document to encourage further input, dialogue, and improvements in open data. It is divided into four main sections that highlight key findings from the Roundtable discussions. Each section includes a description of the topic, key challenges, and suggested solutions. It also includes commitments that DOT has made and opportunities for the private sector and civil society organizations to partner with the agency on specific data solutions. The annexes provide detailed material on DOT's open data and the Roundtable.

AUDIENCE

The report is designed to be of value to (1) operating administrations and offices at DOT that collect, analyze, and disseminate data, (2) Officials in other government agencies, (3) DOT's data stakeholders, and (4) members of the media and the general public.

The Center for Open Data Enterprise is an independent nonprofit organization that conducts Open Data Roundtables, which are action-oriented dialogues that bring together government agencies and the organizations that use their data. These dialogues between data experts inside and outside of government are designed to help extract the value of each agency's data reserves. The Open Data Roundtables were first launched as an initiative of The GovLab at NYU.

FINDINGS AND SUGGESTIONS FROM THE ROUNDTABLE

The Roundtable included a broad discussion of successes and opportunities for improvement in the provision, management and use of DOT data overall. It also included several specific suggestions for improving the use of safety data, which are noted at the end of this section.

SUCCESSES IN USING DOT DATA

The U.S. Department of Transportation (DOT) is widely seen as providing high-quality data in an accessible way. Data users at the Roundtable commended the Department for:

- Recognizing the importance of data: DOT is an increasingly data-centric agency. The Department has hired a Chief Data Officer, created a two-year plan for data, and overall has treated data as a priority.

- Data access: DOT's data inventory is largely complete, and the Department is starting to better understand its data, as well as the gaps and the overlaps between datasets. DOT provides open historical, geospatial, and granular data to users and has begun to provide some information through APIs (for example, the Federal Motor Carrier Safety Administration (FMCSA) SaferBus API).

- Data quality: DOT has broadly demonstrated a commitment to increasing the quality of its data. Some initiatives mentioned at the Roundtable include the Department's Licensing & Insurance system, as well as datasets produced by the Federal Transit Administration (FTA) and the Federal Highway Administration (FHWA).

- Data standards: DOT is participating in and fostering the development of international standards for sharing its data and information, particularly across the surface transportation domain (beginning with traffic records) and around aviation data (flight information, weather, and aeronautical information). The Fatality Analysis Reporting System (FARS) has already started using standardization across states for ease of data transfer.

- User engagement: Overall, DOT is viewed as very responsive to inquiries from its data users.

OPPORTUNITIES FOR IMPROVEMENT

While many data users praised the Department for the quality of its data and data management practices, Roundtable participants uncovered opportunities for further improvement. Several discussions centered on issues common to many federal agencies that publish data: user engagement, data availability and access, data quality, and data interoperability/integration. Others focused on concerns specific to safety data on surface transportation – specifically, data related to crashes, vehicle and manufacturer information, and traffic and navigation. The following sections outline issues that Roundtable participants identified and suggestions for action.

ISSUES COMMON TO ALL PROVIDERS

User Engagement

Many federal agencies are beginning to think of their data stakeholders – both current and potential – as "customers" for their data. This concept is the basis for developing a more user-focused approach to data collection and dissemination.

While the Department is generally responsive to its data users, participants noted areas where DOT could engage users more effectively. Overall, they believed that DOT could:

- Better publicize new open data initiatives
- Better communicate with regulated industries

- Make data easier to find

- Host a community forum to engage the industries that DOT regulates

Data users also asked for a way they could provide informal feedback on datasets and databases and know that their input is being considered.

Suggestions for DOT:

- Strengthen existing relationships between the Department and multiple Transportation Research Board committees. While many effective connections are in place, formalizing these relationships would help make the data more useful and keep the DOT abreast of the kinds of issues that researchers find when using the data. This kind of feedback works well for the National Household Travel Survey (NHTS), which could serve as a model.

- Open a 'data concierge desk' and provide a single point of access to DOT data.

- Consistent with applicable laws, regulations, and policies, develop ways to collect informal feedback from a broad spectrum of data users. Consider taking advantage of existing institutions, such as the Transportation Research Board, to hold "listening sessions" with data users including insurance companies, car manufacturers, and safety advocates.

- In addition to developing metadata and data dictionaries, create factsheets to help orient users to the data and its location.

- Develop a "no wrong door" approach to help users see, discover, and understand how data from around the Department is related. DOT websites do not provide links to other potentially relevant datasets.

- Use a broad range of communication tools to inform data users of the Department's data activities, such as social media, email listservs, tech blogs, and news outlets.

- Use open source data visualization tools to better understand and communicate the data to a large audience, and to make data gaps more visible. (Examples: CrossFilter, Zoomable)

- Host a community forum in a protected online space and use it to communicate the challenges the DOT and its users are encountering.

Data Availability and Access

Data repositories throughout the federal government use multiple formats and languages to help users access data, which may be stored in a variety of different and sometimes incompatible formats. Different users require varying modes of access to information; and providing multiple approaches can help ensure equality of access regardless of the user's technological capabilities.

Roundtable participants highlighted several key issues:

- Some key datasets still need to be digitized and made available for bulk access.

- GPS data specifically for optimizing truck routes that take into consideration roadway limitations (e.g. weight, lane, bridge restrictions) is difficult to link. This data needs to be obtained from many sources and reconciled; however, it is not available homogeneously across the United States. The Bureau of Transportation Statistics (BTS) and Federal Highway Administration (FHWA)have limited data on transportation risks in certain areas on the highway system.

- Many states do not have data on high-visibility and emerging issues such as distracted driving or drugged driving, and the data that does exist may be inconsistent or unreliable. DOT could do more to provide specialized products or enhance its datasets to facilitate analysis of these issues.

Suggestions for DOT:

- Make data more useful with bulk access, APIs, or additional data release where needed. If data is still in PDF format, ensure that legacy information is truly open and machine-readable.

- Communicate the new datasets that will be made available under MAP-21[1], which has identified new transparency initiatives, especially around highway and transit investments.

- Enable municipal governments to share their flyover data (compiled through private sector, high resolution aerial photography) to stitch together visual data on transportation.

- Consider policy options, including options relating to federal grant funding, that would facilitate improved reporting on high-visibility and high-impact safety matters, such as distracted and drugged driving.

- Set data standards in emerging areas so that states don't each have to invent their own reporting method(s). It may be helpful to have standing committees that can react quickly if new issues emerge – for example, on crash or roadway data standards - between the regular five-year update cycles.

Data Quality

At the Roundtable, data users emphasized the need for cleaner and more complete, valid, accurate, and interoperable data. The Roundtable participants flagged issues around data quality, many of them regarding state-collected data. Concerns include:

- Lack of standardization of state data not collected by the U.S. DOT. While standards exist, not all states are using them, and many state officials have not embraced their roles as data quality managers.

- Lack of consistency between state-provided motor vehicle crash data and Fatality Analysis Reporting System (FARS) and General Estimates System (GES) coding.

- High-quality data sources are not always fully leveraged. For example, the Federal Transit Administration's (FTA) National Transit Database, which is self-reported, has more complete information on transit funding than the Bureau of Transportation Statistics (BTS).

Suggestions for DOT:

- Apply data quality measurements, i.e. benchmark data periodically in order to see when data is deteriorating, and share these results openly. The Federal Motor Carrier Safety Administration (FMCSA) already does this, for example.

- Tie funding to demonstrable results to enforce compliance with data quality standards (for example, for data provided by Emergency Medical Services agencies).

- Evaluate states' data efforts and reward successes and 'champions'. Champions can have any level of technical ability.

- As DOT considers building out its own digital services team, consider state-level pilot projects to improve data collection.

- Improve validation procedures for state-level licensed driver estimates.

- Embed data into API header responses, such as 'as of date' to indicate the timeliness of the data disseminated.

- Initiate a study of the value and quality of crowdsourced data at the U.S. DOT level, focusing on known data gaps. Consider a broad range of crowdsourcing options, including the use of Federal fleet vehicles.

- As DOT considers its priorities for the use of big data, consider the value of "scraping" online forums for insights. One example: consider using Trucker Reports to provide more and better information on the location of warehouses. While this approach requires a lot of machine-learning, the private sector could help.

- Provide information to clarify the meaning and relevance of data on insurance for vehicles and address insurance validity.

---

1   Moving Ahead for Progress in the 21st Century Act (MAP-21), U.S. Department of Transportation. _http://www.transportation.gov/map21_.

Data Interoperability/Integration

Many Roundtable participants voiced the concern that datasets have grown and are managed in silos. Both DOT data stewards and external data users described a need to make datasets more interoperable. Specific concerns include:

- A need for common definitions and standards across the Department. Poor definition of terms and concepts is a problem throughout the data "supply chain." Many felt that the need for a "common language" is even more important than data formats.

- A need to improve state data, which is of variable quality and may not be collected in ways that make it interoperable with other data.

Suggestions for DOT:

- Continue to encourage states to collect more standardized data elements and attributes. Build the case that improved process is valuable both to them and others, and provide greater incentives.

- Develop more common data structures – which define the schema used to store data - and common metadata. In particular, DOT should continue its efforts to have states integrate their safety data and apply this guidance to its national datasets as well. This effort could begin with demonstration projects on state-federal data integration using existing standards such as Model Minimum Uniform Crash Criteria (MMUCC) and Model Inventory of Roadway Elements (MIRE).

- Develop common metadata across the whole DOT. The National Driver Register Problem Driver Pointer System (PDPS)[2] was cited as a best practice example.

- Develop consistent definitions/taxonomies (e.g. crash casualties, definition of "rural", metadata) and encourage their adoption. The goal shouldn't be to replace all the systems but to figure out how to deal with these differences; this is not a technical issue but rather a process issue: data should be in a common 'business language.' For instance, the American National Standards Institute (ANSI) defines motor vehicle crash related data, yet not everyone adopts this standard.[3]

- Keep a log and map changes to data schemas to allow understanding of historical data.

- Explore opportunities to integrate data between DOT and other federal agencies. For instance, the U.S. Consumer Products Safety Commission (CPSC) estimates injuries related to consumer products and issues recall and injury data. The CPSC is considering using data from the Fatality Analysis Reporting System, particularly related to all-terrain vehicles. Other instances of interagency collaboration and/or data integration include: the Department of Labor (working with the Federal Railroad Administration and the National High-way Traffic Safety Administration), the Department of Energy, the Environmental Protection Agency, the Department of Housing and Urban Development, the Federal Energy Regulatory Commission.

- Explore opportunities to enhance DOT data with private-sector data.

- Since DOT data is frequently useful in a geospatial context, address relevant considerations up front during the planning process to facilitate spatial connections and analyses.

ISSUES SPECIFIC TO SAFETY DATA

Crash Data

Roundtable participants discussed the need to improve vehicle crash data and connect it to medical data on crash victims. Crash data, which is collected at the state and local government levels, tends to be lacking in detail. A more sophisticated

---

2 National Driver Register (NDR), U.S. Department of Transportation. *http://www.nhtsa.gov/Data/National+Driver+Register+(NDR).*

3 ANSI D16.1-2007 Manual on Classification of Motor Vehicle Traffic Accidents (7th Edition). American National Standards Institute, August 2007. *http://www-nrd.nhtsa.dot.gov/Pubs/07D16.pdf.*

approach to crash data is necessary, particularly to track the long-term effects of injuries sustained in non-fatal crashes. There would be great value in connecting crash data with hospital data to get a fuller picture of the medical consequences of vehicle crashes and connecting those to vehicle design information and other factors.

There are a number of difficulties in collecting sufficient crash data at the scene of a crash. Hospital records are privacy-protected and could only be connected with crash data if privacy could be maintained. The records themselves may also not contain the information needed to link patients to crash events accurately. The data required by law enforcement and police reports is less detailed than what is needed to fully analyze a crash event. Today, while there are almost 6 million police-reported crashes a year, federal investigators sample 50,000 to gather more detailed information.[4] Without full investigations, the underlying causes of crashes are difficult to determine.

The DOT is making some headway in improving its data collection processes to fill the gaps in information. As DOT reported in a presentation at the beginning of the Roundtable, the Department is now working on several strategies to modernize crash data collection on the field, including using electronic tablet-based forms rather than paper-based forms, enhancing coding of crash reports, and using state data collection capacity rather than local police jurisdictions.

The Roundtable also discussed the technological approach of having Event Data Recorders (EDRs)[5] for each make of car, similar to reading the black box data in aircraft. At this point, however, these recorders are expensive and pose privacy concerns.

Suggestions for DOT:

- Explore options for combining crash and hospital data under secure conditions. Strategies can include use of a data enclave (for data from DOT, Health and Human Services, and insurance companies), anonymization techniques, or other approaches. While the Roundtable did not outline detailed, specific solutions, the need to combine these data sources emerged as a major safety priority.

- Revive NHTSA's Crash Outcome Data Evaluation System (CODES), a method for linking data from automated crash files to medical record information files at the state level to provide detailed data on the costs and consequences of motor vehicle injuries.

- Explore the potential for using private-sector data sources under appropriate conditions. One idea is to explore partnerships with insurance companies and a potential consortium approach to work with industry leaders to solve privacy issues. Other options include exploring work with the AAA Safety Foundation for Traffic Safety and using Waze's crash database.

- Conduct a study to measure the accuracy and completeness of police reports. When underreporting is identified, statistical corrections need to be developed to achieve more accurate estimates. Because police reports are not able to identify all crash factors, many of the current estimates are underestimating the safety impact of difficult-to-measure factors. For example, while NHTSA currently uses multiple imputation to correct for the underreporting of alcohol involvement in crashes, the agency does not yet correct for underreporting of other contributing factors such as driver distraction and fatigue.

- Allow the use of EDR data on a voluntary, opt-in basis – perhaps coupled with legislation that considers who owns EDR records.

Vehicle and Manufacturer Information

The Department is taking several steps to improve safety information on specific makes, models and vehicles. For example, DOT is developing new APIs for consumer complaint and recall data.

DOT and its data users would also like to see the vehicle identification number (VIN) used more effectively. It is currently

---

4   National Automotive Sampling System (NASS) General Estimates System (GES) Analytical User's Manual 1988-2013. National Highway Traffic Safety Administration, December 2014. *http://www-nrd.nhtsa.dot.gov/Pubs/812091.pdf.*
5   Event Data Recorder (EDR). U.S. Department of Transportation. *http://www.nhtsa.gov/EDR.*

necessary to go to several different databases for information on a single vehicle. In addition, the VIN system is limited, as it does not include information on safety features purchased with a particular vehicle. Better data integration could make the VIN a more effective tool for understanding how different features affect vehicle safety. The FHWA will soon begin a pilot project to collect VIN numbers for all registered motor vehicles rather than the aggregate registration data currently collected.

Roundtable participants stressed the importance of engaging manufacturers actively in the collection and use of safety data. While manufacturers use DOT data already, it would be valuable for them to share the data they collect, and to share improved versions of DOT data that they have cleaned.

Suggestions for DOT:

- Improve the VIN system to integrate more data more effectively.
- Engage auto manufacturers (who were not represented at this Roundtable) in a future Open Data Roundtable with DOT on safety.

Traffic and Navigation

Roundtable participants noted the potential to greatly increase the Department's knowledge of traffic patterns, dangerous road conditions, and related issues through data from third-party sources.

Suggestions for DOT:

- Explore public-private partnerships with third parties to improve the repository of traffic-related data. These can include such sources as:
    - ▸ GPS navigation companies, e.g. TomTom, Garmin, Navteq
    - ▸ Crowdsourced mapping enterprises:, e.g. OpenStreetMaps, Waze
    - ▸ Google Maps and similar services
    - ▸ U.S. Postal Service, FedEx, UPS
- Do so with attention to privacy concerns, licensing issues, reliability of the data.
- Encourage state/local efforts to develop traffic/navigation data. For example, while state highways come under federal jurisdiction, bridges are overseen by the states. New York City (the NYDOT) has developed a real-time bridge information system that could be a best practice model for other states as well.

APPENDIX A - BACKGROUND

ACRONYMS

AAA	American Automobile Association
AAMVA	American Association of Motor Vehicle Administrators
AASHTO	American Association of State Highway and Transportation Officials
BTS	Bureau of Transportation Statistics
CDS	Crashworthiness Data System
CIREN	Crash Injury Research
CSA	Compliance, Safety, Accountability
EDR	Electronic Data Recorder
FARS	Fatality Analysis Reporting System
FHWA	Federal Highway Administration
FMCSA	Federal Motor Carrier Safety Administration
FRA	Federal Railroad Administration
FTA	Federal Transit Administration
GES	General Estimates System
HPMS	Highway Performance Monitoring System
HSIS	Highway Safety Information System
ITS JPO	Intelligent Transportation Systems Joint Program Office
L & I	Licensing and Insurance
LTCCS	Large Truck Crash Causation Study
MMUCC	Model Minimum Uniform Crash Criteria
MCMIS	Motor Carrier Management Information System
MIRE	Model Inventory of Roadway Elements
NBI	National Bridge Inventory
NASS	National Automotive Sampling System
NEISS	National Electronic Injury Surveillance System (U.S. Consumer Product Safety Commission)
NHTSA	National Highway Traffic Safety Administration
NIEM	National Information Exchange Model
NiTS	Not-in-Traffic Surveillance
NMVCCS	National Motor Vehicle Crash Causation Survey
NTD	National Transit Database
NTSB	National Transportation Safety Board
RDE	Research Data Exchange
SAFER	Safety and Fitness Electronic Records
SMS	Safety Measurement System
TMAS	Travel Monitoring Analysis System
TRB	Transportation Research Board
VIN	Vehicle Identification Number

KEY DOT AGENCIES AND DATASETS

National Highway Traffic Safety Administration (NHTSA)

- Fatality Analysis Reporting System (FARS) – a census of all fatal crashes in the United States, going back to 1975. The database contains information about people and vehicles involved in a crash, as well as other important details.

- National Automotive Sampling System (NASS) – representative statistical samples of reportable crashes: one contains detailed crash investigation data for passenger vehicles and the other overall data on all severities of crashes involving all types of vehicles.

- Vehicle Recalls – a database of manufacturer-initiated and NHTSA-initiated recall campaigns.

- 5-Star Vehicle Safety Ratings - a database of crash protection and technology assessments for new vehicles.

- Child Safety Seat Ease-of-Use Ratings – a database of car seats that meet Federal safety standards and information about how easy they are to install and use.

Federal Highway Administration (FHWA)

- Highway Performance Monitoring System (HPMS) – a database of all public roads in the United States, key characteristics and, for Federal-aid roads, additional information about pavement conditions and annual average daily traffic.

- National Bridge Inventory (MBI) – a database of all the bridges in the United States, key features about those bridges, and their condition.

- Highway Safety Information System (HSIS) – a research-grade database that integrates crash, roadway inventory, and traffic data from select states.

Federal Motor Carrier Safety Administration (FMCSA)

- Motor Carrier Census (MCMIS) – a database of all companies that have been assigned a USDOT number, their contact information, and their operating authority.

- Safety Measurement System Results (SMS) – a database of motor carrier safety ratings.

- Licensing & Insurance (L&I) – a database with information about the level of insurance and other licensing information for motor carriers.

- Enforcement Cases – a database of motor carrier safety violations that resulted in fines.

DOT OPEN DATA POLICIES

The Department began by developing a data inventory to meet the September 30, 2010 deadline in the Open Government Plan. Since then, DOT publishesa list of its publicly available datasets on its website and data.gov. DOT's next step was the development of a Departmental Data Release policy to guide the release of data in open, accessible formats. The policy was finalized and signed in March 2011.

In November 2013, DOT published a new Data Inventory page, meeting the requirements of the Open Data Policy and the DOT's commitment in its Open Government Plan. As the Department develops its data inventory, each dataset is scored using a set of guidelines for identifying high-value datasets. Datasets are also registered in a DOT registry that captures metadata and scoring information on each dataset. Candidate datasets are identified by:

- Reviewing IT systems and determining what data already resides in those systems

- Reviewing information collections and forms to identify potential data sources

- Surveying DOT websites to see what data is currently published/available

- Asking the members of DOT's data.gov group to identify additional datasets

Over 140 data sets have been registered and scored, creating a prioritized inventory of high-value data.

In August 2014, DOT unveiled a new, free, online search tool for recalls using vehicle ID numbers. It allows consumers to tell whether their vehicle or a used vehicle they are considering is at risk due to an uncompleted recall.

The Department is set to deploy its cloud-based geospatial information system, Visual DOT. It will pilot novel visualizations of the DOT's data, targeted to both enhance transparency and provide a more interactive approach for citizens to engage with the data.

As a service to the developer community, DOT maintains a listserv for updates to announce when new resources are made available.

PRE-ROUNDTABLE QUESTIONNAIRE RESULTS

Prior to the DOT Roundtable, the organizers of the Roundtable circulated a questionnaire to all invitees both from DOT and from outside the department. They were asked to respond whether or not they were able to attend the Roundtable. The organizers received 39 responses, 30 of which were from people who attended the event. While individual questionnaire responses are confidential, the Center for Open Data Enterprise has paraphrased and summarized responses to the questions here.

What changes would make it easier to identify, access and use DOT data?
- API access to a greater number of key datasets
- More states participating in the Highway Safety Information System
- Ability to access DOT data through a central clearinghouse
- More timely DOT open data
- Ability to receive the DOT census file electronically rather than on DVD
- Greater availability of Highway Performance Monitoring System (HPMS) data
- More programmatic and bulk access to files, rather than publishing Excel files annually, which makes longitudinal analysis difficult

What issues regarding standards and quality would you like to discuss?
- Quality of health and transportation data
- Consistency of data standards, format, and structure
- Release of more data even if quality is not ideal
- Access to raw data rather than aggregated/summarized data
- Improving metadata
- Improving state crash data consistency and data transfer

What issues/topics regarding feedback loops and communication would you like to discuss?
- Data submission guidelines and considerations for PII
- Access to a data specialist for help with data issues
- Ability to provide comments through feedback links
- Tools like GitHub and other developer networks for transportation data users
- Approaches to presenting data in jargon-free ways for non-technical audiences

What outcomes would you like to see from this roundtable?
- A roadmap for a comprehensive Transportation Data portal
- Networking opportunities among data owners, stewards, and users
- Greater accessibility of Federal Motor Carrier Safety Administration (FMCSA) data

THE CENTER FOR
OPEN DATA ENTERPRISE

CURRENT DATA USE (In order of most commonly used)

Which DOT data does your organization currently use?

> National Highway Traffic Safety Administration (NHTSA)
>
> Federal Motor Carrier Safety Administration (FMCSA)
>
> Federal Highway Administration (FHWA)
>
> Bureau of Transportation Statistics (BTS)

How does your organization currently use the DOT's data

> Provides insights based on open data (e.g. analytics, visualizations)
>
> Processes open data (e.g. aggregation, classification, anonymization, cleaning, refining, enriching)
>
> Develops products based on open data (e.g. APIs, apps)
>
> Provides infrastructure for others to publish open data (e.g. platforms, portals, data stores)

APPENDIX B - U.S. DOT ROUNDTABLE

U.S. Department of Transportation & Center for Open Data Enterprise

DOT Open Data Roundtable Agenda

Tuesday, March 24, 2015

1310 North Courthouse Road, Suite 300, Arlington, VA 22201

9:30 AM	Registration and Coffee
10:00 AM	Welcome
	Maria Roat - Chief Technology Officer, U.S. Department of Transportation
10:15 AM	Structure of the Day
	Joel Gurin - President, Center for Open Data Enterprise
10:30 AM	DOT Presentations

- Data Modernization Efforts – *Tina Morgan*, NHTSA
- Defects Investigation Data – *Frank Borrus*, NHTSA
- National Data Products – *David Winter*, FHWA
- Improving Access to Data – *Jonathan Mueller*, FMCSA

11:00 AM	Break
11:15 AM	Session 1: Using DOT Data for Surface Transportation Safety
12:45 PM	Lunch
1:30 PM	Session 2: Data-Driven Solutions for Surface Transportation Safety
3:00 PM	Break
3:15 PM	DOT Commitments and Next Steps
	Richard McKinney - Chief Information Officer, U.S. Department of Transportation
3:35 PM	Closing
	Joel Gurin - President, Center for Open Data Enterprise
3:45 PM	Adjourn for Reception

LIST OF PARTICIPANTS

GOVERNMENT AGENCIES AND OFFICES

The U.S. Department of Transportation (DOT)

The U.S. Department of Transportation serves the United States by ensuring a fast, safe, efficient, accessible and convenient transportation system that meets our vital national interests and enhances the quality of life of the American people, today and into the future.

Office of the Assistant Secretary for Research and Technology (OST-R)

The Research and Innovative Technology Administration (RITA) coordinates the U.S. Department of Transportation's (DOT) research programs and is charged with advancing the deployment of cross-cutting technologies to improve the U.S. transportation system.

Bureau of Transportation Statistics (BTS)

The Bureau of Transportation Statistics was established as a statistical agency to administer data collection, analysis, and reporting and to ensure the most cost-effective use of transportation-monitoring resources.

Patricia Hu

National Transportation Library (NTL)

The National Transportation Library is a repository of U.S. DOT transportation information, serves as a portal transportation data, and provides library services to U.S. DOT employees.

Amanda Wilson

Volpe, The National Transportation Systems Center

Volpe partners with public and private organizations to assess the needs of the transportation community, evaluate research and development endeavors, assist in the deployment of state-of-the-art transportation technologies, and inform decision- and policy-making through their comprehensive analyses.

John Brewer

Office of the Chief Information Officer

The Office of the Chief Information Officer (CIO) serves as the principal advisor to the Secretary of Transportation on matters involving information and technology. The DOT CIO also has oversight responsibility over the entire Department of Transportation (DOT) IT portfolio of over $3 billion annually, the 6th largest in the federal government.

Richard McKinney
Dan Morgan
Maria Roat

Federal Highway Administration (FHWA)

The Federal Highway Administration is an agency within the U.S. Department of Transportation that supports State and local governments in the design, construction, and maintenance of the Nation's highway system (Federal Aid Highway Program) and various federally and tribal owned lands (Federal Lands Highway Program).

James Pol
Robert Pollack
Stuart Thompson
David Winter

Federal Motor Carrier Safety Administration (FMCSA)

The primary mission of the Federal Motor Carrier Safety Administration (FMCSA) is to reduce crashes, injuries and fatalities involving large trucks and buses. In carrying out its safety mandate to reduce crashes, injuries, and fatalities involving large trucks and buses, FMCSA:

- Develops and enforces data-driven regulations that balance motor carrier (truck and bus companies) safety with efficiency;
- Harnesses safety information systems to focus on higher risk carriers in enforcing the safety regulations;
- Targets educational messages to carriers, commercial drivers, and the public;
- Partners with stakeholders including Federal, State, and local enforcement agencies, the motor carrier industry, safety groups, and organized labor on efforts to reduce bus and truck-related crashes.

Jonathan Mueller
Steven Smith

National Highway Traffic Safety Administration (NHTSA)

Frank Borris
Augustus 'Chip' Chidester
Christina Morgan
Terry Shelton

NHTSA was established by the Highway Safety Act of 1970 and is dedicated to achieving the highest standards of excellence in motor vehicle and highway safety. It works daily to help prevent crashes and their attendant costs, both human and financial.

Bureau of Labor Statistics, U.S. Department of Labor

The Bureau of Labor Statistics is the principal fact-finding agency for the Federal Government in the broad field of labor economics and statistics.

Christen Byler

Maryland State Highway Administration

The Maryland SHA owns and maintains all non-toll, numbered roads in Maryland's 23 counties.

Eric Tabecek

National Center for Health Statistics, Centers for Disease Control and Prevention

The mission of the National Center for Health Statistics (NCHS) is to provide statistical information that will guide actions and policies to improve the health of the American people. As the Nation's principal health statistics agency, NCHS leads the way with accurate, relevant, and timely data.

Holly Hedegaard
Margaret Warner

U.S. Consumer Product Safety Commission (CPSC)

CPSC is charged with protecting the public from unreasonable risks of injury or death associated with the use of the thousands of types of consumer products under the agency's jurisdiction. CPSC is committed to protecting consumers and families from products that pose a fire, electrical, chemical, or mechanical hazard.

Jean Mah

COMPANIES, NONPROFIT, ACADEMIC AND RESEARCH ORGANIZATIONS

AAA

AAA is a not-for-profit member service organization with more than 53 million members which provides services such as travel, automotive, insurance, financial, and discounts

Nathan Warren-Kigenyi

Alan Pisarski Consulting

Consulting, writing and analysis in the fields of transportation research, policy and investment.

Alan Pisarski

Amazon Web Services (AWS)

Amazon provides cloud computing services through Amazon Web Services to a range of clients.

Ariel Gold
Sharon Luther

Battelle Transportation

Battelle provides innovative technology and technical services with an emphasis on development of intelligent transportation systems, technology transfer and commercialization, and hardware/software product development.

Michael Waisley

Center for Open Data Enterprise

The Center for Open Data Enterprise is an independent nonprofit organization that develops smarter open data strategies for governments, businesses, and other nonprofits by focusing on data users.

Audrey Ariss
Katherine Garcia
Joel Gurin
Gustavo Magalhaes
Laura Manley

Consumer Reports

Formed as an independent, nonprofit organization in 1936, Consumer Reports serves consumers through unbiased product testing and ratings, research, journalism, public education, and advocacy.

William Wallace

DC News & Data LLC

DC News & Data is an information search and retrieval services company.

David Cleveland
Gary Plummer

IHS

Through a combination of information, analytics and expertise, IHS enables smarter, better decision making in everything from day-to-day operations to long-term investments.

Richard Stucky

Microsoft

Microsoft develops, manufactures, licenses, supports and sells computer software, services, devices and solutions that help people and businesses realize their full potential.

Geary Brummell
Jean-Sebastien Brunner
Gene Burrus

PricewaterhouseCoopers

PricewaterhouseCoopers is a network of firms in 158 countries that delivers quality in assurance, tax and advisory services.

Jim Dreyer
Joe Gulisano

Socrata

Socrata's mission is to connect people to the government data they need and want.

Kathy Hsu
Joe Pringle

The Center for Advanced Transportation Technology Laboratory (CATT Lab), University of Maryland

The Center for Advanced Transportation Technology Laboratory at the University of Maryland was originally established in 2002 as an academic applied research and development lab to support national, state, and local efforts to solve important transportation, safety, and security problems.

Nikola Ivanov
Michael Pack

Transfix.io

Transfix is a fully automated freight brokerage company.

Jonathan Salama

Transit Labs

Built on decades of reported data and billions of real-time transit data points, TAP empowers transit systems, DOTs, and MPOs with actionable intelligence to make data-driven decisions.

Farhan Daredia

University of Michigan Transportation Research Institute (UMTRI)

The University of Michigan Transportation Research Institute is dedicated to achieving safe and sustainable transportation for a global society. UMTRI is committed to interdisciplinary research that will ultimately increase driving safety and further transportation systems knowledge.

Carol Flannagan

VHB

VHB works to improve mobility, enhance communities, and balance development and infrastructure needs with environmental stewardship.

Robert Scopatz

SPONSORS

The Center for Open Data Enterprise thanks its sponsors for their support of the Open Data Roundtables.

OPEN DATA PARTNER

Microsoft is a worldwide supplier of devices and services that help people and businesses realize their full potential. Customers use Microsoft's products to find creative solutions to business problems, develop breakthrough ideas, and stay connected to what's most important to them. For more information, see *www.microsoft.com*.

SERIES SPONSORS

Amazon Web Services offers a broad set of global compute, storage, database, analytics, application, and deployment services that help both large enterprises and start-ups move faster, lower IT costs, and scale applications. Amazon Web Services Worldwide Public Sector is helping government and education customers employ cloud services to reduce costs, drive efficiencies, and increase innovation across the globe. Public Sector organizations of all sizes use AWS to build applications, host websites, harness big data, store information, conduct research, improve online access for citizens, and more. *For more information, see www.aws.amazon.com/gov.*

PricewaterhouseCoopers is a network of firms in 158 countries that delivers quality in assurance, tax and advisory services. PWC helps federal agencies meet the challenge of integrating financial and performance systems to help make decisions and improve accountability. Clients look to the PricewaterhouseCoopers Public Sector Practice to bring direct hands-on knowledge of federal standards for systems, internal controls, and financial reporting. The Practice assists clients through creating interactive data, developing an agile reporting and analytic framework, and identifying and implementing improvements to the data and information supply chain. To find out more, visit *www.pwc.com/publicsector*.

EVENT SPONSOR

Socrata is focused on democratizing access to government data. Its solutions help government leaders improve transparency, modernize citizen access to information and bring facts into every decision, with unprecedented speed and cost savings. Delivered as turnkey cloud services, Socrata's data consumerization products unlock data in enterprise data silos and transform it into useful information that everyone can easily access, visualize, share and reuse. For more information, visit *www.socrata.com*.

THE OPEN DATA ROUNDTABLE SERIES

The Open Data Roundtables (opendataenterprise.org/convene) are action-oriented dialogues that bring together government agencies and the organizations that use their data. By hosting the Roundtables, the Center for Open Data Enterprise offers a low-tech solution to a high-tech problem. The Roundtables are designed to:

- **Identify high-value datasets** so agencies can address them as a priority;
- **Develop solutions** to make data more accurate, complete, usable, and useful; and
- **Connect data providers and users** for ongoing collaboration.

After each Roundtable, the Center for Open Data Enterprise issues a public report summarizing the participants' discussion of data challenges and opportunities, their proposals for solutions, and agencies' commitments to action. The Open Data Roundtables were originally launched as an initiative of The GovLab at NYU.

THE CENTER FOR OPEN DATA ENTERPRISE

The Center for Open Data Enterprise is an independent nonprofit organization that develops smarter open data strategies for governments, businesses, and other nonprofits by focusing on data users. We believe that open data can support economic growth and social good around the world. Our user-centered approach aims to improving the open data ecosystem in three ways. We **map** the uses of open data from around the world; **convene** data users and providers to identify challenges and opportunities; and **implement** solutions driven by user input. For more information, see *www.OpenDataEnterprise.org*.