2nd Annual USDOT Data Palooza

June 3rd – 4th, 2014
Hilton Crystal City
Arlington, VA

Summary of Proceedings

Notice

Table of Contents

List of Acronyms

AADT	Annual Average Daily Traffic
AASHTO	American Association of State Highway and Transportation Officials
API	Application Programming Interface
APTA	American Public Transportation Association
ATIP	Automated Track Inspection Program
BTS	Bureau of Transportation Statistics
CoP	Community of Practice
DOT	Department of Transportation
EDR	Event Data Recorder
EPA	Environmental Protection Agency
FHWA	Federal Highway Administration
FMCSA	Federal Motor Carrier Safety Administration
FRA	Federal Railroad Administration
FTA	Federal Transit Administration
GIS	Geographic Information Systems
GROW AMERICA	Generating Renewal, Opportunity, and Work with Accelerated Mobility, Efficiency, and Rebuilding of Infrastructure and Communities throughout America Act
GTFS	General Transit Feed Specification
GTMA	Geospatial Transportation Mapping Association
HPMS	Highway Performance Monitoring System
ICM	Integrated Corridor Management
IDCC	International Digital Curation Center
ITS	Intelligent Transportation Systems
JPO	Joint Program Office
LiDAR	Light Detection and Ranging remote sensing technology
MAP-21	Moving Ahead for Progress in the 21st Century
MPO	Metropolitan Planning Organization
MTS	Metropolitan Transit System
NHTS	National Highway Traveler Survey
NHTSA	National Highway Transportation Safety Administration
NTD	National Transit Database
NTL	National Transit Library
NTSB	National Transportation Safety Board
OMB	Office of Management and Budget
PHMSA	Pipeline and Hazardous Materials Safety Administration
PTC	Positive Train Control
RDE	Research Data Exchange
SAE	Society of Automotive Engineers
SHRP2	Strategic Highway Research Program 2
TCIP	Transit Communications Interface Profiles
TFHRC	Turner-Fairbanks Highway Research Center
TPM	Transportation Performance Management
TRB	Transportation Research Board
USACE	United States Army Core of Engineers
USDOT	United State Department of Transportation
V2I	Vehicle-to-infrastructure
V2V	Vehicle-to-vehicle
VMT	Vehicle Miles Traveled

Executive Summary

Purpose

Effective data collection, analysis, standardization, and dissemination are essential elements in the transportation industry's desired shift toward more performance-driven decision-making processes. The Second Annual U.S. Department of Transportation (USDOT) Data Palooza brought together transportation professionals from a variety of sectors and disciplines to seek common ground in this arena.

Held July 3-4, 2014, in conjunction with the Geospatial Transportation Mapping Association's (GTMA) Annual Meeting in Arlington, VA, Data Palooza emphasized the significance of the influx of data being collected today and the larger implications that come with amassing such immense amounts of information. While data collection efforts are soaring, data analysis tools and standardization rules are struggling to catch up. USDOT concentrated the Second Data Palooza on developing action items for the Federal government to assist public and private sector stakeholders in sharing and analyzing data.

Data Palooza organized morning and afternoon workshops on two topics: 1) Standards for Sharing Transportation Data and 2) Analyzing Integrated Data. John Thomas from the Utah Department of Transportation provided the Keynote Address on Data Integration, highlighting Utah's three-year effort to implement an integrated geospatial mapping and data network platform. The conference concluded with breakout discussions on Innovation Action Planning and Commitments for both transit and highways, centered on developing action items for public and private sector collaboration in improving transportation data.

Themes

The **Standards for Sharing Transportation Data** workshops focused what is already occurring within the industry and areas upon which to improve. The development of data standards is important for the efficient and consistent analysis of transportation data. It is an arduous task as it requires building consensus with data users who each have different data purposes, but it is necessary to lay the groundwork now before partners move too far in different directions. Participants enunciated a role for USDOT in creating and enforcing national baseline data standards and establishing clearinghouses to more easily integrate and analyze different data sources.

The public and private sectors are working to make the mass amounts of data collected available to the public, but there are several challenges to this process. Presenters and participants articulated the following **barriers to data sharing**:
- Data are not maintained to research quality due to human and financial resource shortages.
- Practitioners are fearful of the consequences of sharing bad data.
- Data users may mishandle the information to develop misleading analyses.
- Many data are proprietary and therefore inaccessible to the public.
- Data sharing within an agency must be resolved before creating a public-facing warehouse.
- Data privacy is a sensitivity to be mindful of when sharing information publicly.

Participants noted that while public and private agencies are working internally to integrate and share data, greater strides can be made in sharing data externally.

During the **Analyzing Integrated Data** sessions, presenters showcased the latest data applications within USDOT and discussed new analytical research methods. As Plenary Session panelist and TRB Executive Director Robert Skinner noted, "We are entering a data revolution." The amount of data being collected today is more than ever imaginable due to advances in technology and the use of non-traditional data sources. With more data comes more potential for transportation agencies to apply data to improve the condition and performance of their respective systems. Current USDOT initiatives focus on innovative ways to display data in a visual manner comprehensible to the general public. Mapping tools enable multiple

datasets to be overlaid to display a comprehensive picture of system condition and performance. Better visualization and analysis tools help identify specific problems areas and enable more innovative and effective solutions.

Presenters and attendees identified the following problems and issues related to analyzing data:
- Data benefits are realized over a longer timeframe, but budget decisions are predisposed to favor activities with near-term results.
- Data sets are not maintained to research quality, thus making integration more difficult.
- Big data maintenance and computation are time-consuming and may have interpretive biases.
- Big data can overwhelm a system and agency; therefore, data processing must be driven by purpose and necessity to gain the most benefits.

More information is available in areas such as travel demand and travel behavior for both transit and highways. Big data bring even more possibilities, but more research needs to be done to confirm the effectiveness and significance of some of the data being collected. While non-traditional data sources coming from social media, smartphones, crowdsourcing, and digital images can fill in information gaps and provide novel insights, it is often harder to clean and process these data. If maintained and applied appropriately, these new data sources can move transportation projects from reactive to predictive and enable a more holistic program evaluation.

Transit and highway professionals face common problems related to data and institutional capacity. Practitioners articulated these challenges during the concluding discussion sessions on **Innovation Action Planning and Commitments**, among them:

- Practitioners need to be able to effectively communicate the value of data to leadership.
- Demands on the transportation system have been increasing while funds are decreasing, making the need for innovative and economic data applications even more pertinent.
- Although more data brings more potential, it also means there are more challenges for the industry to overcome.
- The complexities and costs involved with data maintenance and thorough analysis are overlooked or go unresolved.

Action Items

USDOT has processes and tools available to help address the technical, institutional, and cultural issues related to data. Attendees identified the following action items for USDOT to help bridge many of the issues identified in the sessions:
1. Break down modal data silos;
2. Promote understanding of the importance of data collection and data management among executives to support innovation;
3. Support multi-modal expansion of data coordination and collaboration among agencies;
4. Establish data communities of practice;
5. Support data clearinghouses;
6. Establish processes for data standardization;
7. Provide best practices for data maintenance; and
8. Provide training to help address skills shortage in this field.

Data Palooza created a forum to highlight and begin addressing data challenges faced within the transportation industry. With data influencing the day-to-day operations of agencies, it is important to sustain and support the conversation to find solutions to data needs and barriers. Data standards, sharing, and analysis must be built upon a coordinated approach. The public and private sector can work together to generate consensus, open data sources, and advance innovative applications. USDOT can be a leader in this effort and in continuing the dialogue. **Select presentations slides can be found at https://speakerdeck.com/datapalooza.**

Meeting Agenda

The 2nd Annual Data Palooza was a two-day event held at the Hilton Crystal City in Arlington, VA. The event began at 8:00 AM with Welcoming and Opening Remarks from Acting Deputy Secretary of Transportation Victor Mendez. Deputy Secretary Mendez announced the winners of USDOT's Data Innovation Challenge (http://www.transportation.gov/fastlane/dot-data-innovation-challenge-attracts-talented-field). Following this, the Plenary Panel Session brought several prominent transportation leaders together to discuss the future role of data in transportation. Day One focused on standards for sharing transportation data and had two sessions from 10:30 AM -12:00 PM and 3:00 PM – 5:30 PM, with the Keynote Session from 2:00 PM – 3:00 PM. Data Palooza - Day Two morning sessions (8:30 AM - 10:00 and 10:15 AM – 12:00 PM) concentrated on the variety of ways transportation data is being analyzed. The GTMA likewise sponsored related data collection and mapping applications sessions during the morning and afternoon of both conference days. Two of those GTMA-sponsored sessions emphasizing transit data issues are highlighted in these Data Palooza proceedings. To close the two-day event, a break out session for innovation action planning was held from 1:30 PM – 2:50 PM. The attendees split into a transit-focused group and a highway-focused group. Both breakout groups discussed and formulated ways the private and public sectors could improve data for transportation applications. The conference participants reconvened from 3:00 PM – 4:00 PM to report out on the actions and commitments discussed.

Welcome and Opening Remarks
Tuesday, June 3, 2014, 8:00am-8:30am
- Victor Mendez, Acting Deputy Secretary of Transportation, USDOT
- Peter Stephanos, Director, Transportation Performance Management Office, Office of Infrastructure, FHWA, USDOT
- Robert Dingess, GTMA President and CEO

USDOT's Data Innovation Challenge Winners:
- RideScout (multiple option travel information mobile application)
- Choices and Voices – Delaware Valley Regional Planning Commission
- FAST Dashboard – Regional Transportation Commission of Southern Nevada

Plenary Panel Session: The Future Role of Data in Transportation
Tuesday, June 3, 2014, 8:30am-10:15am
Moderator: Larry Copeland, Transportation Reporter, USA Today

- Robert F. Skinner, Executive Director, Transportation Research Board (TRB)
- Michael Trentacoste, Associate Administrator, Research, Development, and Technology, FHWA
- Richard McKinney, Chief Information Officer, USDOT
- Kenneth Leonard, Associate Administrator and Director, ITS Joint Program Office
- Rolf Schmitt, Deputy Director, Bureau of Transportation Statistics (BTS)

Standards for Sharing Data Part 1
Tuesday, June 3, 2014, 10:30am-12:00pm
Moderator: Patrick Son, ITS America

- John Brophy, Chief of Crash Investigation Division, NHTSA, USDOT, "National Automotive Sampling System (NASS) – A Data Modernization Project"
- Yolanda Braxton, Program Analyst, Pipeline and Hazardous Materials Safety Administration (PHMSA), "PHMSA Data Initiatives"
- Zeke Bishop, Enterprise Architect, Oracle, "Enterprise Level Data Integration - Data Governance"

Keynote Session: Data Integration – Breaking Down Barriers to Change…
Tuesday, June 3, 2014, 2:00pm-3:00pm
John Thomas, UPLAN Project Manager, Utah DOT

Standards for Sharing Data Part 2
Tuesday, June 3, 2014, 3:00pm-5:30pm
Moderator: Dag Gogue, CEO, Transit Labs, "Developing a Transportation Data Clearinghouse to Handle Data, Data Experts, and Transportation Research"

- Chris Allen, Chief, Office of Highway Policy Information, FHWA, USDOT, "Map-21 and the All Roads Network: Geospatial Data for Transportation and Beyond"
- Michael L. Pack, Director, CATT Laboratory, University of Maryland, "Improving Transportation Operations and Management through Data Integration"
- Waldo Jaquith, U.S. Open Data Institute, "President's Open Data Initiative"
- Mary E. Moulton, Digital Librarian, National Transportation Library, "What's on the BTS Data Clearinghouse and Why?

Analyzing Integrating Data Part 1: US DOT Initiatives
Wednesday, June 4, 2014, 8:30am-10:00am
Moderator: Maria Chai, Senior Community Planner, FHWA – New York Division

- Rolf Schmitt, Deputy Director, BTS, USDOT, "BTS' GIS Portal - The National Transportation Atlas"
- Tianjia Tang, Chief, Travel Monitoring and Surveys Division, Office of Highway Policy and Information, FHWA, USDOT, "FHWA's Integrated Data Analysis Efforts and Plan"
- Steve Smith, Director, Office of Analysis, Research, and Technology, FMCSA, USDOT, "FMSCA Data Sources and Uses"
- Nicole Katsikides, Transportation Specialist, Office of Freight Management and Operations, FHWA, USDOT, "FHWA's Vehicle Probe Data (National Performance Management Research Data Set)"

Analyzing Integrating Data Part 2: Analytical Research Methods
Wednesday, June 4, 2014, 10:15am-12:00pm
Moderator: Joe Moyer, Quality Assurance Manager, Turner Fairbanks Highway Research Center, FHWA, USDOT

- Adella Santos, NHTS Program Manager, FHWA, USDOT, "National Household Travel Survey: Data Program (Long Distance Travel Origin-Destination Data Apps)"
- Laurie Schintler, Associate Professor, School of Public Policy, George Mason University, "Big Data for Travel Demand Modelling (Overview of Using Big Data to Explore Travel Behavior)"
- Stephen Crim, Director of Research, Mobility Lab, "Big Data, Travel Behavior, and Transportation Demand Management"

GTMA TransData Expo
Wednesday, June 4, 2014, 8:30am-12:00pm

Rail / Transit Data Management Issues
Moderator: Bill Toothill, Manager, Global Mapping Solutions / Technology Services Group, DBI Services

- Raquel Hunt, FRA, USDOT "Improving Railroad Asset Data Management"
- Keith Gates, Strategic Planning and Analysis Division, Office of Budget and Policy, FTA, USDOT "Performance and Asset Data – National Transit Database"
- Robert Borowski, GISP, Business Systems Analyst, San Diego Metropolitan Transit System "Transit Asset Data Collection Case Study"

Using General Transit Feed Specification [GTFS]
Moderator: John Giorgis, Director, Strategic Planning and Analysis Division, Office of Budget and Policy, FTA, USDOT

- Andrew Owen, Director, Accessibility Observatory, University of Minnesota "Total Accessibility Calculations and GTFS"
- Alexander Bell, Senior Planner, Renaissance Planning Group "EPA Smart Location Database and Access to Jobs and Workers via Transit Tool"
- Kurt Raschke, Mobility Lab "One Bus Away: Suite of Open Source Tools"
- Adie Tomer, Senior Research Associate and Associate Fellow Metropolitan Policy Program, Brookings Institution "Leveraging GTFS Data to Improve American Transit"

Data Innovation Discussions: Action Planning and Commitments
Wednesday, June 4, 2014, 1:30pm-4:00pm

- Transit Data Breakout Discussion
- Highway Data Breakout Discussion
- Report Out of Action Plans and Commitments

Data Palooza 2014 and GTMA Presentations

A. Welcome and Opening Remarks

Robert Dingess, GTMA President and CEO

GTMA President Rob Dingess welcomed the attendees of the GTMA Second Annual Meeting and Data Palooza Event. Mr. Dingess posited whether Congress will pass the transportation authorization bill soon. He recounted how MAP-21 was actually passed quickly. He ended by welcoming Deputy Secretary Victor Mendez and looked forward to his thoughts on new transportation technologies and data applications that present opportunities to improve the transportation network.

Peter Stephanos, Director, Transportation Performance Management Office, Office of Infrastructure, FHWA, USDOT

Director Stephanos noted that the USDOT is in the middle of an exciting time transitioning to performance-based decisions. There are now greater opportunities for States to support MPOs. All sessions within this two-day event will look at how to use data for smarter decision making. Better decisions should yield improved performance based outcomes in highway operations, transit operations, and public safety, to name a few areas. Four of the eleven performance or asset management rules are complete or out for comments. Through the Data Palooza and other stakeholder events, the FHWA is seeking input on identifying data gaps and how to best address these gaps.

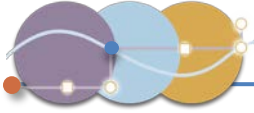**Victor Mendez, Acting Deputy Secretary of Transportation, USDOT**

Deputy Secretary Victor Mendez began his remarks by discussing the MAP-21 Performance Measures rulemakings process currently underway. He noted that all rules and rulemaking processes are cumbersome and takes time to implement. However, the process has a purpose and must be followed.

He noted that MAP-21 legislation is set to expire in September 2014. The $332 Billion surface transportation proposal known as GROW AMERICA seeks to provide a wise transportation investment with the taxpayer dollars. It includes bold approaches such as a major investment in rail. He noted that each modal agency needs a strategic focus and cannot be operated through a series of continuing resolutions. FHWA, for example, is running a $40 Billion business each year. Deputy Secretary Mendez commented that the dwindling Highway Trust Fund that is projected to be insolvent by as early as October 2014 presents a serious issue for moving the country forward. All of the current issues facing the USDOT will require higher standards for the funding uses which will raise the focus of transparency and accountability.

Moving to the topic of data, Mr. Mendez relayed that the value of data is well understood by Secretary Foxx. It is incorporated within the Secretary's seven focus areas. The USDOT is a data-driven organization. The new performance rules require that rationale programming and policy decisions are made from data. Data is now being seen in a different manner than two decades ago. Big data is changing the transportation industry on its own. How can cell phones be used to gather and disseminate transportation data, but avoid a major safety issue of distracted travel? Data will be necessary to meet the safety challenges. How can safety be used to enable people to drive safer? Deaths are down, but 33,000 people still died on our highways in 2012. A new emphasis is on bicycle and pedestrian deaths and injuries, which are now spiking.

In February, Secretary Foxx launched DOT's Data Innovation Challenge, a three-month quest to see what app developers could do to improve transportation by taking advantage of new access to multiple sources of transportation data. Deputy Secretary Mendez announced the results of DOT's challenge. While there were numerous submissions, the field was narrowed down to three winners:
- RideScout (*Transportation Access* category) - a mobile app that provides available transportation options in real time including transit, taxi, car share, bike share, parking, and walking directions;
- Choices & Voices (*Transportation Access* category) - a web-based, long range planning tool that educates users on the linkages between land use and transportation, the cost of maintenance, and the consequences of not investing in transportation; and

- FAST Dashboard (*Traffic and Congestion Management* category) - a congestion analytics dashboard that provides an easy-to-understand online user interface allowing the public and transportation professionals access to real-time and historical freeway monitoring and performance data.

B. Plenary Panel Session: The Future Role of Data in Transportation

Overview

An expert panel with a diverse set of transportation backgrounds discussed their perspectives on the future role of data in transportation. Moderated by Larry Copeland, USA Today's transportation reporter, panelists emphasized the significant influence new technologies have played in amassing more data and the opportunities that comes with applying this new information. While opportunities are plentiful, there are still many technical, cultural, institutional, and security issues that need to be addressed as data usage within the industry progresses. Data is affecting the research, operations, and performance of transportation. Speakers acknowledged that innovation is supported, but a shift should be made toward advocating for implementation of new techniques and tools.

Panel Information

Moderator: Larry Copeland, Transportation Reporter, USA Today
Moderator Copeland issued some thought provoking questions on data and transportation that was directed to the panelists. What is the future of data? There are a myriad of ways that data affects life. What roles does data play in the transportation system? What technologies are changing the transportation systems? Who are developing the National Data Goals? What will they be? How can data be used to direct transportation funding decisions?

Robert F. Skinner, Executive Director, Transportation Research Board (TRB)
Society is now in the midst of a revolution in using data and technologies. The Internet has provided new efficiencies for utilizing data for research and operations. We continue to move to produce faster and cheaper and even unimaginable – not so long ago – applications. To go along with these dramatic changes, the transportation industry is still looking to the Federal government for consistent data policies and a national leadership to plan for these new innovative opportunities. There are three broad areas of research on utilizing data in transportation:
1) Baseline Information – baseline data is hard to obtain. How can new technologies be used to assemble baseline information in a cost effective and robust manner?
2) Applying Data to Daily Work – data should play a central role in planning, designing, constructing, and operations of transportation systems. There is a need to devote more resources to data collection and applications. The IT revolution has put data on your desktop enabling new data manipulation and applications.
3) Research Using New Data – SHRP2 innovations have looked into how to make increasing number and sized databases available to more professionals. What new methods are available to analyze data and new datasets? We need research into how to overcome the cost and complexity of disseminating and sharing data. How to best manage data also requires research.

Michael Trentacoste, Associate Administrator, Research, Development, and Technology (Turner-Fairbanks Highway Research Center – TFHRC), FHWA, USDOT
The FHWA has started an initiative called the Research Data Exchange (RDE) that seeks to integrate and share mass amounts of data for the vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communications and use for mobility, safety, and environmental benefits. RDE 2.0 will integrate 11 different data sets to apply to the Connected Vehicle research and operations.

Pre-dating the RDE, the FHWA has been collecting information on pavement for 25 years and is beginning to expand the use of that data. The question being asked now is, how can that historic and current data be used to make decisions on a day-to-day basis and in long-term design concepts that can help the industry? The FHWA is developing a data hub, InfoPave, comprised of information from a variety of sources for agencies to access and use information in daily operations or long-term design.

TFHRC started an initiative called Virtual Data Access that is a framework and platform to provide agency and regional data and information for agencies to use for region-wide planning (the Kansas City region is the first Virtual Data Access pilot demonstration area). The challenge is how best to integrate data to get the best value from the wide variety of datasets available.

Richard McKinney, Chief Information Officer, USDOT
As CIO, Richard McKinney wanted to understand the cyber-security policies at USDOT and create a holistic security plan. During this process, he saw the opportunity to make the infrastructure lean, smart, and secure so that supporting data will be protected and cost-effective as it moves through the system. USDOT is ensuring money is invested wisely. A modal CIO working group was established and is working collaboratively on a cyber-security plan and helping determine the future path of the USDOT. The future of transportation data is open and a two-way conversation between user and provider. Connected vehicles and automated vehicles are accelerating the need for a two-way conversation on data.

Kenneth Leonard, Associate Administrator and Director, Intelligent Transportation Systems Joint Program Office (ITS JPO), USDOT
Data is critical at the DOT. We are in the middle of an information age; we are all generating data, sharing data, and applying data. There is now a great demand for Big Data. But as data sources increase, so do the challenges. We have the opportunity to build transportation connectivity which utilizes a number of ITS elements, all that demand data. ITS and connected vehicles yield an unprecedented amount of data. The RDE that AA Trentacoste spoke of showed the value in sharing "cleaned" data.

Connected vehicles and integrated corridor management (ICM) systems are two examples of the ways in which the ITS JPO are working towards building a robust system that will inform commuters on making better decisions and enable providers on being more responsive to problems. Two ICM program sites in Dallas and San Diego demonstrate how multi-modal and multi-agency sharing of data can provide both operators and commuters with real-time and predictable traveler information for highways, transit and arterials. Data applications are focusing first and foremost on safety, while also examining applications for transit, mobility, freight, and environmental.

Rolf Schmitt, Deputy Director, Bureau of Transportation Statistics (BTS), USDOT
Transportation is about moving people and stuff, but we need a greater understanding of what and who are moving and why. We have had a number of performance requirements that have put increasing demands on what type of information needs to be collected. New and exciting data sources are available to us, but these sources offer huge amounts of data that address narrow questions. Deputy Director Schmitt advanced that smart utility meters at each house could be linked to transportation providers to better indicate real-time and systemwide travel demands. A big challenge with big data is how do we clean it without destroying the outliers and unexpected real information? Developing more intelligent ways of cleaning and integrating the data is important for future applications and improving performance. Additionally, the visualization of data is essential to explaining the performance story of the transportation system. Maps are incredibly powerful in communicating the story of transportation.

Plenary Panel Session Q&A

The Q&A session following the panelists' speeches covered a variety of topics prompted by the panelists' diverse areas of expertise. The audience took advantage of the opportunity to ask questions to the array of transportation leaders. Questions covered everything from funding to private versus public sector roles to future USDOT data management plans.

- **Audience**: How has data changed during your time at TRB?
  - **Robert Skinner**: Personal computer, Internet, and instant accessibility to information has made a huge difference in efficiency of operations and research.
- **Moderator**: Transportation as an information provider, Richard, can you elaborate on that?

- o **Richard McKinney**: Two-way provision, internally for analyses and project planning for smarter investment and externally for effective commuting, safety, environmental, economic choice. Citizens can provide transportation agencies information and government can also provide them data. The City of Boston's Street Bump application for tracking road surface smoothness is an example of citizens sharing data to improve public operations.

- **Audience**: Congress asking for us to provide more information but cutting funding, they don't understand the value of data, why is there a gap between the importance of data and the understanding both public and policy-related of the value of data?
  - o **Michael Trentacoste**: Data is part of research; however, data collection is not as critical for innovation for implementation, it is important for baseline information that goes towards analysis. There is a disconnect between the innovation and legislation to make the program work.
  - o **Robert Skinner**: Major problem is that we have a complicated story to tell. We do have advocates that can only communicate so much and there are competing priorities to the various research programs. UTCs are a capacity building program versus other structured research transportation programs trying to achieve an objective. Each research program has different niches and the need to secure funding. Research, in general, needs a clear cut message that is easy to understand. The technology revolution within transportation needs to link the public sector research to the private sector benefits.
  - o **Kenneth Leonard**: Timeframe is an issue in that research payoff is very long, this is part of the research communication problem. People look for the low-hanging fruit (early benefits), but research needs deep roots to build the environment that will show results. Benefits are only derived if research continues to be funded and after a ten to twenty year horizon.

- **Audience**: Basic benchmarking data, are we spending enough time enhancing that dataset that we often overlook as opposed to the "exciting" stuff?
  - o **Rolf Schmitt**: Made enormous strides and figuring out how to integrate new data sources into the traditional datasets. Interesting balancing game overtime, but in much better shape than years ago in some areas such as long distance freight, local metropolitan area passenger travel information. Need to know basic information that serves multiple purposes and trying to put a value on it is tough because it supports everything; it is the data infrastructure and could always use more money but trying to use the best we can.

- **Audience**: Differences between USDOT big data versus other third party data sources, are there any thoughts on incorporating this data into the existing baseline data?
  - o **Michael Trentacoste**: Local and state agencies are now buying data from third parties rather than collecting their own. From the Federal perspective, we would like to know if this data is compatible with any Federal datasets. Was the data collection methodology sound? How is the data cleaned and quality controlled? New transportation planning models, such as those now available through Google, are being developed by third parties and public sector needs to advance those that provide quality data and information.

- **Audience**: No plan to maintain data once collection is completed. The inevitability in using stale data on old sites seems to be a large hole in how we plan and program this transportation research. Do we need a new program that includes data maintenance and long-term visions?
  - o **Robert Skinner**: In regards to SHRP2, there is a 5-year plan for the data. Invested $60 million, so there is an interest in maintaining this data. SHRP2 data is much more extensive, there does need to be some thoughtful effort at how to get started managing data, a strong and flexible research program that can confront these questions and help not just researchers but operating agencies that are collecting and using the data.
  - o **Richard McKinney**: Data should not be a snapshot, but should constantly be kept fresh and updated. My office needs to think about storage and storage requirements moving forward. Data infrastructure needs to keep up to ensure that all the real-time and big data efforts remain current.
  - o **Rolf Schmitt**: Even data snapshots are challenging to maintain. BTS manages the National Transportation Library, one place to find USDOT reports. Technology is changing ways to archive data with hardware and software. BTS needs to make sure these data programs remain in an accessible readable format to be useful as historic information.

- **Audience**: Do you think we will ever get to the point where we treat data storage, analytics, etc. as an integral part of the system that it will receive a stable funding source?
  - o **Richard McKinney**: USDOT aggregates a lot of data from the state and local levels. Everybody collects data differently so it is difficult when trying to aggregate the data. Let's sit down with industry stakeholders to agree on data standards to simplify integration and aggregation moving forward. USDOT, in its role as aggregator, not developer or collector, should convene this conversation now, at the cusp of a data tools revolution, and before we put too many resources into it.
  - o **Rob Dingess**: GTMA is willing to bring people to the table and have that conversation, just not sure what it will take to have that happen.
- **Audience**: Role of government is to make data available for the private sector to innovate. How do you see the government and the private sector working together and where do you draw the line between the respective roles?
  - o **Michael Trentacoste**: The White House Office of Science and Technology is examining public research investments. Agencies need to look at how data will be made available to the public in the future. How can data from research be made accessible both internally among the USDOT administrations, but also externally? I think USDOT will be more than just the data providers because the mission is to create innovation and research results, including collecting and utilizing data. The more USDOT data is made available, the more the sponsoring research can added value out of the data. USDOT needs to get its head around the data that they currently utilize and then get it out.
  - o **Rolf Schmitt**: At BTS, any data that can be released to the public is released, but with lots of restrictions. If we bought the data then BTS cannot disseminate it. There are privacy and security issues that restrict data availability. If you are not an experienced data user, there are issues that are not evident. Not all the data is the same unless you have an eye to examine it. BTS seeks to explain to the user how to use the data. BTS is interested in seeing creativity from the community on how to use its data. Once you give someone access to a dataset, how do you help a user acquire the sophistication to use it? BTS hopes the public uses the data effectively.
- **Audience**: Kenneth, what is the biggest difficulty with moving forward with connected vehicles and what is the timeframe?
  - o **Kenneth Leonard**: On February 3rd, 2014, NHTSA released an announcement on proposed connected vehicle regulation under this administration. Every production vehicle by 2020, not just in the U.S. but globally, will be required to have dedicated short-range communication (DSRC) technology so they can communicate basic safety messages. Automated vehicles are moving very quickly on its own with R&D from universities and private companies, such as Google. There are different levels of automation, not completely self-driving cars yet. FHWA (and ITS JPO) will soon be issuing infrastructure guidance. Approximately five to ten percent of the FHWA budget will be used for connected vehicle or ITS infrastructure. These are investment decision made at the state and local levels. Many states are now willing to get out front with connected vehicle infrastructure, so the FHWA wants to get guidance out soon to help states make those decisions and to ensure interoperability of the systems and the data being collected and used.
- **Audience**: What is the vision for data in transportation? What is the most immediate need the private sector can assist with?
  - o **Richard McKinney**: Modal CIOs met and conversed about data in May. The modal CIOs tend to look at data within their individual silo. Who is having the overarching view of transportation within the department? What are the multiple uses of the USDOT databases? CIO McKinney intends to appoint a chief data officer within USDOT, modal CIOs all agreed that someone needs to sit at the enterprise level to look and coordinate data within the department. There is a need for authoritative database of roads and addresses and street names across the country. For example, with NextGen911, there is a need to geo-locate people in trouble. USDOT to begin to take enterprise view.

C. Standards for Sharing Transportation Data Roadmap Workshop

Overview

Data standards are crucial for exchanging and analyzing information across all levels of industry. This brings many questions to the forefront – how do you standardize data, what role should the Federal government play in standardization, what have others already done to begin the standardization process, and what format will be selected and why? Although all of these questions were not answered at Data Palooza, the event provided a forum for attendees to further discuss these problems and potential solutions. Practitioners presented on various efforts of integrated data, allowing attendees to discover noteworthy practices and ask questions about lessons learned.

Several USDOT data sharing efforts were presented at Data Palooza. USDOT agencies such as National Highway Transportation Safety Association (NHTSA), Pipelines and Hazardous Materials Safety Administration (PHMSA), National Transportation Library (NTL), and FHWA's Office of Highway Policy Information informed the audience of their data systems that are either currently or soon to be available to the public. As public agencies, it was noted that there is an amount of responsibility to ensuring data is shared with the public, but there is also an awareness that public agencies need to protect sensitive information, especially regarding privacy. Furthermore, data is often collected by private sector companies and purchased by public entities. Therefore, the data cannot always be made available publicly. Some presenters acknowledged difficulties with obtaining data and getting agencies to agree to provide their data. Despite some of these challenges, agencies have built or are building new, automated systems that will better assist informing policy decisions, expanding research, and evaluating performance. In addition to some raw data, public organizations are also publishing performance reports and dashboards for the public to view and analyze.

Data management difficulties were a topic brought up several times during the two Standards for Sharing Transportation Data workshops. One reason cited for hesitancy on data sharing is that agencies are nervous their data is bad and will be scrutinized by leadership or the public. Agencies do not have the resources to consistently maintain the large amount of data collected. Additionally, most organizations do not have a data management plan or data systems that can handle the vast quantities of data gathered. More attention within agencies should be diverted to data management to improve the quality of data and therefore enhance the applications of data. A conversation on data governance presented a solution for agencies to internally develop rules and definitions for improved and sustained data management.

Although data sharing is important, the lack of data standards in the industry diminishes the value of data that is publically available. Unstandardized data means that companies and public agencies code and process data differently, making it a challenge to integrate different datasets of different formats. While data may be available to the public, some data sources require particular computer programs to open and read the data. The absence of data standards is a barrier to the industry in furthering data applications and analysis. Some presenters and attendees mentioned it is difficult to impose data standards as data is needed and used for many different purposes. A continued dialogue on baseline transportation data standards will be useful in ensuring collaboration and generating consensus. It is important to lay the groundwork now on creating widely-accepted and used data standards before the industry has gone too far in individual directions.

C.1. Standards for Sharing Transportation Data Roadmap Workshop Part 1

Panel Presentation Summary

Moderator: Patrick Son, ITS America

National Automotive Sampling System (NASS) – A Modernization Projects
John Brophy, Chief of Crash Investigation Division, NHTSA, USDOT

- NHTSA is a data-driven agency that uses data for basic research, rulemaking, developing, and evaluating programs, allocating grants, and setting safety targets
- Crash reports are not standardized across states; field workers manually fill out reports of crashes
- Congress appropriated funds in 2012 to modernize the system and will be implemented in 2016
    - Paperless, less manual labor, less errors, more widely accessible

PHMSA Data Initiatives
Yolanda Braxton, Program Analyst, PHMSA

- HAZMAT Intelligence Portal (HIP) established in 2008 to share incident and accident data across government agencies and private partners – modal, environmental, security
    - Gathers data and enters into database in uniformed format
- HIP feeds performance reports and dashboards
- Unified enterprise information helps make operational decisions and identify events that pose greater risks

Data Governance
Zeke Bishop, Enterprise Architect, Oracle

- Data governance makes sure data is managed properly and acts as an internal auditing function for information, authority combined with policy to ensure proper management of collected data
- Treat data as an asset
- Data governance requires a behavior change effort that fits into the business operating model.
- Need early buy-in from leadership then can define and socialize data governance into organization to build a mature program overtime.
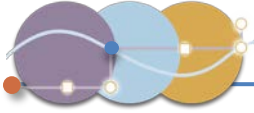
Standards for Sharing Part 1 Q&A

During this session, questions were asked following each presentation. Most of the questions asked during this session focused more on the technical side of data collection and dissemination as it pertained to each example.

- **Audience**: Vehicle technology has also improved. There is other information that can be captured through the vehicle infrastructure, is there any thought in revising EDR standard and increasing the scope of EDR data into the databases NHTSA maintains?
    - o **John Brophy**: NHTSA does not write EDR standards, but it helps advise SAE and is in the middle of a comment phase. NHTSA collects and codes it, but there are no standards for it from the manufacturers. Need to wait on the rulemaking and SAE. The new system will be more adaptable.
- **Audience**: What are we going to do with the risk profile to ensure that the specific mode of packaging, travel, etc., are mitigated?
    - o **Yolanda Braxton**: In the works, adding in the packaging data to get the better feel associated with commodities and packaging.
- **Audience**: Are you deploying the risk profiles to inspectors and field workers?
    - o **Yolanda Braxton**: Yes, available on mobile device. Also some of the dashboards are available to the public. Worked with industry to get feedback on what they would like to see and what information is useful to them – examined website hits and data requests
- **Audience**: Do pipelines fall under PHMSA's domain?
    - o **Yolanda Braxton**: Yes, I work on Hazmat side, but pipeline has similar tool.
- **Audience**: Your data had 1 million pieces that did not match – is there an effort?
    - o **Yolanda Braxton**: There is a team going through the manual labor and we have a process in place. It is labor-intensive
- **Moderato**r: You bring in data from various sources and integrate, how effective has that been? Do you need people to change the way you are submitting the data?
    - o **Yolanda Braxton**: Not too big for us, we have had to ask partners to re-fresh and add data elements for new perspectives. Ask partners if they want to add new requirements to the dashboard or see anything new. We are constantly adding new data and changing the dashboards.
    - o **John Brophy**: The crash we investigate, we collect data with field workers. States that send in the data, they need to fill out uniformly through particularly data submission platform, but they still have varying levels of data because we only require a minimum amount of information.

C.2. Standards for Sharing Transportation Data Roadmap Workshop Part 2

Panel Presentation Summary

Moderator, Dag Gogue, CEO, Transit Labs

Develop a Transportation Clearinghouse to Handle Data, Data Experts, and Transportation Research
Dag Gogue, CEO, Transit Labs

- Demand is increasing and funding is decreasing for public transportation, need to find ways to minimize inefficiencies in the system
- Encouraging transit agencies to use Census data to draw comparisons with statistically similar peer groups
- Need efficient way to communicate with decision-makers where the failures in routes are occurring and why, Transit Labs process uses:
    - Open source technologies(street map layers) and data sources
    - NTD, APCs, GTFS, Census
    - Goal is to show flow of goods and people in one map layer, including transit in map of entire transportation system as a whole

MAP-21 and the All Roads Network: Geospatial Data for Transportation and Beyond
Chris Allen, Chief, Office of Highway Policy Information, FHWA, USDOT

- Four siloed data systems trying to get integrated to mitigate labor and resources
- Kicking off data governance initiative in FHWA, currently working on the definition
- All Roads Network: guidance has not been sent out yet, but states should submit the data as soon as they are able to send it

Improving Transportation Operations and Management through Data Integration
Michael L. Pack, Director, CATT Laboratory, University of Maryland

- Washington, DC has many transportation entities and traffic management centers that fall under different jurisdictions
    - Each organization uses different data vendors, making data integration and coordination difficult
    - Standards were challenging because it is costly and time consuming
- CATT Laboratory's mission was to facilitate operations coordination to reduce costs for vendors, agencies, and encourage innovation
    - Archive and make data available to the public
    - Creation of RITIS as a data warehouse in Washington, DC
- Problems in convincing the agencies to give the data to the warehouse and share with researchers
    - Scared of submitting bad and outdated data
- Provided agencies with a suite of tools that made it easy for them to understand their own data ; web-based, understandable, easily accessible format

President's Open Data Initiative
Waldo Jaquith, U.S. Open Data Institute

- Need to develop data standards now to lay the groundwork for the future
- White House's Open Data Initiative: striving for open data policy and therefore open and transparent government
  - Agencies unwilling to share data because intimidated, fear of mistakes and the consequences of using bad data
  - Need a culture change to encourage publicizing first small then larger datasets even with mistakes
- Producing software for agencies to easily share, convert, and upload data
- Policy could end after President Obama's term because it is an executive order

What's on the BTS Data Clearinghouse and Why?
Mary E. Moulton, Digital Librarian, National Transportation Library

- National Transportation Library (NTL) is under the Bureau of Transportation Statistics to house reports and datasets, acts as central repository for USDOT research
- NTL Data Clearinghouse: implement national transportation data archive as part of MAP-21
  - Data life cycle – helps format data management plan for agencies, working on preservation guidelines for version control and develop version control
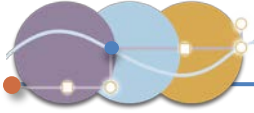 - Data repository for State DOTs because they typically do not have one

Standards for Sharing Part 2 Q&A

This Q&A session focused on the challenges presenters faced in developing their data clearinghouses. Presenters discussed their experiences trying to collect data and the resistance they faced. Data quality issues are prevalent in the industry; more resources and a culture change could help in improving and disseminating data. Complications with data standards were also discussed. Some suggestions to creating baseline data standards focused on having a collaborative and iterative process that evolves with the times and with best practices.

- **Moderator**: Waldo, what challenges have you faced in trying to create a <u>data clearinghouse</u>?
  - **Waldo Jaquith**: Getting easier to open data because tools are now developed for conversion to share data properly. What used to take a lot of time is now much easier than it used to be. There is no competition so the software it slower to develop. The crucial datasets are also managed by one person who and doesn't build the system from scratch, but builds upon and develops the one that exists.
- **Moderator**: Creating a clearinghouse, although technology is not the biggest problem, but a clearinghouse necessitates a certain technology standard. How did Michael go about <u>standardizing various datasets</u>? And can it be expanded to other datasets?
  - **Michael Pack**: There are no standards in the domain that I employ. Every department thinks they are unique so something is slightly different. General policy when acquiring data is not to force them to abide by any standard that we developed because they don't have the budget to do that. As it becomes easier for agencies to convert their data it will make it all easier to integrate. We store data we receive twice – the original way we received it and then after we manipulated it into our own format to aggregate analysis and performance measures. We store the data in our system for ourselves, if you subscribe to our data feeds you see the standards we use. Transit data already has a group that went through this process, GTFS, so we just adopted that standard.
  - **Moderator**: Important to look at what everyone has or is working on and see the formats that worked best for more established organizations to decide which is worthwhile from duplicating.

- o **Keith Gates, Federal**: There are different customers out there and smaller users need just what they want, but others want more robust systems. Use the data for different reasons so it is hard to create standards to satisfy all those needs. Being a data provider is very labor intensive.
- **Moderator**: Who are the different data users and how do we <u>aggregate data</u> and render it in powerful visual ways? Mary, as a librarian you have a wide-range of customers you service, can you talk about this?
  - o **Mary Moulton**: Librarians share everything and they have a global standard developed at the Library of Congress. The search is interoperable because records are shared with each other forever. Important to have events like this where the different stakeholders talk to each other and maintain a dialogue. It is easy when you don't work with information every day to envision how it fits in with other perspectives and audiences. It is interesting to see people from different disciplines come together and agree on standards. Be aware of places to share information, many places on the Internet have data available for sharing. I always talk about metadata because I think it's important to have your information discovered by search engines and other initiatives. Lots of initiatives all over the world that are working to advance this. You can submit your metadata scheme to the International Digital Curation Center (IDCC) and then other people can discover it and harmonize with other efforts.
- **Audience**: It was interesting to hear Michael and Waldo talk about the <u>difficulties in acquiring data</u>. What if you convert it to a standard and push it back out through IDCC and get comments on it. Why did you not talk about the University challenges and programs? How does all that fit together?
  - o **Moderator**: Always important to talk among the owners, processors, and users or data. I like the notion of an iterative process of converting, publish, and getting feedback.
- **Audience**: Wanted to clarify, it seems like Michael said more datasets they got the more users they got and Waldo said the more data they put out there the fewer users they are getting. I don't see the meaning behind getting data out there if there are <u>few users</u>.
  - o **Waldo Jaquith**: Look at the quality of use that the data gets rather than the number of downloads or hits.
  - o **Michael Pack**: The 3,500 users is a small number of users when you think about popular data applications, but I know those users are really using and digging into the data. People are driven to some datasets because of the tools we developed rather than the data we put out there.
  - o **Moderator**: How are you tracking usage of the data?
  - o **Michael Pack**: We can only track the number of downloads
- **Audience**: Is there any <u>feedback mechanism</u> for the users of your data, Michael?
  - o **Michael Pack**: We created user groups and forums. We have quarterly webinars and listening sessions where they can discuss what they do and don't like.
- **Moderator**: Multiple datasets that have different organizations feeding the data. How do you <u>prioritize the data visualization</u> to make sure the users get what they wanted?
  - o **Chris Allen**: We haven't started producing the visualization, but that decision would be up to Justin depending on what people are requesting. In FHWA we've been making decisions for a long time, trying to get people to understand that we have the data but how would you like it processed? FHWA is having listening sessions for internal and external users to ask them what is good, where are gaps, what do you need?
  - o **Moderator**: What is the make-up of that user group?
  - o **Chris Allen**: Internal FHWA leadership and staff in various FHWA-designated disciplines who are providing data to decision makers. TPM office, Field Service offices, etc. trying to reach everyone
- **Audience**: Have not heard a lot about <u>privacy concerns</u>, is it not a big deal or do we not want to touch it? What is the role of privacy and what are some solutions?
  - o **Michael Pack**: 99% of data that agencies put out is not sensitive and does not release any identifying information. Sometimes the consequences and sensitivity of data can be over-inflated.
  - o **Moderator**: On transit side, there is some sensitive information on origin destination data and cell phone tracking, even though the data can be scrubbed and be made more private, there are ways to identify people. These are concerns that need to be discussed especially in developing a data clearinghouse.

- **Audience**: Government-created data has been discussed, there is some very useful data in standard formats non-government, such as open street maps. Have your organizations dealt with any <u>non-government gathered/processed data</u>?
    - o **Moderator**: Transit Lab uses open street map and can layer government data on it. These discussions tend to be focused around government and research universities, Transit Lab was excited to come here and talk as a start-up and meet needs beyond the public sector.
    - o **Chris Allen**: FHWA is exploring open street map and trying to understand how it fits in and it looks like a viable opportunity.
    - o **Waldo Jaquith**: U.S. Open Data Initiative works with private sector as well. Stock markets publish data constantly and these are private companies that know the value of that data is only when it's open. Google Maps will guide me home tonight because it is gathering data from data sources that they bought, but others cannot download. Private business can do whatever they want with their data, they are not accountable publically. They can move faster, but make more mistakes. There is no apple to apple comparison of public versus private data.
    - o **Michael Pack**: Quality issues go both ways, everybody has data quality issues. The companies complaining about bad data aren't releasing theirs for the same reason.
    - o **Shashi Ranganath, Private**: Google Maps is so widely used because of many data points. Google's vision is to have a globally managed data system that is organized and useable.
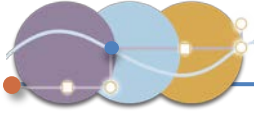
D. Keynote Address: Data Integration – Breaking Down Barriers to Change

Overview

John Thomas of the Utah Department of Transportation (UDOT) discussed the agency's experience in developing their integrated data sharing platform, U-PLAN. In the early 2000s, the agency, at the request of their Executive Director, created a custom application to display traffic demand geospatial and safety data that was available to a wide range of people. Because it was a custom application developed by UDOT, the agency soon started feeling overwhelmed trying to maintain and process the increasing number of layers. UDOT hired ESRI, a private firm, to assist in further development of the tool. The web-based tool is now on an ArcGIS platform and universities were able to create mobile applications from the data. It is a collaborative information site where users can upload their data and create searchable layers. UDOT used Mandli data, but it was too robust for their data systems. To face this problem, UDOT with the assistance of Oracle and other data partners created a data roadmap and updated their data warehouses into the new system. UDOT now has easily accessible and searchable data that has improved their decision making and project delivery. The success of UPLAN demonstrates the importance of strong leadership, close collaboration with both public and private partners, and that data management planning has helped advance the agency's effectiveness and efficiency.
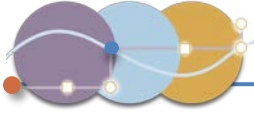
Keynote Address Q&A

As audience members digested the success story of UDOT, the questions focused mostly on understanding the thought processes and extent of the UPLAN program.

- **Audience**: Is this giving your folks additional confidence as they use this information for projects and are other teams getting excited about using these integrated datasets?
  - o **John Thomas**: Now that it's being used, they are getting excited to use it. It is pretty fresh information. FHWA does environmental review, but for rail it was a different group and they felt more comfortable to partner.
- **Audience**: Data integration being used in safety analysis?
  - o **John Thomas**: The transparency of this process to have a self-serve mapping and clear data accessibility allows people to make decisions in a way to be useful.
- **Audience**: Is there an overarching argument that was used in Utah to try to adapt these methods?
  - o **John Thomas**: IT discussions and leadership discussions. The message is there at the grassroots level, we didn't wait for the top-down approach, but now going forward the statewide top-down approach is happening. Be collaborative and open, UDOT does not know everything and still learning, but open to sharing their experiences with technologies that are available to others.
- **Audience**: You must collect some reusable data, is there a common infrastructure being used across UDOT?
  - o **John Thomas**: Philosophy of collect it once, use it multiple times. Partnerships, such as Oracle, helped us understand how to modernize and use and develop data connections

E. Analyzing Integrated Data Workshop

Overview

Integrated data that is being collected and processed today gives agencies the opportunity to tell a more comprehensive and complete story of the transportation system. Presenters during the two Analyzing Integrated Data workshops focused on new analysis and integration techniques being used and developed. New technologies and non-traditional data sources are collecting more and more data for agencies to apply to the day-to-day operations and overall performance of the transportation system. It is important that practitioners discuss these opportunities and collaborate where appropriate to gain and communicate the most value out of the data.

USDOT Initiatives

Part 1 of Analyzing Integrated Data focused on USDOT Initiatives. USDOT agencies are working to integrate many data sets to provide various applications. From daily operations to program evaluation to data visualization, Federal transportation agencies are trying to process data efficiently for a thorough analysis of the transportation system that is understandable at all levels of expertise. Each agency collaborates with public and private partners to gather data. Although there is some amount of labor and time required to integrate the various datasets, some of the agencies see progress being made in streamlining these efforts. Currently, the agencies process data based on the request made by users or based on their internal needs. It was noted that not only is it important for the public to be able to access and understand the data, but also policymakers. Practitioners need to be able to communicate the needs of the transportation system to policymakers effectively. Greater advancement in this area will be made if the dialogue continues on data integration and analysis techniques.

Analyzing Research Methods

The USDOT Initiatives session was followed by a workshop on Analytical Research Methods. The presenters discussed the idea and applications of big data and non-traditional data sources. Due to limited resources, it is important for the industry to look for alternative methods of collecting and analyzing data. Social media, smart phones, and other technologies provide mass amounts of data in fields that often required more in-depth survey or research methods. Big data has many benefits, but presenters also reminded the audience that big data does present new problems. Big data has the potential to fill data gaps; however upkeep, privacy, and statistical significance of big data are two large issues. These new data sources can inform travel behavior and travel demand models more than ever before, but there are hurdles to overcome. Further research and discussions about the integration of big data with traditional data sources will increase its feasibility and reliability.

E.1. Analyzing Integrated Data Part 1: USDOT Initiatives

Panel Presentation Summary

Moderator: Maria Chau, FHWA Division Office - NY

The National Transportation Atlas
Rolf Schmitt, Deputy Director, BTS, USDOT

- Atlas displays data once it was pulled together
  - Applications include: transportation facilities map, thematic maps, deficient bridges web applications, state transportation facts and figures application
  - National Transportation Atlas web application
- Application is in beta phase
  - Drew layers from a variety of sources: Census, EPA, etc.
  - Select features and display summary statistics
  - Some quality issues, let BTS know of any problems to fix

FHWA's Integrated Data Analysis Efforts and Plan
Tianjia Tang, Chief Travel Monitoring and Surveys Division, Office of Highway Policy Information, FHWA, USDOT

- Objective program is to provide data in a manner that is complete, comprehensive, and understandable
  - Other program offices, organizations, and businesses are responsible for performing analysis on data for their own needs and scenarios
  - Data is available for others to use, strives for easy access and timely scheduled publication
    - Data is published in a wide range of formats
  - Locational and temporally linked data

FMCSA Data Sources and Uses
Steve Smith, Director, Office of Analysis, Research, and Technology, FMCSA, USDOT

- FMCSA collects data form multiple sources
  - 500,000 commercial motor carriers spread across the country
  - Receives DOT number that is a unique identifier to track activity
  - Summarized data is available to the public, some information for agency use only
- Applications
  - Safety Measurement System that quantifies performance in 6 areas
  - Annual report of progress and performance
  - Develop priorities and learn to optimize resources
  - SaferBus: Integrated databases for mobile application
    - Customer for a bus company can see detailed information about safety performance so consumers can make informed decisions

FHWA's Vehicle Probe Data

Nicole Katsikides, Transportation Specialist, Office of Freight Management and Operations, FHWA, USDOT

- 600,000 trucks embedded with technology to capture real-time data
- Freight performance measure based on truck probe data
    - Monthly and quarterly performance reports, supports corridor analysis, and as hoc analyses (weather event, emergency closure)
    - Access is primarily for FHWA, but can support external analysis
    - Data not integrated with other FHWA data
    - Customized software applications used to process for relational and geospatial systems
- National performance measures dataset
    - Assist states with performance measures and target setting – archived travel time data for passenger, freight, and all traffic
    - Available to FHWA, Federal partners, States, and MPOs
- Expansive dataset, but it is cleaned so no information on type of truck or commodity being carried

Analyzing Integrated Data Part 1 Q&A

The questions asked following the presentations during Analyzing Integration Data Part 1 covered both technical and big picture issues. Presenters discussed everything from specific calculations used to their experiences collaborating with many different entities.

- **Audience**: Does FMSCA data capture property damage?
    - **Steve Smith**: Yes, if there is a tollway but not otherwise because then it would not be reported to FMSCA
    - **Audience**: States can require crash forms to be filled out if there are a certain amount of dollars or more damaged.
    - **Steve Smith**: FMSCA has been putting together guidance on how to report crash data.
- **Audience**: Types of trucks are not distinguishable, so you cannot derive duty cycles, average speeds or anything?
    - **Nicole Katsikides**: We can do that if we want to look at a specific truck. Access to granular data might provide someone information about a competitor's freight movements and therefore their business practices. The data comes with a unique identifier so we can track that truck overall, but not the fleet size or type of truck or commodity. We have done preliminary service hours analysis.
- **Audience**: How do you compute vehicle miles traveled for passenger cars with the data you collect?
    - **Tianjia Tang**: Good question. We have HPMS data and that is submitted by the state annually other than local and rural minor roads. All other roadway networks are reported with segments lengths and AADT (annual average daily traffic). Multiply segment length by AADT you get VMT. The States works with local MPOs or county for the summary data for the local and rural roads and then we can add those together to get VMT. We are dealing a lot with future VMT to see what we will deal with in 20 years? That is the model result.
    - **Audience**: How accurate are the VMT numbers? Do you do any analysis on the AADT numbers reported?
    - **Tianjia Tang**: There is always room for improvement for data quality. From a statistically significant standpoint, we are very confident with the AADT and VMT data. There are very specific sampling procedures and every State exceeds the sampling accuracy requirement. We do realize there is an issue there, especially with permanent traffic counting stations. If you could move them around and as technology gets cheaper, there is potential to widen the sampling.
- **Audience**: Integrating all this new data, facilitation of data sharing must occur. Are any of you familiar with sharing data as a service to other people who use it, and is there any vision of a USDOT enterprise data sharing service?
    - **Rolf Schmitt**: data.gov or dot.gov/data. The ArcGIS applications I showed depend on the services of other agencies sharing their data. In the GIS world there is a lot of data sharing occurring right now. USDOT is

making sure to inventory their datasets and then making as many public as possible. There are proprietary issues and privacy issues so there are confidential datasets. At the plenary, the CIO mentioned the address issue. There are two government organizations that know where your address is, Census and Post Office. It is protected under privacy conditions that apply to those organizations. USDOT is working with States on the all public roads network for a complete and integrated highway network. Bringing integrated datasets together is not that easy. Someone asked what the grand vision for data is? For me, it is Star Trek – ask the computer merge this with that. Just to create the tables for the NTS, fatalities by mode is a lot of work to deal with inconsistencies of definitions and overlap. As we improve access to the data behind the tables, we need to figure out how to communicate to the users how to use the data effectively. More eyes on the data also mean people can discover any problems or quality issues with the data.

- **Moderator**: Each of you has some publication for the public, how often do you need to <u>tailor data</u> for some specialized need and some variation of an audience?
  - o **Tianjia Tang**: I do it on a daily basis, legislatures ask for trends and scenarios in addition to our schedules annual and quarterly publications.
  - o **Steve Smith**: We have an internal data request system where we have over 700 requests each year from congressional staff requests or State partners or businesses. Our publications we put out, but we work the databases each and every day. Sometimes we don't have to completely reconstruct them, but sometimes we do.
  - o **Nicole Katsikides**: We look at the data differently depending on the type of request we get. Freight Analysis Framework online gets a lot of hits and we worked with BTS to get that out there for people to use and gather data themselves. The FPM data requires different analyses and excited to see questions we get because we can learn something new. We try to be as accommodating as we can.
  - o **Rolf Schmitt**: We go to the complete extreme to make ourselves available for data requests: answer@dot.gov and have a 1-800 number.
- **Moderator**: Can you tell us more about your <u>collaboration with other agencies</u> on data sharing? What is it like within USDOT and externally?
  - o **Rolf Schmitt**: The biggest data collection we do is a commodity flow survey with the Census. There are lots of internal discussions about data collection methods. BTS is not limited to the U.S.; Canada and Mexico share information with us for trans-border issues. There is a lot of collaboration that goes on with the other statistical agencies and other countries.
  - o **Tianjia Tang**: FHWA with collaborate within FHWA, State DOTs, and MPOs. A lot of their information is collected by private sector contractors. If we need technical expertise we seek it with BTS. DOE and HUD work closely with us on data collection.
  - o **Steve Smith**: We work internally with USDOT with NHTSA, FHWA VMT information, NTSB, Department of Homeland Security.
  - o **Nicole Katsikides**: We do things for freight with USACE, their AIS data for tracking vessels and trying to merge that with ours. Major workshop with TRB freight fluidity and Canada looks at end to end supply chain movements and want to apply that here in the U.S. to see how it can worth throughout North America. Department of Commerce, Department of Agriculture to try to figure out how to fit data together or how we can use it. Multi-modal performance measure on the flow of commodities.
- **Audience**: Can you tell us more about <u>Canada's freight fluidity model</u>?
  - o **Nicole Katsikides**: Canada came up with this supply chain analysis that is very good, you can Google it to learn more. I can send you presentations they made if you give me your email address. They are looking at the water movements into their ports, transfer on rail to truck. They use probe and rail data as well as marine data pieced together that they can look for economic development and system performance opportunities along their corridors.

E.2. Analyzing Integrated Data Part 2: Analytical Research Methods

Panel Presentation Summary

Moderator: Joe Moyer, Turner-Fairbanks Highway Research Center, FHWA, USDOT

National Household Travel Survey: Data Program
Adella Santos, NHTS Program Manager, FHWA, USDOT

- NHTS is conducted every 5-7 years, authoritative source on American public travel
  - Funded by FHWA (25,000 sample size) and State and MPOs can purchase larger sample sizes
- Challenges to current methodological and analytical approach:
  - Complex survey design
  - Budget constraints
  - Probabilistic-based survey required by OMB to make scientifically-justified conclusions
- Exploratory research to figure out better design for high-quality data despite constraints
  - Smartphones and Facebook opt-in application when going on long distance trips
  - Data fusion with non-probability sources
  - How do you make it easy and attractive to use?
  - Created algorithm to understand trip purpose when questions were skipped

Big Data for Travel Demand Modeling
Laurie Schintler, Associate Professor, School of Public Policy, George Mason University

- Big data: high volume, velocity, and variety; low veracity
  - Many sources – online, physical, crowdsourcing, digital
  - Structured and unstructured data
  - Can be incomplete, biased, anomalies/outliers – time consuming and expensive to process
- Changing big data landscape for transportation because many different applications
  - Travel demand analysis for planning, policy, management, and solutions
  - Traditional sources of data can integrate new big data
  - Big data is very detailed and fine grain
- Challenges: computational complexity, statistical issues, privacy

Big Data, Travel Behavior, and Transportation Demand Management
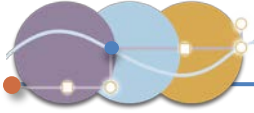Stephen Crim, Director of Research, Mobility Lab

- Potential of big data to fill gaps previously not understood by the transit industry
  - Create profiles for individuals based on big data to target messages and outreach to particular group to influence mode choice
    - What personal characteristics are related to transportation and mode choice?
    - GPS, spending habits, social media
  - Assist with program evaluation – effectiveness of outreach and return on investment
- Challenges: survey fatigue, self-report bias, sampling bias, costs, privacy

Analyzing Integrated Data Part 2 Q&A

The use of non-traditional data sources sparked many questions after the presentations. The audience seemed to be seeking greater clarification on the extent and application of big data and non-traditional sources being used today. Discussion on the future application of these new data sources is important for USDOT to understand and outline the role of big data in transportation planning, operations, and analysis.

- **Moderator**: If these sources of big data are able to provide insight into habits of the public, are we still going to need travel demand models?
  - **Laurie Schintler**: One argument about big data is that we are collecting everything about everything, so we have all the data we need. Do we need travel demand models that rely on underlying theory? I argue that they are complementary to each other, we need the data and the theory.
  - **Stephen Crim**: How do we get at what we could be? I'm concerned that with the proliferation of data we are continuing to fuel a mindset of understanding the status quo and optimizing that without thinking about new ways of organizing urban development and making radical changes to how we develop and live. Still room to experiment and ask what if questions.
- **Audience**: Trying to model every single individual behavior, do we really need that type of detailed data? Do we really need some of this data? Are we chasing those details and missing the big picture?
  - **Adella Santos**: Yes, we still need all that data to tell a story so we can read it in a way that makes more sense to different audiences. Understanding travel behavior, the more accuracy you have to understand what people do and why. It might help for a particular issue, just take the data you need but another researcher may need another. Big data is a terminology now and still need to understand what it provides.
  - **Sarah Sun**: Always good to have detailed data, in ideal world we collect all the data you want. We live with limited funding and resources, do we really need to collect that much detailed data to understand the general trend of transportation?
  - **Adella Santos**: You're right about cost, that's why we have statistical models to do some area of a portion of population to do those estimates. A census is not necessary to be repeated for everything and statistical models can step in.
  - **Stephen Crim**: I appreciate that question and it is similar to Laurie's work comparing the big data source with something more macroscopic like the NHTS. If there are big discrepancies, then maybe we don't need big data sources work that is more macroscopic instead of microscopic. Bringing back up the privacy issues again, information is generated because of our actions and there should be a larger conversation about us as individuals generating the amount of information that we do.
- **Audience**: I've thought about this a lot, big data can substitute the traditional data. I'd love to get a copy of Laurie's presentation because it explains big data very well. There is a third type of data generated by commercial means and is not a public data source and is strategically purchased by the government to get a bigger picture. Big data collected and owned by the government can be made public, but data purchases from commercial sources cannot. I read about synthetic data, if we could work together across government to have a synthetic population dataset that everyone would use and see the insufficiencies and find out where to augment with big data sources. Over time the traditional data sources become older and are less applicable. Is there is a way to update the traditional sources with a synthetic source and it is agreed upon by all agencies and it's a nationwide program updated by all agencies?
  - **Adella Santos**: That is in use, such as the long distance exploratory research. We have to look at different measures in which to collect data, but that is not NHTS's role right now. Perhaps it is an issue we can explore more. Research is being done right now using NHTS as a foundation.
  - **Laurie Schintler**: We need to not abandon traditional sources of data, but they can work together to see what one provides and what one doesn't. It would be a good project for USDOT to fund.
- **Audience**: I work in safety so this question is about congestion data versus trip data. With big data, cell phones have a high concentration rate so we can use the location services to track them. Vehicle and GPS data are not as

concentrated. Is it easy for you to figure out that two people are sharing a vehicle with two data points with comparable data and use it for vehicle demand modeling? How do you <u>break out the data</u>?

- o **Moderator:** It's always a question when looking at the data and seeing how to break it out. We've been discussing some of the issues using non-traditional sources of data.

- **Audience**: People are doing more, asked to do more, with less and must look to change what they are doing. There needs to be value in collecting, compiling, and analyzing data. What would you offer to a local agency or State DOT to <u>change how they are doing business</u> now so they can get at that value proposition for collecting and analyzing data?

  - o **Laurie Schintler**: Don't want to start processing a dataset without any purpose behind it. For a local agency it is important to see how local sources of data are not meeting their needs and hone in on the issues they want to solve and the ways big data can assist decision makers. I am strong believer in collaboration, so government should collaborate with academia, private sector, public, and community leaders.
  - o **Sarah Sun, Federal**: There is a point of diminishing return. Take the time to plan out and understand why the data is necessary. We don't want to end up with no data at all.

- **Moderator**: Good point of how do you <u>discriminate between valuable data</u> and how can local agencies go about doing that. Is there a starting point?

  - o **Stephen Crim**: Arlington County is working on an inventory of Mobility Lab's dataset to get a sense of where we are generating data points and where we aren't. Local jurisdictions still dealing with internally generated data. Mobility Lab has the general purposes outlined and measurements to achieve and questions looking to ask, then see how we can draw together all of the internal sources together to make that more useful and take that as a first step. Smaller scale using "small" data and easier to control, then see how that process works before going out and purchasing the big data.
  - o **Audience**: We can continue this discussion at the TRB Annual Meeting, ABE30 (Transportation issues in major cities) has a call for papers in Big Data and Addressing Urban Cities.

F. GTMA TransData Expo

Overview

During much of the two-day data showcase event, the GTMA-sponsored sessions ran concurrent to the Data Palooza sessions. Two of these sessions, both on day two, examined data management systems for transit and transit data applications. These GTMA sessions combined with the four Data Palooza sessions to provide a good indication of multi-modal data uses and needs. The sessions offered a basis for discussion during the Data Palooza's final facilitated session – Data Innovation Discussion: Action Planning and Commitments – held on the afternoon of day two.

GTMA TransData Expo

Rail / Transit Data Management Issues (Day 2 – June 4, 2014)
Moderator: Bill Toothill, Manager, Global Mapping Solutions / Technology Services Group, DBI Services

- Raquel Hunt, FRA, USDOT "Improving Railroad Asset Data Management"
- Keith Gates, Strategic Planning and Analysis Division, Office of Budget and Policy, FTA, USDOT "Performance and Asset Data – National Transit Database"
- Robert Borowski, GISP, Business Systems Analyst, San Diego Metropolitan Transit System "Transit Asset Data Collection Case Study"

Rail / Transit Data Management Issues (Day 2 – June 4, 2014)
Moderator: Bill Toothill, Manager, Global Mapping Solutions / Technology Services Group, DBI Services

Improving Railroad Asset Data Management
Raquel Hunt, FRA, USDOT

- 175,000 route miles of rail in the United States
- FRA oversees Amtrak passenger rail service
- Safety is the largest contributor to the FRA GIS system
- Automated Track Inspection Program (ATIP) provides foot-by-foot detail
- Highway-Rail Grade Crossing inventory is part of the FRA Asset Data. Rail Crossings Locator Map is created from the inventory. Each Crossing is identified through location and address standards in compliance with the National Emergency Numbering Association 9-1-1 numbering system.
- New mapping mandates tied to the Positive Train Control (PTC) Initiative

Performance and Asset Data – National Transit Database
Keith Gates, Strategic Planning and Analysis Division, Office of Budget and Policy, FTA, USDOT

- FTA Office of Budget and Policy does not retain a large GIS program
- National Transit Database's (NTD) annual budget is $3.45 million
- FTA is anticipating expansion of data collected and provided to the NTD with the Asset Management and State of Good Repair active rulemakings. Rulemaking will expand Assets reported, which are currently on a minimum

number of facilities. Assets currently collected include miles of track and number of highway-rail grade crossings in the system.

- All systems provide the number of vehicles operating in maximum service
- FTA and FHWA issue Conditions and Performance Report. The 2013 report was just released.
- Transit / public transportation systems reporting to the National Transit Database (NTD):
    - 821 Urbanized Area systems
    - 547 mid-to-large systems report monthly
    - 1341 rural area subrecipients (average of 20-30 per state)
    - 134 tribal transit program recipients
- NTD has monthly data from January 2002, Rural data from 2007, Safety data since 2002
- Are revising the uniform system of counts data process
- Service info is used by APTA to augment some of their figures collected

Transit Asset Data Collection Case Study
Robert Borowski, GISP, Business Systems Analyst, San Diego Metropolitan Transit System

- MTS collects Trolley Asset Management Inventory using LiDAR (LIght Detection And Ranging remote sensing technology). There are 8080 individual assets collected.
- System provides 3D LiDAR point data
- MTS procured a consolidated ERP / EAM system
- Why use LiDAR? The benefits include minimizing safety when collecting data on an operating transit system based on the traditional collection methods. Also much faster and cheaper to collect. Much less labor intensive.
- Milepost issues arise with numbering consistency. MTS will need to correct any inconsistency in their Asset Management System.
- MTS currently has an RFP issued for the development of its Asset Management System

Rail / Transit Data Management Issues Q&A

- **Audience:** How do you reduce the opportunities to manipulate the NTD data use in incorrect context?
    - **Keith Gates:** A number of requests come through the Congressional Affairs Office and the FTA needs to be upfront regarding the use of its data. Recently we had to explain that incidents impacting safety are not reported to the NTD.
- **Audience:** NTD provides a great resource of data. GTFS also is a great resource for transit. How can they be connected?
    - **Keith Gates:** GTFS is not part of the NTD's statutory mission. However, the FTA does not want to miss valuable opportunities to enhance its database. FTA is willing to look at integrating GTFS within its reporting system to reduce reporting demand on the transit agencies. May be able to enrich the NTD mobility and accessibility data with GTFS data.
- **Audience:** Can scatterplots be created from the FTA's NTD data? Can you show operating cost variations between agencies?
    - **Keith Gates:** Can insert data into existing charts. Can make scatterplots using downloadable NTD Excel spreadsheets. Operations data has not proven the most accurate within NTD, so not always easy to make comparisons among agencies for operational numbers.
- **Audience:** Has San Diego MTS considered tracking assets by linear reference points instead of mileposts?
    - **Robert Borowski:** For some newer items are going to linear reference points. Not for established assets such as switches and grade crossings.

Using General Transit Feed Specification [GTFS] (Day 2 – June 4, 2014)

Moderator: John Giorgis, Director, Strategic Planning and Analysis Division, Office of Budget and Policy, FTA, USDOT

- Andrew Owen, Director, Accessibility Observatory, University of Minnesota "Total Accessibility Calculations and GTFS"
- Alexander Bell, Senior Planner, Renaissance Planning Group "EPA Smart Location Database and Access to Jobs and Workers via Transit"
- Kurt Raschke, Mobility Lab "One Bus Away: Suite of Open Source Tools"
- Adie Tomer, Senior Research Associate and Associate Fellow Metropolitan Policy Program, Brookings Institution "Leveraging GTFS Data to Improve American Transit"

Transit Accessibility Calculations and GTFS

Andrew Owen, Director, Accessibility Observatory, University of Minnesota

- Calculated transit accessibility within 30 minutes; what can be reached per unit of travel time?
- Investments in transit improvements and service frequency show direct employment (economic) benefits over time
    - States and local areas can set targets to improve performance, but difficult because accessibility has not been measured before in most areas and no baseline is available.
    - Can be used by planners to also show the reverse that service cuts and reduced transit investments could result in lost job access.
- Minnesota DOT has six years of location accessibility data per mode.
- Calculations include approximately 16 other metropolitan areas:
    - Washington, D.C. (heavy rail distribution)
    - Atlanta, GA (heavy rail distribution)
    - Seattle, WA
    - Minneapolis – St. Paul, MN
    - San Antonio, TX
    - Chicago, IL
- Critical data sources:
    - LEHD (Longitudinal Employer-Household Dynamics) – provides block level employment estimates from the Census Bureau
    - Open Street Map
    - GTFS – shows plan, not performance
        - Automated Vehicle Location (AVL) data is only available in GTFS format
        - GTFS Data Exchange is backed by Amazon S3 (internet storage system). This is the only location where old GTFS data can be found.

EPA Smart Location Database and Access to Jobs and Workers via Transit Tool

Alexander Bell, Senior Planner, Renaissance Planning Group

- EPA Smart Growth website (http://www.epa.gov/smartgrowth/smartlocationdatabase.htm) includes the Smart Location Database and the Access to Jobs and Workers via Transit Tool.
- GTFS can be used to produce value-added data products
- National reviews at regional level where transit stops exist
    - In Tampa, someone who is carless can only access seven percent (7%) of jobs with 45-minute travel time
    - Durham-Chapel Hill and Raleigh data shows abutting regions with overlapping transit service
- Difficult to make cross-metro comparisons for jobs access
    - Compare single metros over time
    - Varying data collection sources, AVL and others

- Not all properties utilize GTFS or share data on GTFS Data Exchange

One Bus Away – Suite of Open Source Tools for Transit Data Aggregation and Dissemination
Kurt Raschke, Mobility Lab

- One Bus Away began as a transit traveler information application research and development product by the University of Washington. Now available for all transit systems in the Puget Sound region. MTA in New York now has MTA Bus Time which uses fully open-source software.
- One Bus Away utilizes a suite of open source tools. Inputs include:
    o GTFS (General Transit Feed Specifications)
    o GTFS–RT (GTFS-Real Time)
    o SIRI (the Service Interface for Real Time Information, a European standard for real-time transit data)
- One Bus Away outputs include:
    o Web and mobile applications
    o API (application programming interface) for developers
- Common standards and formats leads to reusable software; build software for standards not for cities
- Free license is provided to use One Bus Away software. An agency sets up One Bus Away server to run program software.
- Challenges:
    o Static data collection and availability
    o Real-time data collection and availability
    o Data quality problems, many times created because of staffing shortages for QA/QC, consistency checks, and regular updates. Proprietary systems many times provide dirty data that an agency is required to fix.
    o Adoption of open source versus conventional proprietary solutions.
- Apps are only as good at the underlying data. Getting access to the data feed is the greatest challenge for most applications developers.

Leveraging GTFS Data to Improve American Transit
Adie Tomer, Senior Research Associate / Associate Fellow, Metro Policy Program, Brookings Institution

- 95 percent of the U.S. transit passenger miles are in the top 100 metro areas that includes 371 transit systems.
- Brookings study document transit use and jobs access from a large number of metropolitan areas. Brookings has manually coded over 150 GTFS feeds.
- Transit falls short on connecting people to jobs in cities or suburban areas:
    o Large cities provide better access than smaller cities
    o 94% of large city jobs are accessible via transit
    o 41% of central city (urban) jobs are accessible via transit
    o 22% of suburban jobs are accessible via transit
- GTFS-supported decision making show the benefits from transit investments. GTFS feeds provide better comprehensive metro planning. Great support in orienting Transit-Oriented Developments (TODs) in cities and suburbs. New transit-job access metrics can be used as a gauge for discretionary grants.
- Need Federal and metropolitan collaboration to improve data through a standardized national geospatial transit database. A standard compatible and convertible language is needed to overcome current interoperability issues.

Using GTFS Q&A

- **Audience:** What is the <u>budget</u> for the transit accessibility study?
  - **Andrew Owen:** There is a Pooled Fund for the Annual Accessibility Report. The Pooled Fund provides from $100,000 to $150,000 per year to continue (and expand) the Report. The accessibility study is also calculating accessibility via automobiles.
- **Audience:** Does the Accessibility Study examine any <u>links to realty</u>?
  - **Andrew Owen:** Realty use hasn't come up yet. Accessibility success comes through partnering with transit advocacy groups and constructing new stops or lines. We haven't had the opportunity to examine beyond planned alternatives.
- **Audience:** Any additional work published that uses the <u>EPA process</u>? How does the <u>transit-jobs access database</u> calculate travel time?
  - **Alex Bell:** All current work using this process has been done under contract to EPA. For autos we have detailed street coverage, but not travel time. EPA might publish the travel times. To calculate travel time involves processing a lot of data. Most of it was done in GIS or in an Access database; it requires a lot of manual labor. The EPA Smart Growth database did not use Open Street Maps, but did incorporate NavTech "all streets" database that EPA purchased.
- **Audience:** What is the <u>benefit of One Bus Away</u> if we already have Google Transit?
  - **Kurt Raschke:** The big thing about Google Transit is that it's proprietary data, so there is no control over what these applications do. These providers (Google, Bing, Apple, etc.) are free to present the data how they want, use what they want, and they could even stop collecting and sharing when they want. It's not a platform for research or development, it is first, and foremost serving their own commercial needs. One Bus Away looks at the local data needs in a more pragmatic method.
- **Moderator:** How does <u>Mobility Lab</u> get involved? It is offering a free application. How is it supported?
  - **Kurt Raschke:** It's a loose relationship between partners, various other academic institutions and interest groups. Right now Mobility Lab is hosting One Bus Away applications for the DC area. One Bus Away does not have a permanent home right now.
- **Audience:** What kind of <u>interoperability</u> is sought with "standardization" of geospatial formatting or languages. Is it that the data being provided through GTFS is not good enough?
  - **Adie Tomer:** GTFS data is more than good enough. It is pretty easy to validate. It gives you what you need for research-based analysis. But it's about analysis and making smarter decisions for long-term plans and improvements. If transit agencies are required to make this data available and MPOs can easily analyze it, then benefits will be achieved.
- **Moderator:** What are the <u>standardization needs for transit</u>?
  - **Adie Tomer:** Bad data is worse than no data. Users will wait for buses and be disappointed or have a bad experience, so transit professionals have a higher bar because public relies more heavily on that information. If you know what you're dealing with in the quality of the data, then processing and validating is easier and more accurate. Standardization improves data quality and consistency.
- **Audience:** Is One Bus Away using <u>crowd source data</u> to support, validate, or compare with the data provided (ground truth)?
  - **Kurt Raschke:** The One Bus Away mobile apps do have a feedback mechanism built in and closing the loop for the next release of the agency's GTFS. There are examples of real-time information from cell phones, but it's not guaranteed and no substitute for doing a comprehensive effort.
  - **Moderator:** The World Bank sponsored open source tools in developing countries to create additional GTFS sources. Many of these new locations had no public transit agencies so they typically had to work with privately run businesses. In addition, Carnegie Mellon University is looking at crowdsourcing to provide additional transit data.
- **Audience:** Are you aware of <u>TCIP</u> (transportation communications information protocol) developed by APTA? What is FTA's contribution to your studies?
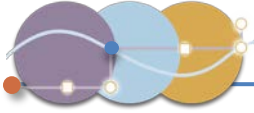
- o **Adie Tomer:** GTFS is straight forward and best format available for route data. There is more critical mass around it as compared to TCIP.
  - o **Kurt Raschke:** We mostly work with the NTD and try to coordinate the process. We looked at TCIP, too, but we couldn't figure it out.
- **Audience:** <u>Applying standards</u> universally for research is tough. Is the FTA the appropriate agency to move GTFS forward and can it happen?
  - o **Moderator:** It's finding the right mix between ease and quality. FTA does effect standards already such as through the NTD, but geospatial data has been outside the mission of the FTA. However, policymakers may expand that view.
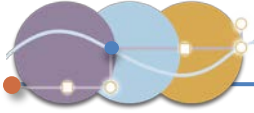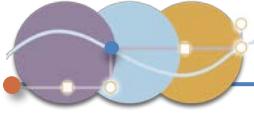
G. Data Innovation Discussions: Action Planning and Commitments

Overview

Data Palooza concluded with a break out session that focused on discussing highway and transit data issues, concerns, collaboration opportunities, and next steps. This session captured participant takeaways and the role of the Federal government in filling any data gaps. The conversations also focused attention on areas where the private and public sectors can collaborate on standardizing and better analyzing data. These action items show the shared responsibility and commitment of the private and public sectors to collaborate to improve the sharing, standardization, and application of transportation data. FHWA will consider incorporating the action items developed during the highway and transit discussions into its future data and project plans.

The action planning and commitment discussion sought answers to the following questions:

- **Standards for Sharing**
  - Public sector: When it comes to standardizing highway transportation data for sharing, what does the **public sector** do well? What are the top items the **public sector** can improve upon in the next two years?

  - Private sector: When it comes to standardizing highway transportation data for sharing, what does the **private sector** do well? What are the top items the **private sector** can improve upon in the next two years?

  - Public and Private sector collaboration: How can the **public and private sector** work together in this area to improve collaboration and results over the next two years? For each suggestion, who should be responsible (private, public, joint venture, other)?

- **Analyzing Integrated Transportation Data**
  - How would you like to collaborate with the **private sector** to address current limitations and advance the practice of analyzing integrated transportation data?

  - How would you like to collaborate with the **public sector** to address current limitations and advance the practice of analyzing integrated transportation data?

  - How can the **public and private sector** work together in this area to improve collaboration and results over the next two years? For each suggestion, who should be responsible (private, public, joint venture, other)?

G.1. Public and Private Sector Strengths and Improvements

Participants identified the strengths of the public sector as its ability to clearly communicate needs and vision to the wider public and private industries and to use the tools and processes in place to develop standards and data guidelines, such as following AASHTO's standardization procedures. Weaknesses of the public sector surrounded the lack of resources, structure, and organization for data management internally. Data management is underfunded and employee data literacy within agencies is limited, resulting in data that is not maintained and therefore the full benefits are not realized. USDOT has the foundation and processes in place to assist Federal and State governments to both build data skills and take the lead in working with stakeholders to develop and build upon existing data standards in order to obtain significantly more value of the data being collected.

The public sector does not have to work alone in this endeavor; it can collaborate with private partners to assist in addressing some of their public sector needs. Participants agreed that the private sector excelled in their willingness to learn and adapt to changes to position themselves to be at the forefront of researching, building, and implementing new technologies. Private and public sector attendees both expressed their interest and enthusiasm in working together to advance the state of the practice. Developing public-private partnerships through a multitude of innovative outlets is mutually beneficial and essential to ensuring an operative role of data in the future of transportation.

G.2. Data Innovation Discussion and Report Out: Transit Data

David Jackson (USDOT – Volpe Center) facilitated the transit discussion. The transit professionals who participated appreciated the opportunity to focus only on data concerns specifically from a transit perspective. A variety of issues were brought up surrounding data sharing, standardization, and analysis. Participants acknowledged the leadership role the Federal government should play and the importance of collaboration and coordination with the private sector in ensuring that advancements are made in the use of open data, data standardization, and integrating data.

The participants' data issues can be summarized into the following action items for FTA and their partners in the FHWA:

Break Down Modal Data Silos
The issue of data silos was a theme heard throughout the conference, and was again emphasized during this breakout discussion.. Transportation organizations have many different divisions encompassing multiple modes with separate databases and distinct data needs. If data silos are removed, increased data sharing within an agency, among public agencies, and across modes will provide a more holistic picture of the transit system and leaders can make more informed, data-driven decisions. The Federal role is seen as providing leadership and guidance as to how to benefit from the sharing of multi-agency, multi-modal databases.

Promote Understanding among Transit Executives to Support Innovation
Transit professionals noted the difficulty in trying to implement innovative technologies because funds are focused on operating and capital costs. The Federal government needs to encourage and support integrating innovative solutions in the transit world by communicating the importance to leadership and setting aside funds for data management, including upgrades to data collection, storage, and analytical systems.
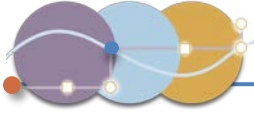
Support of Multi-modal Expansion
Participants representing both the transit side and the highway side agreed on the need to expand the multi-modal focus. MAP-21 emphasized expanding capital planning and project selection decisions to incorporate a more multi-modal transportation network perspective, yet guidance is needed as to how to achieve this goal. Greater data coordination and collaboration among the Federal government agencies is seen as a first step to attaining multi-modal databases that can support capital and operating decisions for a full transportation network. Leadership and guidance from the Federal government can support a holistic, rather than mode-specific, transportation network. They would alleviate transit agencies' concerns in cooperatively supporting new project selection processes that also include other modes and transportation functions such as pedestrian, bicycle, parking, freight, air, and water.

Establish a Transit Data Community of Practice
Participants expressed a desire for a forum for transit professionals and transit data users to talk and discuss best practices and new technologies. This would facilitate greater communication about practices being undertaken at other agencies and how the implemented technologies worked. With this data sharing forum, participants could make better decisions for their individual agency and help create a more informed and innovative national transit system. While professional associations were identified as the best facilitator of the transit data Community of Practice (CoP), the Federal government was seen as a necessary component to ensure that the conversations and innovations that are being discussed were expanded beyond the core group within the CoP.

Support a Transit Data Clearinghouse
It was recommended that the Federal government work across agencies to identify transportation-relevant databases and initiate a consolidated list of these databases and data points collected. This should be a multi-agency effort (FHWA, FTA, FRA, etc.). From the collection and assessment of the data already coming to the Federal government, the identification of data gaps can occur. At that point, both professional associations and the private sector can have an opportunity to provide input to various transportation forums (state, regional, local, research, etc.).

Data Innovation Discussions: Action Planning and Commitments

The attendees advocated for more open data as well as a platform for sharing this data. While there are many Federal, private, regional, State, and local data sources, most are not readily available because it is too expensive and time-consuming to maintain open data. Limited funds are allocated to data management, except in larger transit agencies. Even sharing data within a transit agency is a challenge and requires resources to collect and process the data from the various divisions. There are many groups and stakeholders interested in transit data which will require a large amount of collaboration in forming a data clearinghouse. It was suggested that Federal funds should be set aside for making transit data more open. If Federal funds are used, however, the data will fall under open data regulations, such as greater privacy rules.

The GTFS was suggested as a model to apply to other types of data. GTFS was a private sector project that helped fill a public sector need with expanded traveler information. Similar collaboration with private partners could assist in making other transit data open. Several participants also mentioned the possibility of expanding data gathered and reported by the NTD. The NTD could include a field where agencies can indicate whether or not to publish the data. More exploration should go into finding the best solution for a transit data repository.

G.3. Data Innovation Discussion and Report Out: Highway Data

Michael Nesbitt (FHWA – TPM) led the highway discussion. Highway professionals discussed takeaways from the conference as well as areas the Federal government can improve and lead in facilitating transportation data standardization and analysis. Participants appreciated the opportunity to meet new people and learn new industry practices. Several attendees commented on the robust amount of data available today because of new technologies and data sources. While more data is preferable, it also presents problems in figuring out how to manage and apply the data to get the most value. Agencies are having trouble keeping up with the vast amount of data available and might be missing opportunities to realize the full potential of some datasets.

Based on their comments, the following action items summarize recommendations for FHWA in helping move highway transportation data forward:

Federal Leadership Needed in the Data Standardization Process
Both the States and private sector acknowledged the need for guidance in the data standardization and management processes. The standardization process will be a long and iterative one with the participation of many stakeholders, but it is an important dialogue that needs to be facilitated by the public sector because the Federal government is the entity that creates and communicates national goals and needs.

Help Identify Data Gaps to Meet Current and Future Needs
Participants noted they learned about many data collection technologies as well as a variety of data sources available. With this much information available, it would be useful to find the areas where there is limited data in order to focus resources more effectively. To first identify data gaps, the Federal government can encourage a more open culture of data sharing so agencies do not become guarded when sharing their data problems. The identification of data gaps will reduce the amount of resources being invested in duplicative efforts and further transportation data analysis efforts. An attendee noted that identifying data gaps will help FHWA and USDOT organize their data and provide an opportunity to break down data siloes or hindrances to data sharing. More organized data will lead to a better understanding of the problem being solved and this can then be communicated more clearly and quickly to private companies who can assist in addressing that need.

**Provide Best Practices for Data Maintenance**
Data maintenance was a challenge often brought up throughout the conference and again during the highway data discussion. Data is most valuable when it is maintained and updated regularly. The benefits of transportation data will not be realized if it continues to be inadequately maintained. As mentioned above, data organization and management is important to knowing the needs of an organization and driving decision making. FHWA can act by providing capacity building support for data management programs.

Communication of Data Value
There lacks communication to leadership on the value and importance of transportation data, making it difficult to acquire the funds necessary for thorough collection, regular maintenance, and effective analysis. One State DOT stated that while leadership within the agency may understand the importance of data collection, data management is often not factored into the budget and not allocated sufficient human resources. With vast amounts of data being collected with the intent to apply it to the decision-making process, it is essential that this data is maintained and cleaned properly to maximize its value. Leadership may not see this larger picture of data and the FHWA should assist States in building and sharing an understanding of the value in data.

Provide Training to Help Address Skills Shortage in this Field
There is a need for public sector employees to become data literate. Many participants agreed that there are not enough employees with the right skill sets to manage and process data. This responsibility is usually assumed by the IT department; however, they typically do not have the capacity to handle that amount of data and workload. A guidance or training document created by the Federal government would ensure that employees are learning basic and standard data management
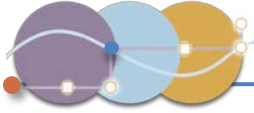
elements. One participant noted that private sector vendors may be able to help in this role of providing training to staff on how to manage the data and move beyond just delivering the data to agencies.

H. List of Exhibitors

Each GTMA and Data Palooza exhibitor was briefly interviewed to share why they participated in the conference and their perspective on the state of the industry as well as any recommendations they may have for USDOT moving forward. The exhibit halls were open throughout each day of the conference and gave the opportunity for vendors and researchers to showcase their work. One room held displays of private vendors while the other hall displayed university and government data applications. As professionals collecting and using the data, it was important to give them the chance to discuss the challenges they experienced and the opportunities they see. Several shared themes emerged from these conversations. It appears that both the public and private sector acknowledge that partnerships are the way forward and each plays a different, but vital role in advancing the transportation system. The exhibitor comments are summarized below:

Reasons for attending GTMA/Data Palooza:
- Interested in seeing the other work going on and opportunities for collaboration
- Share technology, research, and work with others

Concerns/challenges/issues facing the industry:
- Cost of maintaining data
- Databases are siloed within agencies and not interoperable
- Ensuring data does not release private information
- Identifying the right partners to coordinate and collaborate

USDOT, public sector stakeholders, and industry roles:
- Industry advances data collection and sharing technology for the public sector to employ
- Industry provides the inventory of assets and condition
- Public sector stakeholders analyze data to fit their prioritization, programming and performance needs
- USDOT facilitates and maintains open dialogue among a wide array of stakeholders
- USDOT maintain national transportation dataset repository
- USDOT coordinate establishing baseline data standards

Exhibitor	Description
Advanced Mobile Asset Collection	AMAC is a mobile, accurate, and automated system that measures traffic signage and pavement marking retro-reflectivity. It compiles an asset inventory and condition assessment.
Alliance Transportation Group, Incorporated	Alliance offers engineering, planning, modeling, and public outreach services to clients in both the public and private sector.
American Traffic Safety Services Association	ATSSA promotes moving towards zero deaths and the advancement of roadway safety.
Cubic Transportation Systems	Provides both front and back office services for transit fare ticketing.
Data Transfer Solutions LLC	DTS provides asset management, GIS, and transportation planning solutions for the transportation sector. They offer a suite of tools that assist with data collection, analysis, and storage.
Dynatest Consulting, Incorporated	Dynatest are pavement engineering specialists and technology providers dedicated to pavement engineering services, pavement testing equipment, pavement performance analysis software, and pavement management.
Esri	Esri is mapping software that assists organizations with integration, design, and inventory needs through visualization and analysis.
FHWA Long-Term Bridge Performance Program and Rutgers University Center for Advanced Infrastructure and Transportation, RABIT Bridge Inspection Tool	RABIT is a robot that inspects bridge deck condition. It includes two cameras and takes photos every two feet. The presenters demoed the RABIT Bridge Inspection Tool for attendees.
FHWA Office of Operations Research and Development	FHWA recently rolled out the Research Data Exchange that is part of the ITS Program Office. It is a web portal to share research data collected from connected vehicle prototype applications.
FHWA Turner-Fairbanks Long-Term Pavement Performance Program InfoPave	InfoPave is a relational database to assist States and localities select the best pavement for their unique situation.
FHWA Turner-Fairbanks Vehicle Emissions Due to Intersection Designs	A researcher at Turner-Fairbanks analyzed emissions data at a two-way stop intersection before and after it was changed to a round-about. The data was processed into a visual format to better communicate the emissions benefits.
International Cybernetics Corporation	ICC manufactures non-destructive test equipment for the governmental and consultancy industry.
GeoDigital	GeoDigital delivers engineering solutions for LiDAR and imaging technology to help agencies develop project schedule demands within budgetary constraints.
Google Enterprise	Google Enterprise offers a suite of products for businesses that uses the cloud to assist with data storage and sharing internally.
LimnTech Scientific	LimnTech is a striping technology that streamlines the process as well as enhances the accuracy and safety of roadway striping.
Mandli Communications, Incorporated	Mandli assists agencies with the design and development of highly specialized digital imaging, data collection equipment, and operational methodologies. Mandli provides advanced data collection hardware as well

Exhibitor	Description
	as a variety of comprehensive mobile application for data modeling.
PaveMetrics	PaveMetrics supplies high resolution 3D imaging and sensors of road condition. Through PaveMetric sensors, transportation and infrastructure managers can select optimal maintenance strategies and budgets.
Roadway Safety Foundation	As a non-profit organization, Roadway Safety Foundation seeks to reduce the frequency and severity of motor vehicle accidents, injuries, and fatalities. It is primarily a consortium of professional associations.
SHRP2 (2nd Strategic Highway Research Program) - Administered by the TRB, under a MOU with FHWA and AASHTO	SHRP2 was authorized by Congress to address some of the most pressing needs related to the nation's highway system: the high toll taken by highway deaths and injuries, aging infrastructure that must be rehabilitated with minimum disruption to users, and congestion stemming both from inadequate physical capacity and from events that reduce the effective capacity of a highway facility. SHRP2 launched 12 new products in May 2014, three of which are related to data. Offering both financial and technical incentives for states to adopt safety, capacity, and reliability technologies.
Terrasolid Limited	Terrasolid provides software for LIDAR processing that can apply data to improve pavement condition, design, and performance.
Transit Labs	Transit Labs is a data analytics company that seeks to improve data collection, reporting, analysis, and modeling for better transit decision-making.
University of Kentucky Transportation Center - Safe Routes to School Program (SRTS)	UKTC administers the SRTS for the State of Kentucky Transportation Cabinet. UKTC collects SRTS performance measure and outcome data for the National Center for Safe Routes.
University of Maryland CATT Laboratory	The CATT Lab designs user-centered software and information visualization systems. It develops real-time systems that collect and integrate various data sources to create a comprehensive database.
USDOT Volpe National Transportation Systems Center	The Volpe Center works on a variety of GIS projects and applications for different transportation agencies.
Velodyne Acoustics, Incorporated	Velodyne specializes in laser distance measurement and developed technology for visualizing the environment. The company has enhanced and commercialized their LIDAR scanner.
Virginia Tech Mechanical Engineering Department	The department has a project collecting data on gravel roads for terrain modeling that will assist in predicting the performance of different vehicle models.
Virginia Tech Transportation Institute	VTTI is collecting datasets from different sources and housing it to make it publically available. VTTI maintains and manages the data to make it research ready for the public.