NEW YORK STATE OF OPPORTUNITY™ | Department of Transportation

NYSDOT Research Peer Exchange:

The Use of Data Management and Analytics to Improve Agency Decision Making and the SPR Research Program

Feb 2-4 2021

DISCLAIMER

This report was funded in part through grant(s) from the Federal Highway Administration, United States Department of Transportation, under the State Planning and Research Program, Section 505 of Title 23, U.S. Code. The contents of this report do not necessarily reflect the official views or policy of the United States Department of Transportation, the Federal Highway Administration or the New York State Department of Transportation. This report does not constitute a standard, specification, regulation, product endorsement, or an endorsement of manufacturers.

1. Report No. SR-20-01	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle NYSDOT Research Peer Exchange: The Use of Data Management and Analytics to Improve Agency Decision Making and the SPR Research Program		5. Report Date: September 2021 6. Performing Organization Code
7. Author(s) Jonathan Hill, New York State Department of Transportation, Office of Policy, Planning and Performance; Ryan Lund, Transportation Research and Development Bureau		8. Performing Organization Report No.
9. Performing Organization Name and Address New York State Department of Transportation, 50 Wolf Road, Albany, NY 12232		10. Work Unit No. 11. Contract or Grant No.: SR21.01.027
12. Sponsoring Agency Name and Address New York State Department of Transportation, 50 Wolf Road, Albany, NY 12232		13. Type of Report and Period Covered 14. Sponsoring Agency Code

15. Supplementary Notes

This project was funded in part with funds from the Federal Highway Administration (FHWA). The following people from NYSDOT served as the research peer exchange hosts and/or participated in the coordination and development of the peer exchange and provided valuable input in the meetings and the final report:

Policy and Planning Division: Ronald L. Epstein, Executive Deputy Commissioner and Chief Financial Officer; Lynn Weiskopf, Director, Office of Policy, Planning and Performance; Jonathan Hill, Senior Administrative Analyst, Office of Policy, Planning and Performance; Jim Davis, Director, Statewide Policy and Performance Bureau; Mike Flynn, Acting Director, Statewide Planning Bureau; Lisa Cataldo, Senior Transportation Analyst and SPR Program Manager, Statewide Planning Bureau; Beth Brown, Transportation Analyst, Statewide Planning Bureau; Matthew Hannon, Senior Capital Program Analyst, Statewide Planning Bureau.

Engineering Division: Robert Sack, Director, Office of Technical Services; Daryl Bushika, Pavement Management and Materials Data Systems, Office of Technical Services; Michael Rossi, Director, Highway Data Services Bureau; Ryan Lund, Professional Engineer, Transportation Research and Development Bureau; Wes Yang, Professional Engineer, Transportation Research and Development Bureau.

Operations and Asset Management Division: Steve Wilcox, Transportation Maintenance Engineer/Manager, Office of Transportation Maintenance.

16. Abstract

The following report summarizes the results of the New York State Department of Transportation (NYSDOT) State Planning and Research (SPR) research peer exchange held on February 2-4, 2021. Managers and staff from NYSDOT's Office of Technical Services of the Engineering Division and the Office of Policy, Planning and Performance of the Policy and Planning Division hosted this peer exchange which focused on the use of data management and analytics to improve agency decision making and the SPR research program.

17. Key Words: DOT research peer exchange, State Planning and Research (SPR) peer exchange, DOT RD&T.	18. Distribution Statement: No restrictions.		
19. Security Classif. (of this report): Unclassified	20. Security Classif. (of this page): Unclassified	21. No. of Pages: 120	22. Price

Form DOT F 1700.7 (8-72)

Table of Contents

Peer Exchange At-A-Glance

"Use of Data Management and Analytics to Improve Agency Decision Making and Research"
February 2-4, 2021
Webex Video Conference

Host Agency
New York State Department of Transportation (NYSDOT)

Guest Agencies
AEM Corporation
California Department of Transportation (Caltrans)
District Department of Transportation (DDOT) – Washington D.C.
Florida Department of Transportation (FDOT)
Kentucky Transportation Cabinet (KYTC)
Maryland Department of Transportation (MDOT)
Massachusetts Department of Transportation (MassDOT)
Vermont Agency of Transportation (VTrans)
University at Albany Visualization and Informatics Lab (SUNY AVAIL)

Purpose
The emergence of new data sources, capabilities, and requirements has forced state DOT officials to find new ways to obtain, manage, analyze, and communicate data for research, planning, and decision-making. This peer exchange was organized to help states identify and implement ways to best leverage data systems, to enhance the effectiveness of their research programs, and to improve agency decision making.

Key Takeaways

- DATA GOVERNANCE: New and expansive datasets and technologies require updated data governance policies and procedures to achieve full research potential and efficient data use.
  - See AEM1, MDOT, FDOT , DDOT
- IT PARTNERS: IT staff are necessary partners in any big data project but are often unaware of research needs and aims.
  - See AEM1, AEM2, KYTC
- MODERN DATA MANAGEMENT: These modern approaches make data widely available through cloud-based storage using flexible storage systems.
  - See AEM1, AEM2
- COLLABORATION: Conducting effective, data-driven research requires collaboration between program areas within agencies as well as external consultants, partners, and stakeholders.
  - See AEM1, MDOT, NYSDOT, DDOT, FDOT, SUNY AVAIL, Caltrans
- PERSONNEL: Cross functional teams have been assembled to develop data analytics tools with modern data management approaches.
  - See AEM1, AEM2, KYTC
- PROOF OF CONCEPT: Pilot projects can demonstrate the utility of big data approaches to research, supporting executive buy-in for expanded research.
  - See KYTC, NYSDOT, AEM2, FDOT, SUNY AVAIL

- ACCESSIBILITY: Consistent and accessible templates, applications, and other points of engagement for non-specialists enhance the utility of data systems and expand capacity.
    - See KYTC, NYSDOT, DDOT, AEM2, Caltrans

Abstract

This report summarizes the results of the New York State Department of Transportation (NYSDOT) State Planning and Research (SPR) research peer exchange, which was held virtually on February 2, 3, and 4, 2021. This peer exchange was organized to help states identify and implement ways to best leverage data systems, to enhance the effectiveness of their research programs, and to improve agency decision making. To achieve this, the peer exchange participants explored modern data management approaches, personnel, data analytics case studies, organizational advancement approaches, and research in the context of data driven decision making. Several key peer states, national experts, and NYSDOT personnel participated in the peer exchange to provide many different perspectives and insights.

Acknowledgments

There are many parties that contributed to the success of the NYSDOT Research peer exchange. FHWA provided support and funds to make the peer exchange possible. Participants and other contributors allocated time out of their busy schedules, joined the conversation and shared valuable insights from their experience through presentations and discussions to advance data analytics, management and research. Participants were also involved with documenting the peer exchange through contributions to this report and providing presentations to share. These contributions were vital to the success of this peer exchange and are greatly appreciated.

Introduction

Purpose
With the emergence of new datasets, new data capabilities, and new data requirements, states are increasingly using resources to obtain, manage, analyze, and communicate data for research, planning, and decision-making. This peer exchange, "Use of Data Management and Analytics to Improve Agency Decision Making and Research," was organized to help states identify, share and implement ways to best leverage data systems, to enhance the effectiveness of their research programs, and to improve agency decision making.

Process
The Peer Exchange was conducted virtually through the Webex meeting platform throughout three afternoon sessions from February 2-4, 2021. The peer exchange began with a brief welcome followed by self-introductions of state and partner participants. The first two afternoons consisted of several technical presentations followed by an "open mic" session for posing questions and answers, engaging in discussion, and capturing key takeaways.

The final afternoon session was a topic-oriented, participant-guided discussion. This began with a brief overview by each state generally describing their state, agency, and research program, as well as identifying at least one key data interest. This was followed by a topic-by-topic discussion of data management procedures to identify current constraints and best practices, as well as opportunities to enhance agency decision making and research through improved data management and analytics. The Webex chat box function was used to capture discussion on all three days.

Requirements
This peer exchange was conducted in accordance with 23 CFR 420.209 (a)(7), requiring periodic peer exchanges between state DOTs on research-related topics as a condition for receiving FHWA planning and research funds.[1]

[1]23 CFR 420.209: "(a) As a condition for approval of FHWA planning and research funds for RD&T activities, a State DOT must develop, establish, and implement a management process that identifies and results in implementation of RD&T activities expected to address high priority transportation issues. The management process must include: … (7) Participation in peer exchanges of its RD&T management process and of other State DOTs' programs on a periodic basis. To assist peer exchange teams in conducting an effective exchange, the State DOT must provide to them the information and documentation required to be collected and maintained under this subpart. Travel and other costs associated with the State DOT's peer exchange may be identified as a line item in the State DOT's work program and will be eligible for 100 percent Federal funding. The peer exchange team must prepare a written report of the exchange."

Technical Presentations

This section describes in detail the presentations provided and key takeaways captured from these presentations.

Session 1: February 2, 2021
Understanding the Data Challenge – Turning Data into Information

AEM Corporation

"Guidebook for Managing Data from Emerging Transportation Technologies: Overview of Best Practices for Modern Data Management"

This presentation provided an overview of the recently completed project NCHRP 08-116 (Research Report 952), "Framework for Managing Data from Emerging Transportation Technologies to Support Decision-Making." This presentation focused on the differences between the traditional data management approaches of most transportation agencies and the modern data management best practices needed to successfully carry transportation agencies into the future.

NCHRP 08-116 served as a foundation supporting discussions on the many challenges faced by departments seeking to leverage the potential of new data types and sources in transportation research. Attendees discussed the main challenges facing their agencies including struggles to handle large and dynamic datasets produced by new technologies; institutional data silos; the need for executive buy-in; and a lack of funding and staffing to build, operate, and maintain data capacity. The presenters offered strategies and resources for developing a framework for managing data from emerging technologies and applying this framework within transportation agencies.

Additional tools and resources can be found in NCHRP 08-116 (Research Report 952).

Takeaways

- Managing new data requires a paradigm shift – it is not something that can be addressed incrementally.
- The nature of new data requires an updated approach – understanding needs; obtaining the necessary data professionals; building the flexibility to handle, analyze, and apply new data.
- Most data research activities at peer DOTs occur across a patchwork of program areas, IT departments, and outside academic and consultant partners.

Kentucky Transportation Cabinet (KYTC)

"KYTC's Big Data Journey Including Various Use Cases and Tools"

This presentation covered the Kentucky Transportation Cabinet's journey into big data, discussing the many benefits along with organizational challenges, technical lessons learned, and the recent migration from on-premise architecture to cloud computing.

KYTC began utilizing big data in 2014 with the introduction of crowd sourced data into snow and ice operations. The system has steadily grown to support use cases and research projects in nearly every engineering division of the Cabinet. The data management practices are also now being used as a template by the enterprise data team.

An overview of the legacy KYTC IT workflow was provided with processing steps on several types of data to arrive at many different visualization tools. KYTC has created a data lake for raw data storage, and employs many different software tools including Elastic search, ESRI ArcGIS and ArcPro, Excel and UrbanSDK to support many different data use cases. The data visualization use cases covered snow and ice, traveler information, crash detection, crash analysis, and work zone monitoring, among others.

KYTC also described its recently employed cloud computing structure and the advantages and disadvantages in its use. The presentation compared on-premise and cloud-based data management, concluding that on-premise data storage requires large investments in servers and data management support services for software and hardware used. These requirements could challenge the capabilities of the IT support services available. Cloud-based processing was found to be relatively less expensive than on-premise storage, although still potentially costly.

Attendees learned how KYTC achieved this transformation in organizational data processes and began to expand capacity to other use cases beyond snow and ice removal. This led to a detailed discussion on the mechanics of activities at KYTC, from building data capacity internally, getting the most out of consulting and licensing agreements, and evaluating different visualization tools and software.

The Kentucky Snow and Ice data analytics development team created many valuable data analytics pipelines and data visualization tools. This team consists of a manager with a background in engineering, surveying, GIS, and IT; a GIS/ITS specialist; and a full stack developer with experience with Java, C#, and Python.

Takeaways:

- Data lake is a misunderstood concept; likely to encounter resistance from IT.
- On-premise data solutions afford greater control, but often exceed the technical capacity and expertise of state DOTs and require large, up-front investments.
- Data literacy or data analytics training is needed to leverage the full utility of data resources.
- Cloud-based and third-party solutions typically lack integrated dashboard interfaces and can pose problems when there is a loss of power or data connectivity.

Maryland Department of Transportation (MDOT)

"Data into Information for Transportation Decision-making – The Maryland Experience "

This presentation described the state of geospatial data governance and management practices at MDOT State Highway Administration (SHA) and efforts to ensure that these practices inform decision making processes. It also detailed some of the challenges and opportunities in turning data into information – getting value from the "data flood."

The discussion focused on the challenges presented by the vast quantity of data now available to transportation planners. These include, for example, the existence of data silos across program areas, lack of uniformity in reporting, and a lack of enterprise systems to allow for "apples to apples" comparison in decision making processes. Traditional procurement methods were also identified as a potential obstacle to data-driven decision making, since they often do not reflect the needs of new data management practices and infrastructure.

Takeaways

- Transportation decision making will require a combination of traditional sources, big data, and crowdsourced data.
- Improved data management practices will require champions, ongoing investment and engagement, and commitment for mainstreaming.
- Collaboration between agencies (DOTs, MPOs), data providers, and the research community is critical.
- The value of improved data analytics must be demonstrated to support more sustained investment and systemic support.

New York State Department of Transportation (NYSDOT)

"Applied Data Governance in the Development and Management of NYSDOT's System of Engagement – Maps and Apps Library"

This presentation provided an overview of NYSDOT's System of Engagement (SoE) project, with a particular focus on the data governance strategies employed to develop the Maps and Apps library. The project focuses on democratizing authoritative data so that it can be used by people throughout the department to make better decisions. By putting the focus on the needs of end users of the data, NYSDOT has been able to expand the data use beyond traditional functions, such as meeting mandated reporting requirements.

The discussion focused on how to translate existing sources of data into useful, reusable, map-based Systems of Engagement that support a dynamic range of department activities. NYSDOT's ever-growing library of maps, apps, and other data services was offered as a model for leveraging and adapting unstructured and tabular data to real-life, geospatially oriented research use cases.

NYSDOT data visualization needs were identified by meeting with personnel throughout the DOT and discussing operational needs. New GIS based apps can be created by building upon existing web services and integrating available data. As NYSDOT presenters indicated, once a visualization need is identified, an app can be developed within weeks, resources permitting.

Takeaways

- Making data useful to non-specialists within the agency is critical.
- Systems of Engagement allow data to be made useful continually in different use cases. The key is to have a reliable and authoritative data source.
- Making tabular data geospatial is a useful way to bring data together, integrate data and allow the user to "visualize" relationships.

District Department of Transportation (DDOT) – Washington D.C.

"Getting Started on Data Governance at DDOT"

This presentation explored data governance development at DDOT and use cases demonstrating why data governance matters. As the largest data owner in the Washington D.C. government, DDOT efforts seek to encompass people, processes, and technology to achieve improved data accessibility, quality, and accountability.

Several benefits of data governance were identified including easier access to data, less time to find and access data, less duplication of data analytics, and enhanced retention of institutional knowledge. Using data governance should result in better data-driven decision making.

A data wiki was developed to house information about DDOT data using the Atlassian Confluence platform, which provides data parameters such as public availability, location, owner, update frequency, format, and access details. Efforts are being made to hire a full-time employee to support this effort, keep the data wiki current, and support the data governance efforts more generally. DDOT has data analysts on staff in a few key groups but many program areas rely on interns to start the shift to more data-intensive efforts, before eventually looking to hire full-time positions (which are sometimes filled by former interns). The discussion focused on DDOT's initial strategies to allow data systems to "talk" to each other, improve efficiencies, reduce duplication, and retain institutional knowledge. Knowing your data and the users is the first step in the process. By reducing the time and effort necessary for staff to track down data, practical data governance protocols can increase efficiency. Institutionalizing these practices and training staff in data governance practices results in better research outcomes.

Takeaways

- Common data templates and centralized datasets facilitate collaboration, wider use of data, and efficient data utilization.
- Lack of data governance results in redundant data collection, ineffective data discovery, and suboptimal data use.
- Collaboration with data owners is critical to effective data use.
- Full-time staff with redundant data capabilities is necessary for thorough data analysis. Agencies need those who can translate between the need for analysis and what data to use to get there.

Session 2: February 3, 2021
Channeling the Data Flood into Decision Support

AEM Corporation
"Roadmap to Big Data"

This presentation built on AEM's contribution in the first day's session, "Guidebook for Managing Data from Emerging Transportation Technologies: Overview of Best Practices for Modern Data Management," which discussed the findings of the recently completed NCHRP 08-116 (Research Report 952). This subsequent presentation described in detail the 8-step Roadmap for Managing Data from Emerging Technologies, with a focus on the big data concepts required in Step 4: Establish an Embryotic Big Data Test Environment or Playground, and Step 5: Develop the Pilot Project Within the Big Data Test Environment/Playground.

Participants discussed big data concepts related to Steps 4 and 5 such as data lake, the cloud, and distributed computing, and began to relate these to practices at their agencies. The discussion focused on how agencies can iteratively develop and demonstrate the benefits of a modern, big data approach to data management, analysis, and decision-making.

Additional tools and resources can be found in NCHRP 08-116 (Research Report 952).

Takeaways

- Organizational change can be realized through iterative steps that demonstrate the value of big data approaches.
- Buy-in from IT is critical to establishing a big data test environment or "playground". This is not a traditional environment for IT which tends to structure systems for specific use cases. The data lake is designed to have use cases applied to the data. Data integrity is important, but users need to be able to apply different questions to the data.
- Test projects must be tailored to the capabilities and needs of the agency, available consultants, and partners. Start small, think big!
- Data management implementation assistance is temporarily available through NCHRP 08-116.

Florida Department of Transportation (FDOT)

"Florida Live Testbed: Data Analytics and Artificial Intelligence for Smart City Transportation"

This presentation described the collaborative development between FDOT and the University of Florida of real-time transportation data system in the campus vicinity. This prototype system can use incident detection, vehicle classification, space-time trajectories, near-miss identification, signal retiming, travel-time distributions, signalized intersection control strategies, and sensor fusion algorithms using real-time video and ground sensor data and artificial intelligence algorithms. The purpose of this project is to produce smarter intersections, streets, networks, and explore the possibilities of connected and automated vehicles (CAVs).

The presentation detailed the I-STREET Trapezium project in Gainesville, which is employing lidar, video, and other sensors at 27 intersections to test new sensors and analytics for improving traffic operations and safety. Lidar and video data can be used to improve the accuracy and speed of detection for vehicles, pedestrians, bicyclists, and other modes. The data can then be processed and synthesized to improve signal and network operations and then stored in a data warehouse.

The team at the University of Florida is developing signalized intersection control strategies and specialized algorithms for optimizing traffic flow for autonomous and traditional vehicles. Short range communication is employed for autonomous vehicles and radar is used to detect traditional vehicles. These strategies are being tested on signals at the Florida DOT's Traffic Engineering Research Laboratory (TERL) facility and intersections in Gainesville.

Artificial intelligence and machine learning algorithms have been developed to detect real-time incidents, classify vehicles, detect near misses, and retime signals while maintaining user privacy. To implement artificial intelligence for operational insights, several components were employed, including data collections system, simulator for generating labeled training data, hardware infrastructure and software infrastructure. Simulated data were used to fill in gaps in the sensor data for training models.

Takeaways

- New, rich transportation datasets require new solutions for data and storage.
- Artificial Intelligence (AI)/Machine Learning (ML) is useful for developing real-world transportation applications.
- Many more potential operational uses exist for AI and ML applications.
- Collaboration between city and state transportation personnel and academic researchers produces dividends.

State University at Albany Visualization and Informatics Lab (SUNY AVAIL)

"NPMRDS – Probe Data Analytics Tools for Transportation Planning: Web-based Analysis and Reporting Tools for NYSDOT and NYS MPOs"

This presentation described SUNY AVAIL's development of a web based congestion, reliability and incident analysis tool with advanced analytics and interactive visualizations, based on the National Performance Management Research Data Set (NPMRDS). This tool has been developed in close coordination with NYSDOT and the State's Metropolitan Planning Organizations. While initially used to meet federal performance management reporting requirements, the tools have grown to support agency and MPO activities such as before/after project analysis, and MPO congestion planning. NYSDOT and the MPOs continue to explore new datasets and new use cases.

The unique partnership between NYSDOT and NYSAMPO, as a technical working group, led to several advancements in congestion and reliability performance management and planning. These advancements include newly developed local system performance measures, insightful data visualizations, and best practices in software database management and user interface design.
Advanced analytics with interactive visualizations and a series of transportation planning case studies were conducted across New York State. Value was extracted from the NPMRDS by using various performance measures, producing Congestion Management Plans (CMPs), and conducting corridor studies.

Takeaways

- Academic researchers can be productive partners in leveraging the utility of existing data resources and assisting the development of internal agency data capacity.
- Partnerships are critical to the success of useful tool development and data research programs.
- Successful data research programs need an ecosystem of support for continuity and utility.
- The NPMRDS, an open data source provided by FHWA, can be very useful beyond Federal performance measure calculations.

California Department of Transportation (Caltrans)

"2021 NYSDOT Peer Exchange Webinar"

This presentation focused on open source and collaborative software development and explored models for supporting the development and integration of new functions and capabilities of evolving analytics.

In seeking to meet the goals of multiple organizational units, this presentation described the challenges of creating a robust, performance-based, real-time transportation management system that is also replicable, scalable, affordable, and maintainable over many years.

Of special interest were the potential benefits of adopting open-source data systems to lower costs, enhance collaboration, eliminate license fees, and enhance accessibility. Participants explored in greater depth the potential benefits of using open-source data solutions at individual states. States also discussed the potential obstacles including, for example, a potential lack of consultant support or lack of personnel or training to maintain and operate these solutions. Open source requires technical resources to maintain and update data analytics software tools. The contracts used by open source portals such as GitHub can pose a legal obstacle for state DOTs. Open source development can encourage community involvement and improvements and allow flexibility to try new ideas and functions.

Takeaways

- The significant potential time and cost benefits of open source solutions should be considered in developing agency data capacity. Any approach to data analysis requires the production of actionable results to justify continued funding and support.
- Open source approaches can attract unanticipated collaborators, enhancing agency engagement. Such collaborations do require ongoing agency support.
- It is important to incentivize consistency and common tools and platforms with partners.
- University research collaboration can help implement open source solutions.

State Overviews

Session 3: February 4, 2021
Supporting Decision Making and Research with Data

On the final day of the peer exchange, several guest states provided a concise presentation on the general characteristics of their state (population, transit service, lane miles, etc.) and their agency data activities, and identify a data issue they were interested in exploring in greater depth.

The following summaries represent a brief description of data activities as well as the data areas of interest identified by each state.

New York Department of Transportation (NYSDOT)

New York's transportation system encompasses all modes – pedestrian, bicycle, highway, transit, rail, aviation, and marine. NYSDOT manages 7,900 bridges, 44,000 lane miles of roads with 7,900 employees in 11 Regions and one main office.

New York State's research program is managed jointly by the Transportation Research and Development Bureau (Engineering Division) and the Statewide Planning Bureau (Policy and Planning Division). These two groups leverage and coordinate their complementary expertise to support department research activities. Both groups are located at the NYSDOT main office in Albany. NYSDOT conducts research with in-house experts and through partnerships with academia; industry; other local, state, and federal agencies; and other states.

INTERESTS: Data analytics applications; Data governance; Modern data management solutions.

Vermont Agency of Transportation (VTrans)

VTrans manages 4,050 structures and 6,511 lane miles with 1,200 employees in three divisions and eight maintenance districts. VTrans research program activities are conducted by two employees in the Policy, Planning, and Research Bureau, part of the Policy, Planning, and Intermodal Development Division. Research is conducted with a budget of less than $1.3 million by leveraging NCHRP, TRB, New England Transportation Consortium, and pooled funds. Research staff coordinate with program area champions to prepare research problem statements, coordinate with executive approval, and then identify qualified researchers. The agency is interested in systems used by other State DOTs to store and manage performance data, as well as national research into open data, especially as it relates to federally funded research. It is currently focused on working with champions across the agency to help prepare research problems.

INTERESTS: Data storage; Data governance.

District Department of Transportation (DDOT) – Washington D.C.

DDOT research activities involve two employees with additional support from Howard University. These are housed within State and Regional Planning Division, which also handles SPR Part 1 and Metropolitan Planning, and which reports to the Chief Administrative Officer. The principal focus of research activities is urban and multimodal research.

INTERESTS: Integrating data across modes and systems.

Massachusetts Department of Transportation (MassDOT)

The MassDOT data research staff works across three groups in two offices: The Office of Performance Management and Innovation and the Office of Transportation Planning. This work supports both MassDOT and Massachusetts Bay Transportation Authority (MBTA), and frequently involves collaboration with universities, advocacy groups, and other organizations. The Research and Technology Transfer group manages a total budget of approximately $8 million (SPR funds with state match) with five staff members. The Office of Performance Management and Innovation conducts internal research to support policy and measure development and promotes data sharing through an Open Data Portal and Data Blog. The SPR-funded research process includes several phases: collect problem statements, select statements, identify principal investigator, develop scope, procurement, research phase, produce deliverable, implementation and evaluate benefits. Several research projects have involved data analytics including Using Mobile Lidar for Automated Asset Inventory and Condition Assessment, Multisource Data Fusion for Traffic Incident Detection, Massachusetts Depth to Bedrock, and Measuring Accessibility to Improve Public Health.

INTERESTS: Institutional structure of data research activities; Using third-party data.

Florida Department of Transportation (FDOT)

The Research Center oversees the FDOT's research program and contracts with state universities and other research service providers. The Research Center also participates in pooled fund studies with other state transportation agencies and contributes to national studies on subjects of benefit to Florida. At any given time, FDOT typically has more than 125 active projects.

INTERESTS: Coordinating data projects across working groups and stakeholders.

Topical Discussion

Session 3: February 4, 2021
Supporting Decision Making and Research with Data

Following individual state presentations, attendees were invited to engage in a topical discussion on any of the ideas presented during the peer exchange. The following preliminary topics were drawn from the preceding state DOT submissions and the presentations, and were used to frame the interests of participants in this discussion:

- Data lake and pipelines
- Data hosting and cloud computing
- Research role in data management
- Integrating third-party data
- Systems used to store/manage performance data
- National research on publishing open data

In addition to these general topical areas of interest, peer exchange organizers crafted a series of questions meant to reintroduce topics and issues back into the conversation that had been raised over the first two days. These were:

What has your State's experience been to date with applying these new streams of data to:

- Research?
- Planning and Performance measurement?
- Operations and maintenance?

How has your research program been applied to help navigate or enable effective use of data and analytics practice?

- Developing improved governance?
- Leveraging university programs?
- Collaboration on open source systems and tools?
- Private data or analytics purchases?
- integration of agency, private and modal data?

What are some common obstacles to the effective use of data for research, operations and planning analytics, and what strategies have you been able to apply?

- Procurement and funding – how are states funding servers, cloud storage, etc.?
- Staffing and training – are states using SPR funds for training?
- IT Support – hosting and tool development and access?
- Organizational silos and data governance?

What research problems or subject matter do you anticipate will be the focus of data needs over the next several years:

- Infrastructure inventory status (resiliency preparedness, asset management)
- Environment (GHG reduction)
- Transportation System Management and Operations (TSMO)
- Modal integration
- Planning and forecasting
- Economic or community development analysis

- Automation

This period of open discussion, informed by two prior sessions of presentations and issues raised in the Webex chatbox, produced a wealth of insight on a range of data issues and topics related to transportation. Drawn together from all three sessions, these are listed below by category:

Discussion Summary

Data Management

- Data silos can exist in large organizations from potentially isolated functionality of different program areas. When a program area conducts its business, it has typically returned high-level information to other parts of the organization as requested. Detailed data is often retained in silos within program areas. Now with the advancements in data analytics, this detailed data is in demand since modern machine learning methods can make sense of it for other applications beyond the initial purpose for collecting this data. This is motivation to move data from silos to date lakes, where it can be identified, and accessed more widely within an organization.
- Communication between program areas helps to promote sharing of data and provides opportunities for the same data to be used for multiple purposes. For example, real time traffic data can be used for operations and planning functions. Some DOTs have benefited from a data strategist to oversee data collection needs, eliminate duplicative data collection, and acquire needed data.
- State DOTs are becoming more motivated to move data into data lakes where they can be widely consumed throughout the agency using data analytics. Data lakes can accelerate the use of this data for research projects, as noted by Kentucky. Data lakes provide for the underlying data to remain unchanged while allowing the user to apply the data to a variety of different analytic purposes.
- Caltrans indicated that the development of a data lake is critical to developing better data analytics tools. Data needs to be available internally within DOTs and shared between partner agencies to provide better access to data.
- Although data lakes have many advantages, there are some challenges also. The process of creating a data lake requires reorganization of siloed data which can get complicated and expensive if existing systems need to be rewritten. Privacy could be an issue with data lakes if the data security is questionable or there are vulnerabilities. Data lakes typically have slower queries compared with data warehouses, which use relational databases.
- Some states have indicated that the costs of third-party data storage can have a variable cost depending on state size and needs. The accuracy of data can be an issue. For example, a data provider indicated that its data is accurate. However, the accuracy was only 40 percent.

Open Source Data Analytics

- The benefits and challenges of open source development of data analytics were discussed. Open source has the benefit of being customizable, not requiring a license, potentially low maintenance costs, and bringing development assistance from a wider base. However, some challenges were identified such as open source requires specialized skills to maintain. This expertise can be challenging to procure if only a couple of organizations have that capability. For example, the University Transportation Center (UTC) that Caltrans approached for open source development was not well suited for open source development. Also, open source development costs can be high depending on how much customization is required.

- There are potentially some legal issues for state departments of transportation (SDOT) regarding opensource development. A legal issue for open source development has been establishing ownership of intellectual property. If a state develops intellectual property with opensource development, the managing entity may retain ownership, which can be problematic. Open source portal legal agreements may require immediate sharing of source code. However, legal departments for state DOTs may not support this.
- AASHTOWare was discussed for open source development, although some participants expressed security concerns over the two-year gaps between software updates.

Personnel

- A discussion regarding personnel for improving data analytics and management established some interesting points. DOTs need to have personnel that can understand data systems and analytics while also having the ability to communicate well with IT and management.
- There are issues with retaining qualified IT personnel since SDOT wages can be lower than those provided by a consultant. In Kentucky, 70 percent of IT services are contracted out.
- The Kentucky snow and ice data analytics team has been very effective at using modern data management and data analytics for many operations use cases. Its team consists of a manager with a diverse background in engineering technology, surveying, and IT; an IT professional with GIS experience; and a software developer with programming experience in Java, C#, and Python.

Summary

This peer exchange was organized to help states identify and implement ways to best leverage data systems, to enhance the effectiveness of their research programs, and to improve agency decision making. The exchange covered a wide array of topics in data management and analytics, with many different perspectives from participating transportation managers, research managers, researchers, data managers, information science experts, programmers, planners, and engineers. The topics covered include research, modern data management, data governance, data storage, data analytics, open source development, collaboration, and personnel.

Staffing, Collaboration, and Support

Several states presented on their research programs, organizational overviews, and topics of interest. These presentations highlighted a common issue facing many transportation agencies – that successful data research programs need an ecosystem of support for continuity. The discussion between presenters and participants identified collaboration with universities, program area champions, and other transportation stakeholders as key to the success of research programs. The peer exchange itself demonstrated how academic researchers can be productive partners in leveraging the utility of existing data resources until such time that internal agency data frameworks can be developed.

Modern Data Management

The ongoing proliferation of new data types provides new research opportunities, but also requires new solutions for data management and storage. Modern data management practices were summarized with presentations provided by the AEM corporation and documented in detail in NCHRP 08-116 Final Report (Research Report 952). This presentation outlined the differences between traditional and modern data systems regarding governance, data storage, user population, and architecture – demonstrating how practical value can be derived from existing data. KYTC provided a working example of how to successfully employ modern data management and analytics to snow and ice applications.

Building Data Lakes

Modern data management is critical to efficient data storage, discovery, retrieval, and processing. As various presenters and participants recognized, data silos in large organizations can negatively impact the functionality of different program areas. To address the problem of program silos, participants discussed the emerging industry consensus that better outcomes can be achieved by moving from data silos to data lakes, where various data sources can be identified and accessed efficiently across an organization. Most participants agreed that creating effective data lakes is critical to developing and leveraging stronger, more efficient data analytics tools. Discussion on the clear advantages of data lakes later turned to the practical challenges that creating data lakes presents. The reorganization of siloed data can be both costly and time consuming, presenting a disincentive to data lake creation at agencies with limited resources and staff. Working through these challenges will require customized and pragmatic responses from agencies wishing to create data lakes.

Leveraging the Cloud

Multiple presenters, such as KYTC, touched on the debate between storing data in-house or using third-party solutions. The main storage options available to transportation agencies are on-premise and cloud-based, the latter of which generally include advantages like specialized processing tools, lower personnel resources for maintenance and security. Cloud-based storage was found to be relatively less expensive than on-premise storage, although still potentially costly.

Cloud computing was considered more expensive than cloud storage, and some states have indicated that the costs can be variable depending on state size and needs. There are some concerns regarding the unique and proprietary nature of cloud resources, the difficulty in transferring between different cloud platforms, and a lack of integrated dashboards. Participants also expressed concern over accessing data in the unlikely event that network connectivity becomes limited or nonexistent. As this discussion highlighted, transportation agencies have to carefully weigh many factors in choosing between on-site and cloud-based data storage.

Data Governance

Data governance provides policies to guide the storage and access of data within an organization, making it easier to find and use data. Common data templates and data sets can help establish and support data governance aims. Providing data to a broad user base can provide significant value within an organization, but it also introduces potential security risks that must be mitigated. The discussion of data governance focused on practical ways to balance data access and security at various kinds of agencies with different institutional cultures and resources. As an example, DDOT described how it is developing data governance through a Data Wiki used to summarize available data. A closeout process is being developed to keep the Data wiki updated. In another example, NYSDOT described how it has developed a Maps and Apps library to visualize department data sources geospatially, adding value to the data by providing access to a broad user population. There emerged a consensus among participants that efforts to establish new data governance policies first require an understanding of data sources and data users.

Data Analytics

Data analytics have the potential to transform data into actionable insights and deliver value to transportation stakeholders. The peer exchange explored some current data analytics and machine learning transportation applications already in use. FDOT described how its team is testing machine learning and artificial intelligence analytics to extract useful information from real time sensor data in a smart city testbed. These advanced analytics have leveraged various data to provide real-time incident detection, vehicle classification, space-time trajectories, near-miss identification, signal retiming, travel-time distributions, and signalized intersection control strategies. SUNY Avail described how its researchers have leveraged open data sets using a state-of-the-art, web-based congestion, reliability, and incident analysis tool with advanced analytics and interactive visualizations. The Kentucky Snow and Ice data analytics team demonstrated many valuable data analytics pipelines and data visualization tools it has created for snow and ice response, traveler information, crash detection, crash analysis, and work zone monitoring, among others. In all cases, it was clear that the software pipelines for data analytics were key to delivering valuable insights from data. Data analytics were conducted using a variety of different platforms including Python, Java, C#, Elastic search, ESRI ArcGIS, Excel, and Urban SDK. Also suggested, and explored in the discussion, was the idea that the success of pilot projects can help build up both data analysis capacity and executive support for additional data projects in the future.

Looking Forward

The development of modern data management practices within an organization was explored using the guidance from NCHRP 08-116 Final Report (Research Report 952), as well as valuable experience from participants. This peer exchange highlighted the critical insight that improved data management practices will require champions, investment, and commitment across an agency, from the executive level down to data practitioners and IT specialists. The utility of new data analytics opportunities will not always be self-evident – there will be an ongoing need to engage and sell executive management on their importance.

Skilled specialists are required to transform data into decisions using modern data management systems and data analytics, but a surge in hiring is not a reasonable expectation in most agency settings. In addition, some agencies have reported difficulties retaining professionals for data analytics and management. For this reason, building cross-functional teams with the personnel and resources agencies already have available to them will be key. Data literacy or data analytics training can expand access to the data, leveraging the utility of both the data and existing staff. Some organizations have a data expert embedded within each program area. Broad data access can reduce agency over-reliance on a single expert staffer.

Background experience or training that has been helpful for some teams includes computer science, IT, GIS, engineering, and management. DOTs need to have personnel that can understand data systems and analytics while also having the ability to communicate well with IT and management. Because of the critical role they play in technological change and transfer, IT staff must be brought on board as active partners in any adaptation to changing data management or analytics needs.

Pilot projects can deliver and demonstrate value but must be tailored to the capabilities of the agency, available consultants, and partners. There is much to be gained by applying data-driven solutions to transportation problems, and an ever-growing range of analytical tools and available datasets to support positive and transformative transportation outcomes for all stakeholders. This peer exchange demonstrated that big institutional change can be realized through iterative steps that demonstrate the value of big data approaches.

Next Steps

As the data landscape for transportation agencies continues to evolve, NYSDOT has begun adapting its business processes and will continue to do so. NYSDOT continues to engage in a successful partnership with the Metropolitan Planning Organizations with the assistance of the SUNY AVAIL team to successfully produce new data insights on the federally provided NPMRDS. This partnership serves as a model for successful collaboration on data projects, utilizing the resources and expertise of the academic community.

In addition, NYSDOT manages contracts with consultants working on a range of data issues. Through this dynamic and research-driven portfolio of projects, NYSDOT is gaining the knowledge and experience necessary to develop more robust data governance practices, partner more closely with IT staff, and identify data storage and security practices suitable to Department needs. Through its many data-related activities, NYSDOT will continue to learn from – and adapt to – the emerging big data environment.

As demonstrated during the peer exchange, NYSDOT staff continue to add new capabilities and interfaces to the Department's GIS-based Maps and Apps portal. This portal was conceived and developed as a collaborative and cross-functional platform for making data useful and accessible to all Department staff and, as appropriate, to the public in easy-to-understand formats. As NYSDOT continues to expand Maps and Apps offerings for a variety of users and use cases, this portal will facilitate further data integration, knowledge sharing, and data accessibility.

The Statewide Transportation Master Plan currently under development at NYSDOT represents a collaboration between consultants and a range of federal, state, and local groups, as well as the public. This long-term plan requires consultants to propose innovative approaches for incorporating new and growing datasets into planning activities. Through this contract, NYSDOT staff and partners are exploring new digitally driven approaches to conducting and measuring public involvement, new metrics to measure and model system performance, and new and accessible interfaces to map spatial information for non-specialist audiences. This project has the potential to enhance Department data processes across a wide spectrum of program areas and working groups.

Through these activities, among others, NYSDOT is constantly implementing and evaluating new data management and application strategies to enhance Department activities. Combined with insights drawn from the innovative work of transportation planners across the country – captured in this peer exchange – NYSDOT will continue to gain knowledge and explore practical solutions to the big data challenges which lie ahead.

Resources

- NCHRP Research Report 952: Guidebook for Managing Data from Emerging Technologies for Transportation
- NCHRP Research Report 952: Executive Summary
- NCHRP Report 814: Data to Support Transportation Agency Business Needs: A Self-Assessment Guide
- TRB Conference on Performance and Data in Transportation Decision Making (2019)

Attachments

Participant List

Name	Affiliation	Role
Benjamin Pecheux	AEM	Director of Information Science
Kelley Klaver Pecheux	AEM	Senior Director of Transportation
Chad Baker	Caltrans	Geospatial Data Officer
Mandy Chu	Caltrans	Chief, Office of Highway System Information and Performance
Nick Compin	Caltrans	Project Manager at Division of Traffic Operations
Stephanie Dock	DDOT	Research Program Administrator
J. Darryll Dockstader	FDOT	Manager, Research Center
Dr. Sanjay Ranka	FDOT	Professor at University at Florida (CISE)
Maria Chau	FHWA	Senior Community Planner
Gautam Mani	FHWA	Community Planner
Anna Price	FHWA	Acting Deputy Administrator
Chris Lambert	KYTC	TSMO Integrated Data Collection, Management and Analytics
Anna Gartsman	MassDOT	Director of Strategic Research
Lily Oliver	MassDOT	Research Manager
Laura Riegel	MassDOT	Director of Data Strategy
Subrat Mahapatra	MDOT	Deputy Director, Office of Transportation Mobility & Operations (OTMO)
Beth Brown	NYSDOT	Transportation Analyst, Statewide Planning Bureau
Daryl Bushika	NYSDOT	Pavement Management and Materials Data Systems, Office of Technical Services
Lisa Cataldo	NYSDOT	Senior Transportation Analyst and SPR Program Manager, Statewide Planning Bureau
Jim Davis	NYSDOT	Director, Statewide Policy and Performance Bureau
Ronald L. Epstein	NYSDOT	Executive Deputy Commissioner and Chief Financial Officer
Mike Flynn	NYSDOT	Acting Director, Statewide Planning Bureau
Matthew Hannon	NYSDOT	Senior Capital Program Analyst, Statewide Planning Bureau
Ryan Lund	NYSDOT	Professional Engineer, Transportation Research and Development Bureau
Michael Rossi	NYSDOT	Director, Highway Data Services Bureau
Robert Sack	NYSDOT	Director, Office of Technical Services
Lynn Weiskopf	NYSDOT	Director, Office of Policy, Planning and Performance
Steve Wilcox	NYSDOT	Transportation Maintenance Engineer/Manager
Wes Yang	NYSDOT	Professional Engineer, Transportation Research and Development Bureau
Alex Muro	SUNY AVAIL	Lead Programmer
Manuel Sainz	VTrans	Performance Analytics Group
Mandy White	VTrans	Data Unit in Operations and Safety Bureau

Overview Documents

NYSDOT Virtual Peer Exchange
February 2021

Topic –

Use of data management and analytics to improve agency decision making and research.

Overview –

With the emergence of new datasets, new data capabilities and new data requirements, states are increasingly using resources to obtain, manage, analyze and communicate data for research, planning and decision-making.

The purpose of this Peer Exchange is to help States identify and implement ways to best leverage data systems, to enhance the effectiveness of their research programs and improve agency decision making.

The Peer Exchange will be conducted virtually over the course of three afternoon sessions. The first two afternoons will consist of several technical presentations followed by a brief question and answer period and participant discussion. The final afternoon session will be a topic-oriented discussion, driven by the participants. It will begin with a very brief overview by each state providing a general context for their state, agency and research program. It will proceed with a topic-by-topic discussion of data management procedures to identify current constraints and best practices, as well as opportunities to enhance agency decision making and research through improved data management and analytics.

What to Expect –

The peer exchange sessions will be conducted through the WebEx meeting platform. They will take place during the afternoons on February 2 through February 4, 2021 (agenda follows).

Day 1 will begin with a brief welcome followed by self-introductions of State participants.

Each presentation on Day 1 and Day 2 will be followed by an "*open mic*" for questions and answers, discussion and to capture key take-aways. The Chat Box function will also be used to capture discussion.

Day 3 will begin with each state providing a very brief overview (5 minutes):

State:	State name and population
Agency:	Number of employees, regions/districts, lane miles and bridges
Research Program:	Number of employees, organizational structure, principle focus
Data Interest:	Identify one key data interest

Day 3 will proceed with a topic-oriented discussion on how to improve decision making and research through improved data management and analytics. Topics (or questions) will be compiled from participant interests provided prior to and during the exchange. Each topic will be addressed as follows:

General Decision Making	Research Activities
• Current practices.	• How does this data practice impact research?
• Constraints, best practices, opportunities.	• How can research improve this data practice? Research needs?

NYSDOT Virtual Peer Exchange
February 2021

Outline & Schedule –

	Understanding the Data Challenge – Turning Data into Information **Session 1 – Tuesday February 2, 2021 (12:30pm-4:00pm EST)**	
12:30 pm	**Welcome, Self-Introductions, Purpose & Process**	
1:00 pm	**Data management: Introduction and Scan of Emerging Practice**	
	Guidebook for Managing Data from Emerging Transportation Technologies Overview of Best Practices for Modern Data Management	AEM Corporation Kelley Pecheux Benjamin Pecheux
2:00 pm	**Getting Value from the Data Flood: Sensors, Probes and Crowdsourcing**	
2:00 pm	***KYTC's Big Data Journey Including Various Use Cases and Tools***	Kentucky TC Chris Lambert
2:30 pm	***Data into Information for Transportation Decision-making – The Maryland Experience***	Maryland DOT Subrat Mahapatra
3:00 pm	**Integrating Data Across a Vast Organization: Data Governance**	
3:00 pm	***Applied Data Governance in the Development and Management of NYSDOT's System of Engagement – Maps and Apps Library***	NYS DOT Steve Wilcox Mike Rossi
3:30 pm	***Getting Started on Data Governance at DDOT***	District DOT Stephanie Dock
	Channeling the Data Flood into Decision Support **Session 2 – Wednesday February 3, 2021 (1:00pm-4:00pm EST)**	
1:00 pm	**Emerging Practices for Big Data**	
	Roadmap to Big Data Steps to implement a data lake and cloud computing pilot to demonstrate the benefits for transportation applications	AEM Corporation Kelley Pecheux Benjamin Pecheux
1:45 pm	**Emerging Practices with Artificial Intelligence**	
	Florida Live Testbed: Data Analytics and Artificial Intelligence for Smart City Transportation	Florida DOT (UFL) Dr. Sanjay Ranka
2:30 pm	**Aggregating and Integrating Data to Tell the Story (Visualization)**	
	NPMRDS – Probe Data Analytics Tools for Transportation Planning: Web-based Analysis and Reporting Tools for NYSDOT and NYS MPOs	SUNY AVAIL Alex Muro
3:15 pm	**Collaborative Research - Open Source Software Development and Use**	
	2021 NYSDOT Peer Exchange Webinar Open source and collaborative software development. Models for supporting development and integration of new functions and capabilities of evolving analytics.	CALTRANS Nick Compin, Ph.D
	Supporting Decision Making and Research with Data **Session 3 – Thursday February 4, 2021 (1:00pm-4:00pm EST)**	
1:00 pm	**State Overviews**	
	States will provide a brief overview of their State, Agency and Research Program.	Participants
1:45 pm	**Improving Decision Making and Research – Topical Discussion**	
	A topic-oriented discussion of data management and data analytic procedures to identify opportunities to enhance agency decision making and research.	Participants
3:30 pm	**Overview of Peer Exchange & Wrap-up**	

Overview Slides

2

Welcome

Introductions – Self Introductions of Participants by State

Purpose

- Help States identify and implement ways to best leverage data systems, to enhance the effectiveness of their research programs and improve agency decision making.

Process

- Technical presentations – Q&A, Discussion
- "Open Mic" and Chat Box – Key Take-Aways

Agenda

Topics of Interest

NEW YORK STATE OF OPPORTUNITY. Department of Transportation

1

Virtual Peer Exchange

Use of Data Management and Analytics to Improve Decision Making and Research

Hosted by New York State Department of Transportation
February 2-4, 2021

NEW YORK STATE OF OPPORTUNITY. Department of Transportation

3

Day 1 Technical Presentations

Understanding the Data Challenge – Turning Data into Information Session 1 – Tuesday February 2, 2021 (12:30pm-4:00pm EST)		
1:00 pm	Data management: Introduction and Scan of Emerging Practice	
	Guidebook for Managing Data from Emerging Transportation Technologies Overview of Best Practices for Modern Data Management	AEM Corporation Kelley Pecheux Benjamin Pecheux
2:00 pm	Getting Value from the Data Flood: Sensors, Probes and Crowdsourcing	
2:00 pm	*KYTC's Big Data Journey Including Various Use Cases and Tools*	Kentucky TC Chris Lambert
2:30 pm	*Data into Information for Transportation Decision-making – The Maryland Experience*	Maryland DOT Subrat Mahapatra
3:00 pm	Integrating Data Across a Vast Organization: Data Governance	
3:00 pm	*Applied Data Governance in the Development and Management of NYSDOT's System of Engagement – Maps and Apps Library*	NYS DOT Steve Wilcox Mike Rossi
3:30 pm	*Getting Started on Data Governance at DDOT*	District DOT Stephanie Dock

Please Use the "Open Mic"
Use the Chat Box
Mute if not talking
Daryl.Bushika@dot.ny.gov

NEW YORK STATE OF OPPORTUNITY. Department of Transportation

4

Day 2 Technical Presentations

Channeling the Data Flood into Decision Support Session 2 – Wednesday February 3, 2021 (1:00pm-4:00pm EST)		
1:00 pm	Emerging Practices for Big Data	
	Roadmap to Big Data Steps to implement a data lake and cloud computing pilot to demonstrate the benefits for transportation applications	AEM Corporation Kelley Pecheux Benjamin Pecheux
1:45 pm	Emerging Practices with Artificial Intelligence	
	Florida Live Testbed: Data Analytics and Artificial Intelligence for Smart City Transportation	Florida DOT (UFL) Dr. Sanjay Ranka
2:30 pm	Aggregating and Integrating Data to Tell the Story (Visualization)	
	NPMRDS – Probe Data Analytics Tools for Transportation Planning: Web-based Analysis and Reporting Tools for NYSDOT and NYS MPOs	SUNY AVAIL Alex Muro
3:15 pm	Collaborative Research - Open Source Software Development and Use	
	2021 NYSDOT Peer Exchange Webinar Open source and collaborative software development. Models for supporting development and integration of new functions and capabilities of evolving analytics.	CALTRANS Nick Compin, Ph.D

Please *Use the "Open Mic"*
Use the Chat Box
Mute if not talking
Daryl.Bushika@dot.ny.gov

NEW YORK STATE OF OPPORTUNITY. Department of Transportation

1

Day 3 Activities

Supporting Decision Making and Research with Data Session 3 – Thursday February 4, 2021 (1:00pm-4:00pm EST)		
1:00 pm	State Overviews	
	States will provide a brief overview of their State, Agency and Research Program. NY, VT, DC, MA, FL	Participants
1:45 pm	Improving Decision Making and Research – Topical Discussion	
	A topic-oriented discussion of data management and data analytic procedures to identify opportunities to enhance agency decision making and research.	Participants
3:30 pm	Overview of Peer Exchange & Wrap-up	

Please *Use the "Open Mic"*
Use the Chat Box
Mute if not talking
Daryl.Bushika@dot.ny.gov

NEW YORK STATE OF OPPORTUNITY. Department of Transportation

2

General Decision Making	Research Activities
• Current practices. • Constraints, best practices, opportunities.	• How does this data practice impact research? • How can research improve this data practice? Research needs?

- **What has your State's experience been to date with applying these new streams of data to:**
  - **Research**
  - **Planning and Performance measurement**
  - **Operations and maintenance**
- **How has your research program been applied to help navigate/enable effective use of data and analytics practice?**
- **What are some common obstacles to effective use of data for research, operations and planning analytics and what strategies have you been able to apply (if any)?**
- **What research problems/subject matter do you anticipate will be the focus of data needs over the next several years?**

NEW YORK STATE OF OPPORTUNITY. Department of Transportation

Topics of Interest - Preliminary

1. Data Governance
2. Data Lake and Pipe Lines
3. Data Hosting and Cloud Computing
4. Research role in data management
5. How is 3rd party data integrated
6. Systems are used to store/manage performance data
7. National research on publishing open data
8. Role of research
9. How to coordinate with stakeholders on data issues

NEW YORK STATE OF OPPORTUNITY. Department of Transportation

Thank You

NEW YORK STATE OF OPPORTUNITY. Department of Transportation

Technical Presentations

AEM Corporation

Guidebook for Managing Data from Emerging Transportation Technologies

Overview of Best Practices for Modern Data Management

Kelley Pecheux, AEM Corporation

Benjamin Pecheux, AEM Corporation

NYSDOT Peer Exchange
February 2, 2021

Overview of Presentation

- NCHRP 08-116 background and objectives
- Challenges to managing data from emerging technologies
- Why should agencies move toward modern data management?
- Overview of Guidebook
- Laying the foundation – best practices for modern data management
- Supporting tools

aem

2

NCHRP 08-116 Background and Objectives

Background:

- New, big, and varied datasets are available to transportation agencies at an increasing pace.
- These data have tremendous potential to offer new insights to transportation agencies.
- The volume, speed, and granularity of these data are unprecedented and will fundamentally alter the transportation sector.

Research Objectives:

- Develop a framework for managing data from emerging technologies, including data from connected and automated vehicles and data linked to new mobility initiatives.
- Outline a process for applying this framework to help agencies incorporate these data into the decision-making process.

aem

3

Transportation Agency Challenges to Managing Data from Emerging Technologies

- Reliance on traditional database management systems. ***Data from emerging technologies are too large, too varied in nature, and will change too quickly*** to be handled by these traditional data systems.
- Struggle to break down ***business unit and data silos***.
- ***Do not fully recognize the value of big data or the eminent need*** to ready for it.
- ***Do not fully understand the uses and benefits*** of cloud-based architecture conducive to handling data from emerging technologies.
- Have ***difficulty hiring and retaining modern data management professionals***.
- Experience a *loss of control to vendors* over data, technology, and service agreements.
- Are ***unequipped to handle this level of big data at an organizational level***.

"Our big data issues are straightforward, we don't have the technology, money, or skills."

– CITY DOT

aem

4

Why Should Agencies Move Toward the Modern Approach to Data Management?

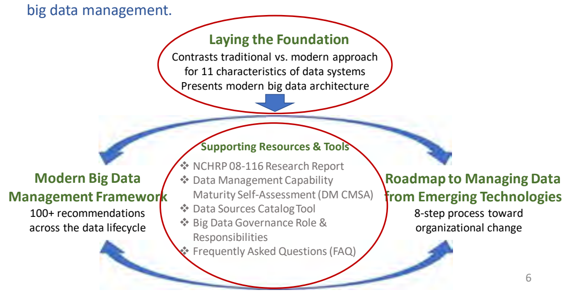
- With increased connectivity between vehicles, sensors, systems, shared-use transportation, and mobile devices, *unexpected and unprecedented amounts of data are being added to the transportation domain at a rapid rate*.
- These new *data offer the potential to uncover insights to drive better decision-making* at all levels of transportation agencies in a way that is simply not happening now.
- The potential *value of these new data cannot be easily or efficiently extracted by traditional methods*; the complexity of the task requires new big data tools and techniques.
- As data sources become more varied and change more and more rapidly, *the traditional approach cannot cope with the complexity and cannot be re-designed quickly or cost-effectively enough* to handle frequent data and business requirements changes.

➢ **Modern big data methods to collect, transmit, store, aggregate, analyze, apply, and share these data need to be adopted by transportation agencies if they are to be utilized to facilitate better decision-making**.

5

This Guidebook Can Help Agencies Shift Toward the Modern Data Management Approach

- Provides a ***modern big data management framework*** that introduces new concepts and methodologies, best practices, and 100+ recommendations for managing data in a modern, flexible, scalable, and sustainable way.
- Lays out a ***roadmap*** on how to begin to shift – technically, institutionally, and culturally – toward effectively managing data from emerging technologies.
- Provides ***examples and references*** of transportation agencies currently exploring or already navigating the implementation of big data, including their challenges and successes.
- Discusses ***common misconceptions*** within the transportation industry regarding big data management.



6

Laying the Foundation

Best practices for modern data management

aem

7

Traditional vs Modern Big Data Management

Managing new and emerging data requires a complete paradigm shift.

These data cannot be handled simply by adding more hardware or processing power.

The nature of the data demands an updated approach.

- The following slides contrast various characteristics of traditional data management practices with their big data management counterparts, providing examples that demonstrate the stark contrast between the current state of practice for most transportation agencies and the ideal state based on data industry best practices.

aem

8

System Design and Flexibility

Traditional Data Systems

- Systems are designed and built for a pre-defined purpose; all requirements must be pre-determined before development and deployment.
- System designed as "set it and forget it;" designed once to be maintained as is for many years. Systems are rigid and not easily modified.

Modern, Big Data Systems

- Systems are designed and built for many and unexpected purposes; constant adjustments are made to the system following deployment.
- System is ephemeral and flexible; designed to expect and easily adapt to changes. Detects changes and adjusts automatically.

aem

9

Hardware & Software

Traditional Data Systems

- As technology evolves, hardware becomes outdated quickly; system can't keep pace.
- System features at the hardware level; hardware and software tightly coupled.

Modern, Big Data Systems

- As technology evolves, system changes to keep pace with innovation. Hardware is disposable.
- System features at the software level; hardware and software decoupled.

aem

10

Data Storage & Processing

Traditional Data Systems

- Data and analyses are centralized (servers)
- Schema on write ("schema first")
- Analysis code is limited to the data organization imposed by the schema

Modern, Big Data Systems

- Data and analyses are distributed (cloud)
- Schema on read ("schema last")
- Analysis code can modify and customize the schema as needed

aem

Resource Allocation

Traditional Data Systems

- Majority of resources spent on the data system, from hardware and software to database design and maintenance.
- Data analysis is limited to the software used to create the database (often tied to commercial hardware and software stack).

Modern, Big Data Systems

- Many fewer resources spent on database design and maintenance (database hardware and software are maintained by the cloud).
- For each analysis, can use the most appropriate (often open source or pay-as-you-go software).

aem

Data Governance

Traditional Data Systems

- Data governance is centralized; IT strictly controls who sees / analyzes data (heavy in policy-setting)
- Uses a tight data model and strict access rules aimed at preserving the processed data and avoiding its corruption and deletion.
- Small number of people with access to data; limits use of data for insights and decision-making to a "chosen few."

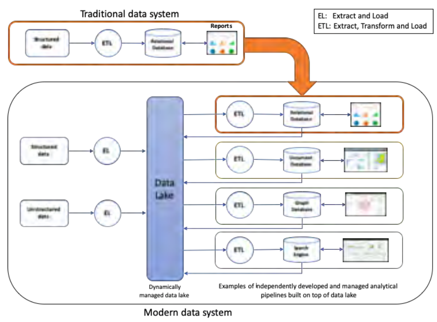
Modern, Big Data Systems

- Data governance is distributed between a central entity and business areas.
- Consider processed data as disposable and easy to recreate from the raw data. Focus instead is on preserving unaltered raw data.
- Many people can access the data; applies the concept of "many eyes" to allow insights and decision-making at all levels of an organization.

aem

Modern, Big Data Architecture

- New architectural patterns need to be adopted to cope with the wide variety of fast changing data
- Flexible and distributed data architecture capable of applying many analytical technologies to stored data
- Data is stored in a "data lake"
- "Schema on read" / "schema last"



aem

Guidebook Supporting Resources and Tools

aem

Guidebook Supporting Resources & Tools

- **NCHRP 08-116 Final Research Report** – *Framework for Managing Data from Emerging Technologies to Support Transportation Decision-Making*, provided under separate cover, documents the research activities and provides supplemental information for reference to support implementation of the guidebook.
- **Data Management Capability Maturity Self-Assessment (DM CMSA)** – offers over 100 questions to allow agencies to gauge their data management practices, as well as identify areas for improvement.
- **Data Sources Catalog Tool** – a tool to catalog existing and potential data sources.
- **Big Data Governance Plan Template** – provides a list of recommendations to consider when developing a modern data governance approach, a description and frameworks for big data governance, and a tool for tracking the big data governance roles and responsibilities within an agency.
- **Frequently Asked Questions (FAQ)** – responses to frequently asked questions regarding big data implementation, management, governance, use, and security.

aem

So, how do we get started?

Tune in tomorrow at 1:00!

17

Kelley Pecheux, Sr. Director, Transportation
kelley.pecheux@aemcorp.com
(703) 350-8487

Benjamin Pecheux, Director of Information Science
ben.pecheux@aemcorp.com
(703) 989-4776

Contact Information

18

Kentucky Transportation Cabinet (KYTC)

Big Data Journey

Kentucky Transportation Cabinet

TRANSPORTATION.KY.GOV

The Need for Change

- 2012-2013 and 2013-2014 Winters
  - Record Snowfalls
  - Record Costs
  - Salt Shortages
  - Interstate Incidents
- Spring of 2014 – KYTC began research for a Snow and Ice Decision Support System
- September of 2014 – KYTC signed with the Waze Connected Citizen Program
- November 2014 – Title 23, CFR 511.301-315 requirements for ITS

Data Sources

- HERE Traffic Speeds
- Waze Incidents
- Waze Traffic Speeds
- iCone Traffic Speeds
- Twitter
- KYMesonet
- CoCoRahs
- Doppler Radar
- NWS Forecasts: Rain, Snow, Ice
- Statewide TMC Reports
- Metro TMC Reports
- Snow Plows (AVL)
- Roadway Weather Stations
- County Activity Reports
- Dynamic Message Signs
- Truck Parking

Linear Referencing System

- District
- County
- City
- Route
- Road Name
- Mile Point
- Functional Class
- Etc…

- Capture Time
- Date Exists
- Date Exists 2min Rounded
- Source Update
- Source Record Update
- Year, Month, Day
- Hour, Minute, Second
- Sub-second (DSRC)

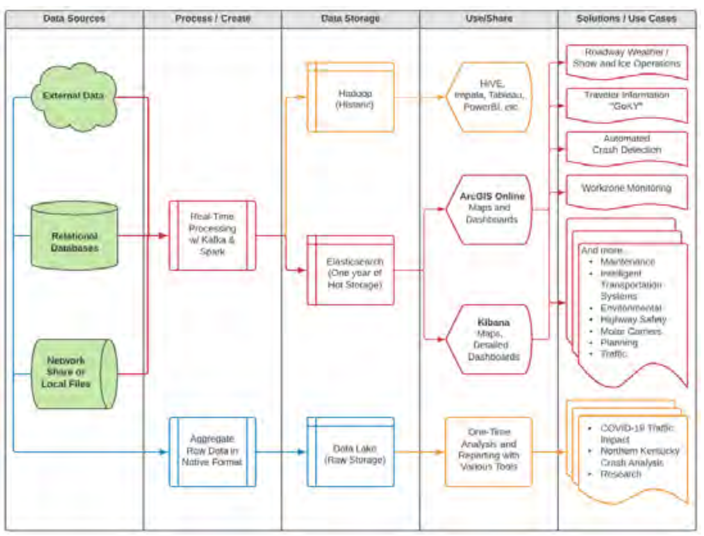
Current and Future Use Cases

- Title 23, CFR 511.301-315 (2014)
- Snow and Ice Management (2014)
- Situational Awareness (2014)
- Traveler Information System (2015)
- Incident Detection (2015)
- Incident Recovery Times (2016)
- Traffic Control Plan Training (2016)
- Environmental (2017)
- Work Zone Monitoring (2018)
- Secondary Crash Analysis (2019)
- Department of Motor Carriers (2019)
- COVID-19 Traffic Analysis (2020)
- Work Zone Performance Committee (2020)

- Maintenance Performance Measures*
  - Mobility*
  - Snow and Ice Operations*
- University Research*
- Business Intelligence*
- Predictive Analytics*
- Secondary Crash Detection*
- Congestion Mitigation*
- Signal Timing*
- Automated DMS Messages*
- Automated Bookkeeping*

*in development

Architecture



2020 Legacy Architecture

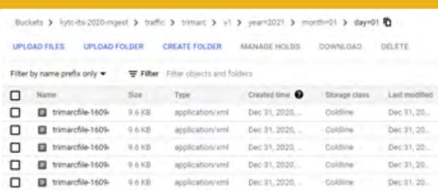
2020 Velocity:
24.6 million records per day*

1 year of rolling data is maintained in Elasticsearch for dashboards and reporting.

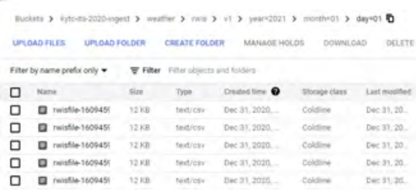
*Snow and Ice Operations generate an *additional* 10 million records per day of incoming data for a statewide event and output also increases to 125,000 records every 2 minutes for mapping and dashboards used in decision support solutions.

THE KENTUCKY TRANSPORTATION CABINET

Data Lake Debacle



- Misunderstood concept
- IT can be resistant or reluctant



- Independent process
- Data is NOT transformed

On-Premise (Nov 2014 - Sept 2021)

- Hardware: $260,000
  - 26 Physical Production Servers
  - 430 Cores
  - 832GB RAM
  - 145TB Storage
- Licensing Costs: $100,000/year
  - Cloudera
  - Elasticsearch
- Server Maintenance: $300,000/year
  - Additional maintenance contract
  - 50% of Developers time to servers

- One time cost for storage
- One time cost for compute/processing
  - 345mil records in 11min using 100% CPU
  - 345mil records in 17min using 20% CPU
- Skills and lessons translated to cloud

- Lack of education and skills
- Difficult to properly setup
- Maintenance headache

Google Cloud (Oct 2020+)

Google Cloud "Unlimited" Use Contract

- $12,000/month for unlimited usage
- Contracts are per use case
- Annual review and adjustment
- Actual usage cost: $4,353.87/month
  - $1,849.77 Compute Engine
  - $625.35 Cloud Pub/Sub
  - $600.12 BigQuery*
    *This will greatly increase! I ran a single $30 query earlier this week.
  - $493.30 Cloud Functions
  - $327.03 Cloud Dataflow
  - $319.95 Storage
  - $138.35 Stackdriver Monitoring

- Positive user experience
- Less time on administration
- SQL friendly
- Costs can be reduced with creativity

- New technology didn't change IT or procurement policies
- Still experience errors and downtime
- Lack of unified reporting/dashboard solution
- Developers are 100% dedicated as we migrate

Visualization Tools

Currently in use by KYTC:

- Elasticsearch / Kibana
- ESRI ArcGIS Online
- ESRI ArcPro
- Microsoft Excel (yes, *that* Excel)
- UrbanSDK (Data Science SaaS)

KYTC Enterprise Data utilizes Business Objects for legacy databases and reporting needs.

Evaluated but dropped:

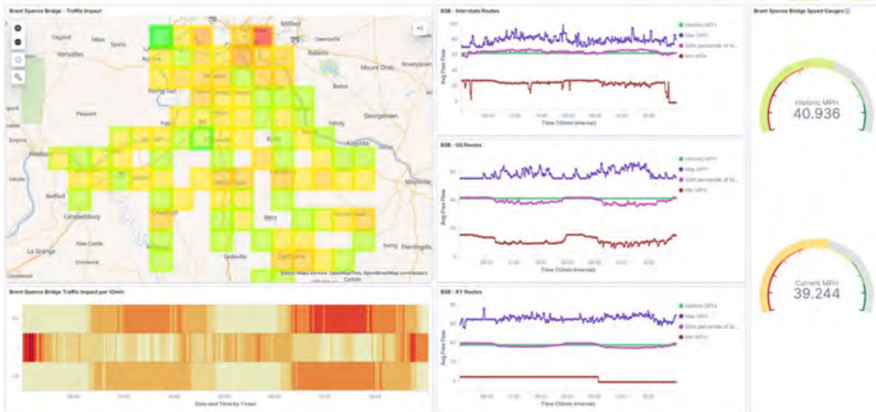
- Tableau
- PowerBI

Still being evaluated:

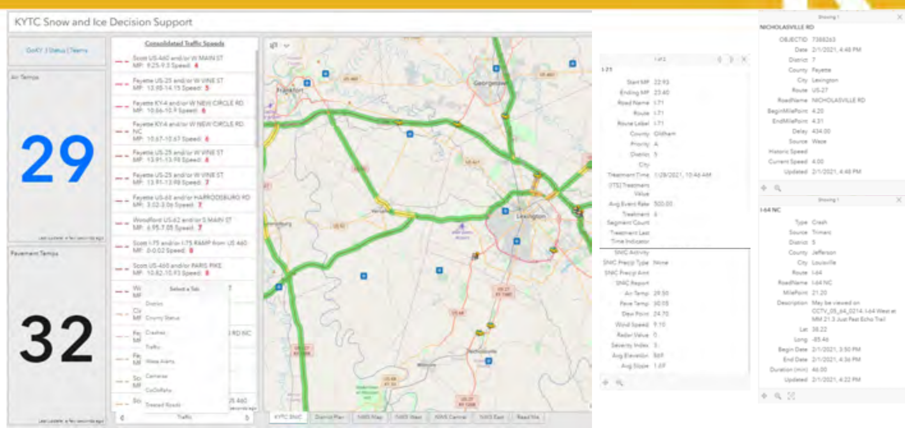
- Google Data Studio
- Looker (purchased by Google)
- Knowi.com

Use Cases

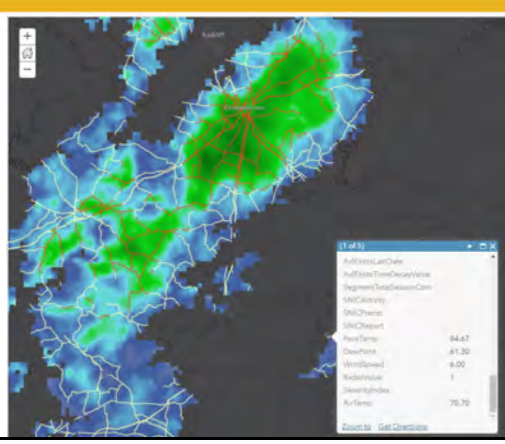
Traffic Impact of Bridge Closure



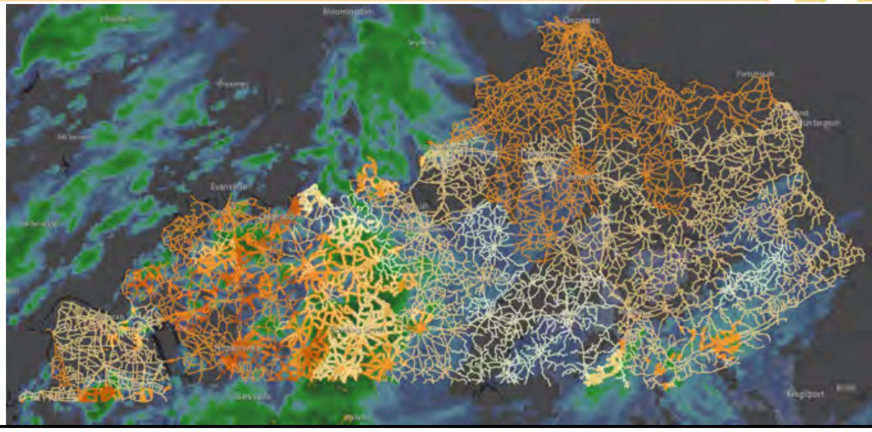
Snow and Ice: Decision Support



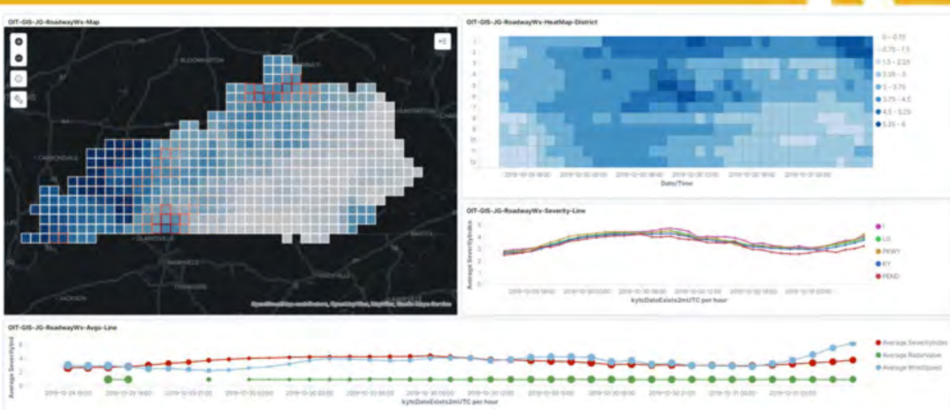
Roadway Weather



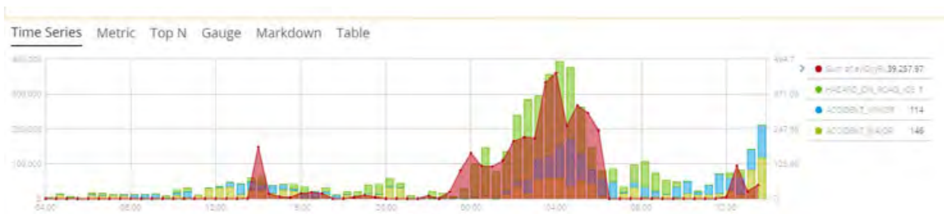
Roadway Weather



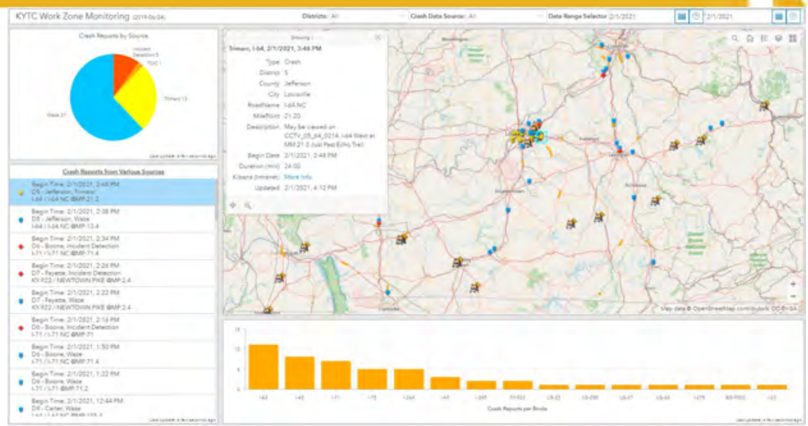
Roadway Weather Performance



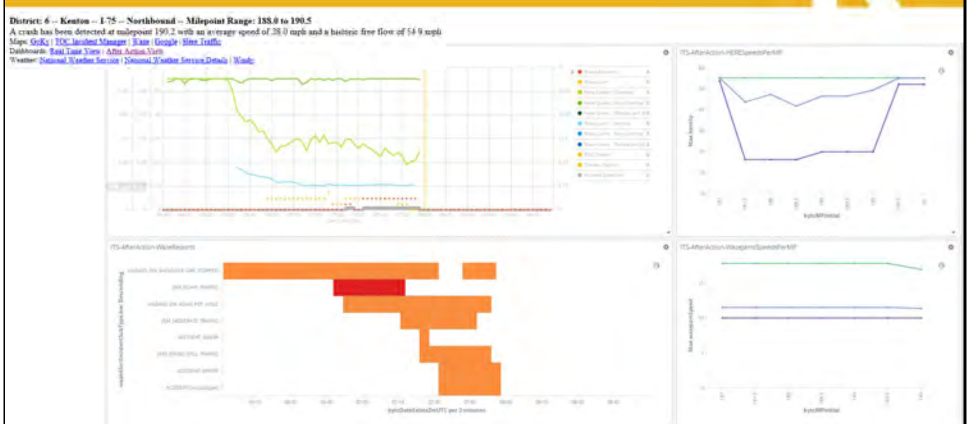
Snow and Ice Performance



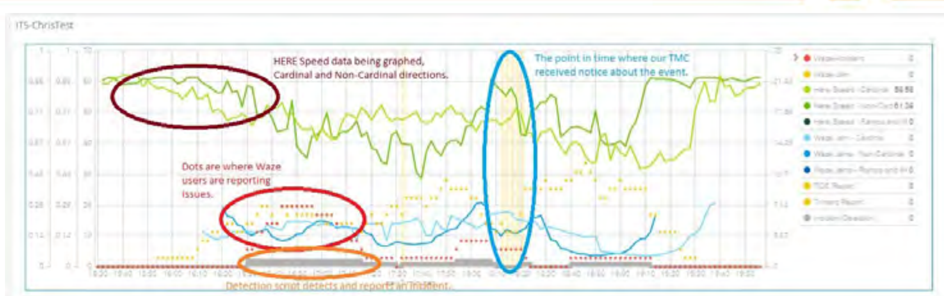
Work Zone Monitoring



Incident Detection

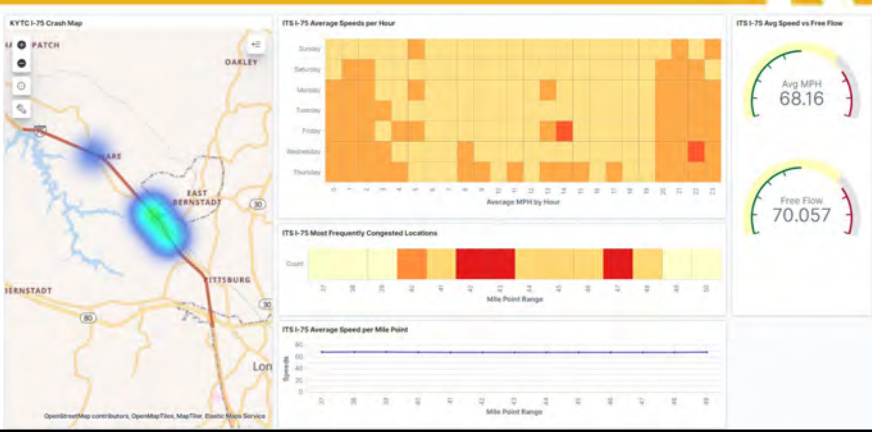


Incident Detection / After-Action

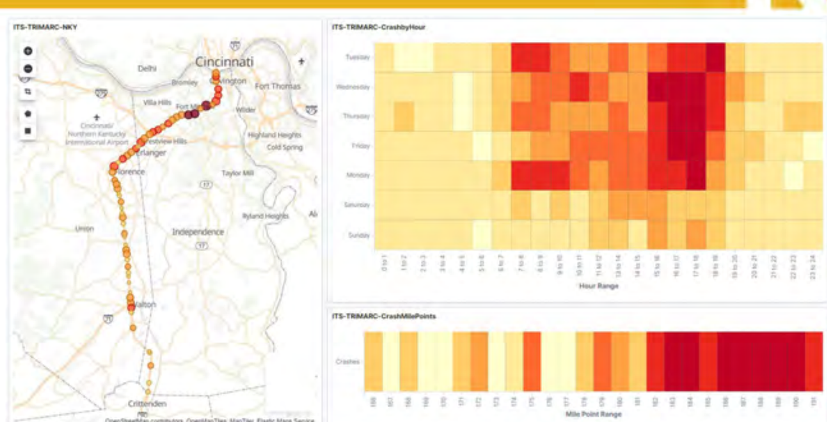


Work Zones: 7 Day View

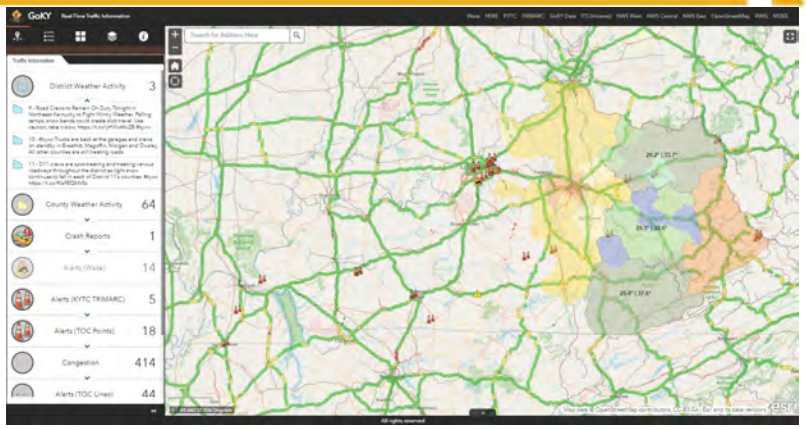
Work Zones: 30+ Day View



Crash History and Mitigation



GoKY: Traveler Information



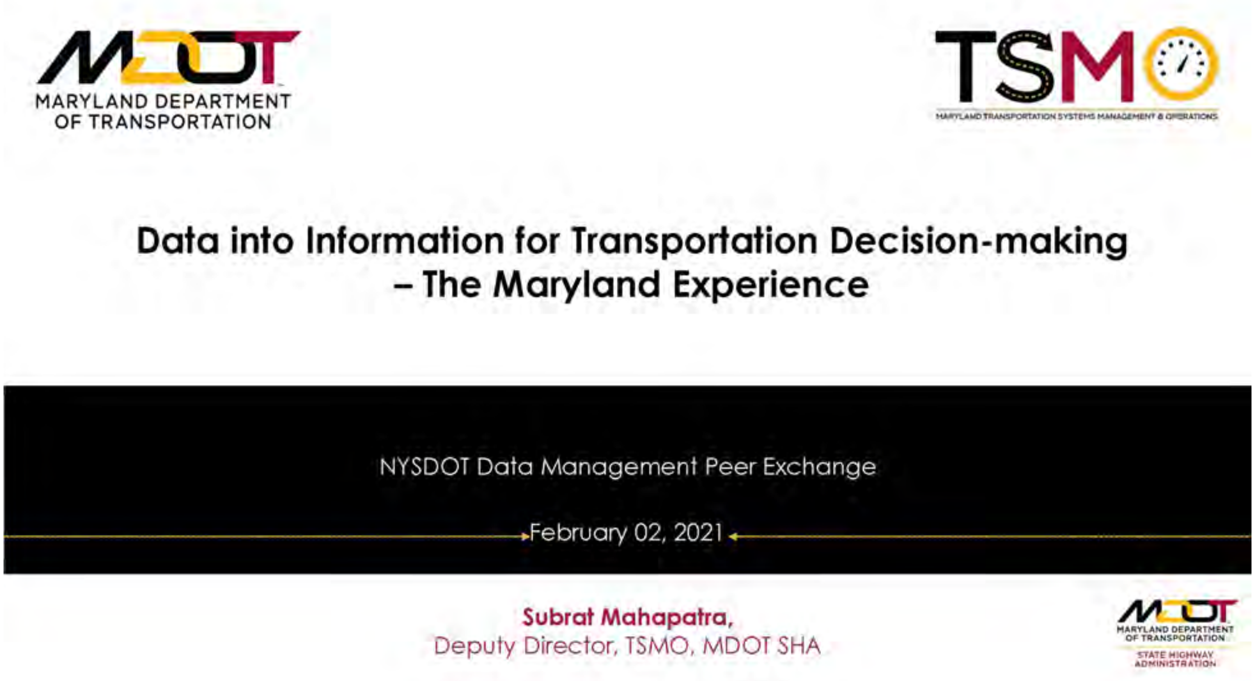
Questions?

Chris Lambert
chris.lambert@ky.gov

Twitter: @KYTC | @ChrisLambertKY
Facebook: /kytc120

THE KENTUCKY TRANSPORTATION CABINET

Maryland Department of Transportation (MDOT)



MDOT & MDOT SHA DATA STREAMS

MDOT comprises of six business units and an Authority. Strategic, Tactical, Operational Activities at the TBUs generates a highly complex set of data streams…

MDOT SHA operates and maintains the numbered, non-toll routes in Maryland - 17,000 lane-miles and 2,576 bridges

Many ways to looks at the Data Streams…

By MODE – Highway, Transit, Air, Water, Other…

By FUNCTION - Planning, Engineering, Operations, Maintenance, Performance Management

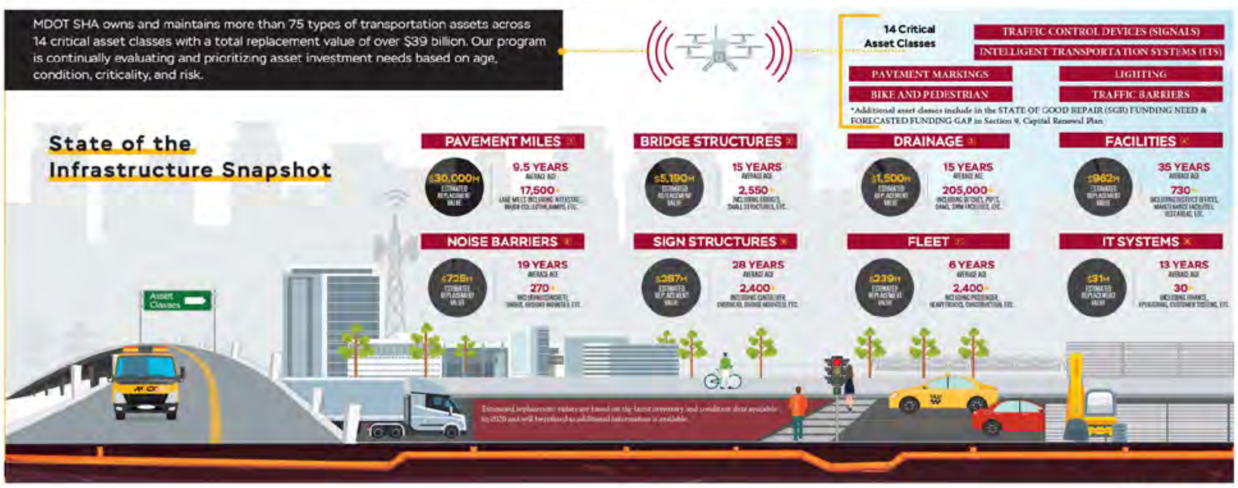
By PERFORMANCE AREA – Safety, Mobility, Asset Management, Administrative, Business Transactions

By SOURCE - Internal vs External Data

MARYLAND DEPARTMENT OF TRANSPORTATION
STATE HIGHWAY ADMINISTRATION

3

MDOT SHA ASSET PORTFOLIO



MARYLAND DEPARTMENT OF TRANSPORTATION
STATE HIGHWAY ADMINISTRATION

EXISTING PROGRAMS & DATA PLATFORMS

EXCELLERATOR

RITIS

Regional Integrated Transportation Information System (RITIS)

APPLICATION CONTEXT IS IMPORTANT

Too Much Data

Crash Data
Incident Response
Fatalities/ Severity

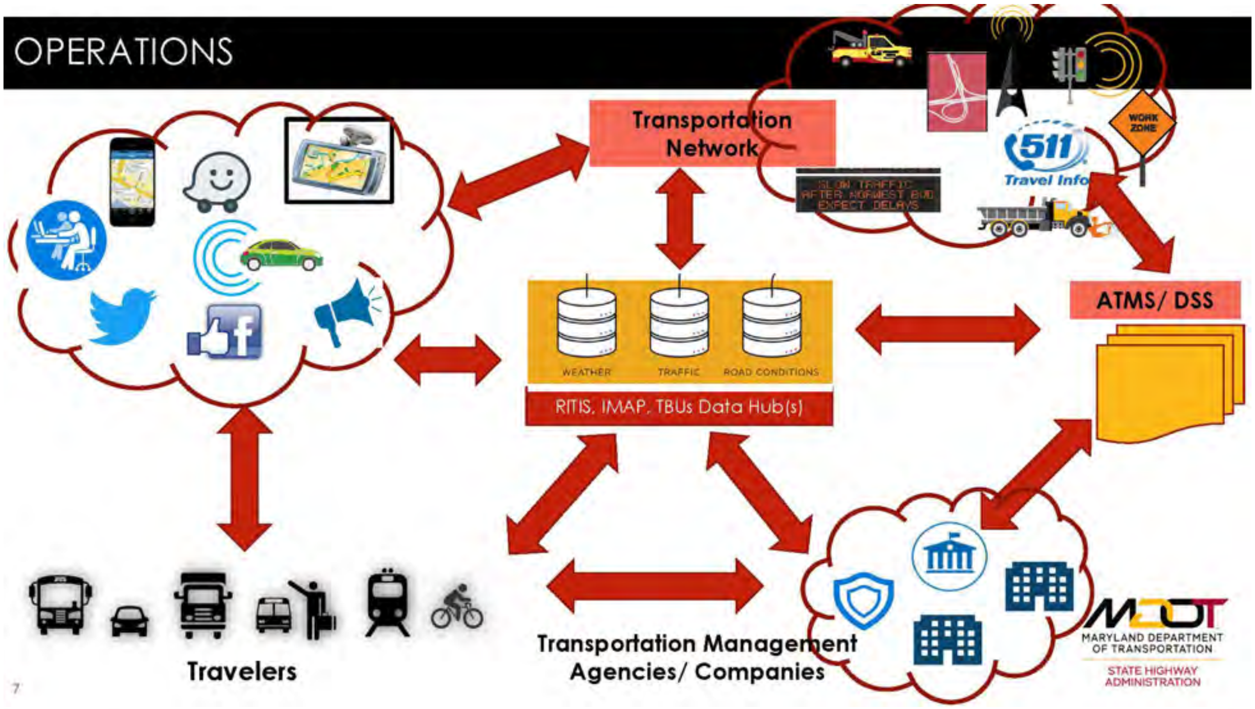
Asset Conditions
System Preservation
Risks/ Criticality

Average Daily Traffic
TTI/ PTI, Travel Times

Projects & Programs
Capital, O&M
Benefit/ Costs

Other Considerations:

- Multi-Modal Networks
- Facilities Operated by Others
- Freight Corridor Priorities
- Local Priorities, Past Studies/ Commitments
- Big Data/ Crowd Sourced Data



PLANNING

Planning Data

- Census Data
- DLLR Employment Data
- Auto/Truck Count data
- NHTS/ACS.MPO HTS Data

- TAZ/Block-Level Socioeconomic Data
- Existing/Future Transportation Networks
- Future Investments (CLRP, TIP, CTP, etc.)
- Sector Plans, TODs, Activity Centers, etc.)

Operations Data

- Signal Data
- Lane Configurations
- Probe/Speed Data
- Turning Movements
- Work Zone Limits
- Incident/Safety Data

Asset Data

- Centerline Attributes
- Pavement Attributes
- Roadway Structures
- Signal Systems
- ITS Assets (DMS, CCTVs, etc.)

Performance Management

- INFRASTRUCTURE CONDITIONS
- SEGMENT, CORRIDOR & SYSTEM PERFORMANCE

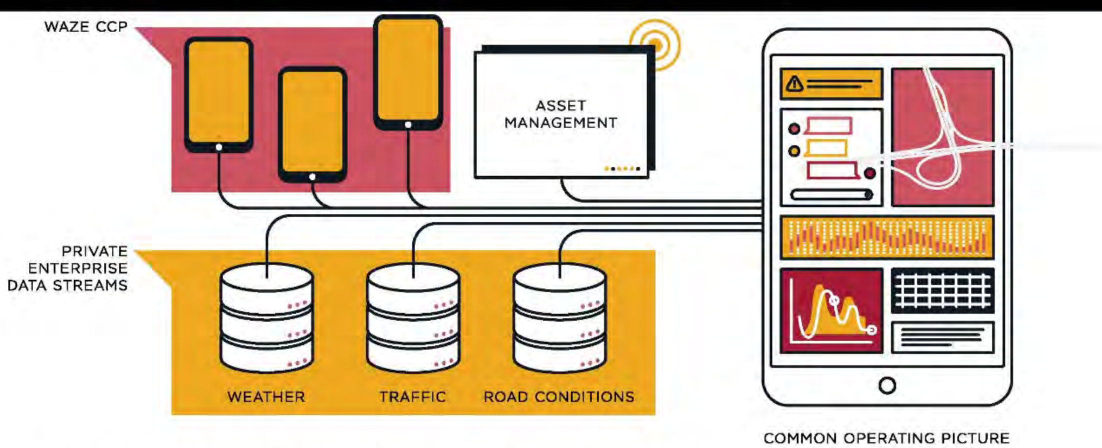
MARYLAND DEPARTMENT OF TRANSPORTATION
STATE HIGHWAY ADMINISTRATION

TRADITIONAL, BIG DATA & CROWD SOURCED APPLICATIONS – SAME DATA STREAMS, MANY USES..

- **Real time applications**
  - System Monitoring
  - Incident Mgmt.
  - ATM/ ICM
  - Maintenance/ SOGR
- **Archived data applications**
  - Planning/ TSMO
  - Asset Management
- **Performance Management**

9

MAINSTREAM USE OF CROWD SOURCED DATA



Use Waze CCP type of data to support TIM, ATM, ICM applications

Asset Conditions Monitoring, Situational Awareness, Customer Requests

10

CHALLENGES

- Data sources and accuracy may not serve multiple needs (Planning, Ops, TSMO etc.)
- Data remains siloed with different levels of maturity for various program areas
- Lack of uniformity in reporting (sometimes vary by entity/ geography)
- Lack of awareness of existing data leads to data redundancy (multiple entities collect same/ similar information)

11

CHALLENGES contd…

- Data collection methods have changed but business processes/ SOPs have not been updated.
- Lack of enterprise systems for "apples to apples" comparison of transportation decisions (capital vs ops, SOGR vs new projects, bridges vs pavements)
- Traditional DOT workforce and procurement methods do not reflect the new data management needs adequately.

12

SOLUTIONS/ OPPORTUNITIES

Basics Questions:

- Is the data informing decisions?
  - Strategic / Tactical/ Operational
- Do we have the business processes and technology infrastructure?
- Can investments in data be justified and sustained?

1. **VALUABLE**
2. **AVAILABLE**
3. **VALIDATED**
4. **SECURED**
5. **CLEAR**
6. **EFFICIENT**
7. **ACCOUNTABLE**

AASHTO Reference: https://data.transportation.org/

Need for Data Governance
Data Management & Data Business Plan

TREAT DATA AS AN ASSET

13

MDOT DATA GOVERNANCE & DATA HUB

- Where Do We Want to Be?
- What Did We Learn?
- What Do We Need to Do?
- How Can We Monitor and Evaluate Data Governance Implementation?

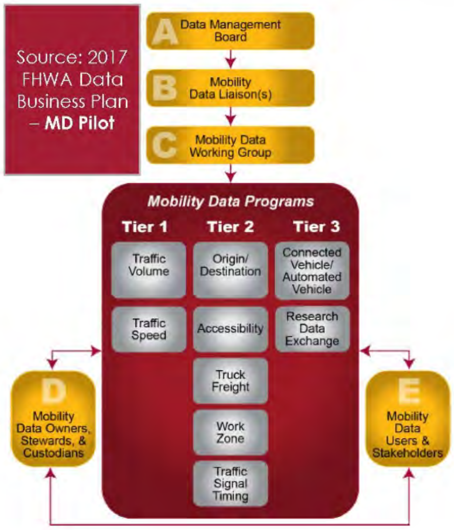
People
- Staff resources and training
- Organization
- Culture

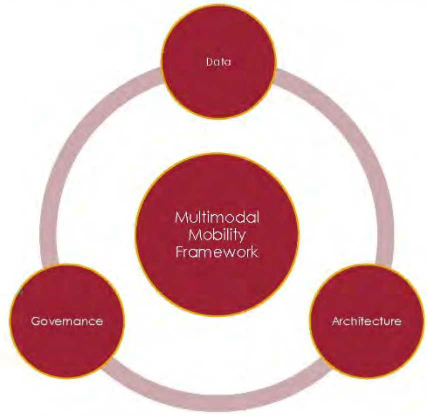
Processes
- Procedures and documentation
- Accountability and incentives

Technologies
- Data assets
- Information systems
- Tools for Data Governance implementation

14

MDOT SHA TSMO DATA GOVERNANCE EXAMPLE





Interaction, structure, and components to integrate and report on mobility data

15

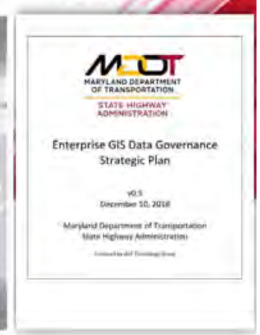
MD & MDOT SHA GEOSPATIAL DATA GOVERNANCE & MANAGEMENT

Policies, standards and procedures to collect, manage, disseminate, utilize, and archive enterprise data and related applications.

- **Data Development**
  - MDOT SHA Geospatial Data Integrity-Data Submission Policy & Procedure
  - Maryland iMap Data Submission Policy
- **Data Sharing & Security**
  - DoIT Account Management Policy
- **Application Development**
  - MDOT SHA Standard Web Map Configuration in AGOL
  - MDOT SHA Web App Documentation in AGOL
- **Enterprise GIS Data Inventory**
  - Documented in standardized template

MD iMAP
Maryland's Mapping & GIS Data Portal

Maryland's OPENDATA Portal



IN SUMMARY…

- Transportation decision-making will require a combination of traditional sources, big data and crowd sourced data. Comprehensive business processes and Data Hub models **still in infancy**.
- Needs champions, investment and **commitment for mainstreaming**
- **Collaboration** between agencies (DOTs, MPOs), Data Providers and Research Community needed
- **Need for Advanced research** to combine traditional data sources with big data for travel behavior analysis
- Investments in data has to demonstrate the **value added for sustained investments**

17

QUESTIONS?
THANK YOU!

10

CONTACT INFORMATION

SUBRAT MAHAPATRA
Deputy Director, TSMO & CATS
Office of Transportation Mobility & Operations (OTMO)
MDOT State Highway Administration

(410) 582-5569
smahapatra@mdot.maryland.gov

MDOT MARYLAND DEPARTMENT OF TRANSPORTATION STATE HIGHWAY ADMINISTRATION

New York State Department of Transportation (NYSDOT)

NEW YORK STATE OF OPPORTUNITY. Department of Transportation

Applied Data Governance in the Development and Management of NYSDOT's System of Engagement

Maps and Apps Library

1

2

NYSDOT Maps and Apps: A System of Engagement

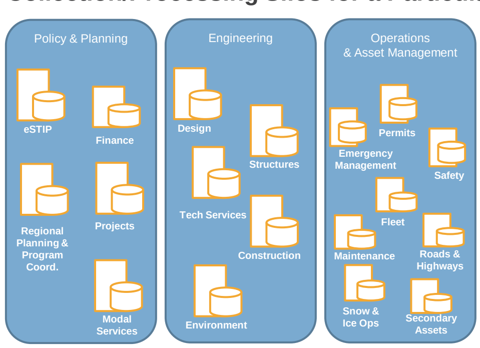
- **Goal**: Make Commonly Requested Data Widely Available
- **Key Topics:**
  - System of Record (SoR) vs. System of Engagement (SoE)
  - Reusable Web Services
  - Viewers, Dashboards, Operations, and Mobile Collector Apps
  - Map-based Access to Tabular Data

NEW YORK STATE OF OPPORTUNITY. Department of Transportation

3

System of Record (SoR) vs System of Engagement (SoE)

- **SoR - Data Collection/Processing Silos for a Particular Program Area**



NEW YORK STATE OF OPPORTUNITY. Department of Transportation

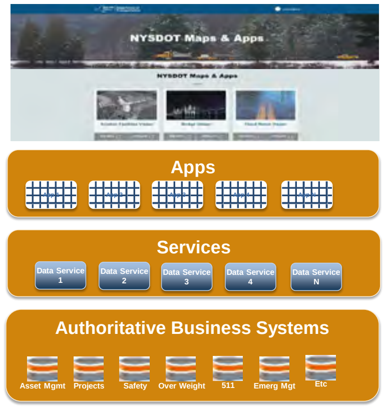
The SoE is a System of Systems

A Single Destination – manage user identities – access on any device – collect new data – share with others

Suite of apps. Apps consume one or more services

Reusable services to read/write data to/from data domains

Maximize the value from existing IT investments and data governance



NEW YORK STATE OF OPPORTUNITY. | Department of Transportation

Data Web Services – the "Legos"

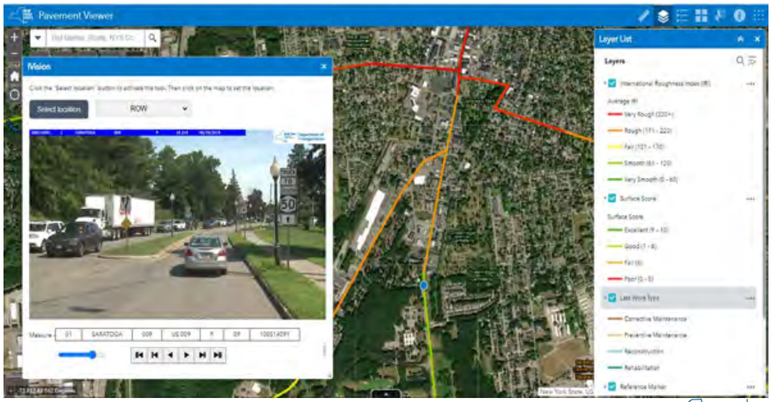
- **Reusable Web Services are the building blocks of your apps**
- **Develop an inventory of specific data elements to be used in the app**
- **Determine the "authoritative source" for each data element**
- **Understand the limitations of various data elements**
- **Align data sets to the location. Where that isn't possible with the base data, find common keys to make the connection**
- **Use "crosswalks" to translate and standardize data fields between systems**

NEW YORK STATE OF OPPORTUNITY. | Department of Transportation

Reusable Web Services

NEW YORK STATE OF OPPORTUNITY. | Department of Transportation

NYSDOT Maps & Apps – Pavement Viewer



NEW YORK STATE OF OPPORTUNITY. | Department of Transportation

SoE – Designed to Meet a Particular Business Purpose

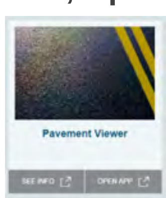
- **User Story:** *As a Resident Engineer, I need to be able to evaluate Safety information and answer potential questions posed by the public, executives, or other stakeholders about the location of future, current, and past safety related projects, investigations, and maintenance activities. I need to be able to view the following layers on a map at my computer or on my mobile device connected to the web:*
  - **Projects** - Planned Projects; current year, last 4 years, and next 4 years
  - **Maintenance** - Anything Safety related in the Last 2 years
  - **Enterprise Asset Management Program** - Network Master
  - **Safety** - 3 years of Raw Crash Data, HALS, PILS/SILS, SDL, PIES
  - **Intersections** - Signalized Intersection Data
  - **Roadway Information** - Number of Lanes, Speed Limit, Functional Class, Shoulder Type, Route ID, Median Type, Shoulder Width, Lane Width

  - We need to be able to navigate and search the map by the following:
    - **PIN**
    - **BIN**
    - **Route**
    - **County**
    - **Municipality**
    - **Reference Marker**
    - **New York State Place Geocoder**

NEW YORK STATE OF OPPORTUNITY. | Department of Transportation

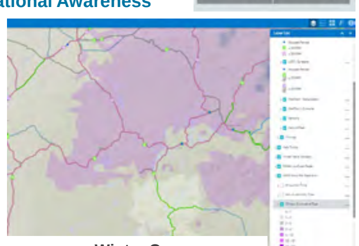
Viewers, Dashboards, Operations & Mobile Collector Apps

Viewers:
- **Map-centric**
- **Maximize query/analytical capabilities**

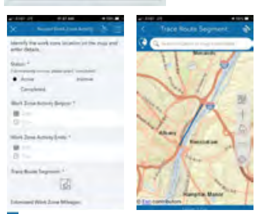


Dashboards:
- **Data-centric**
- **Visualize status/performance**

Operational Apps:
- **Situational Awareness Tool**



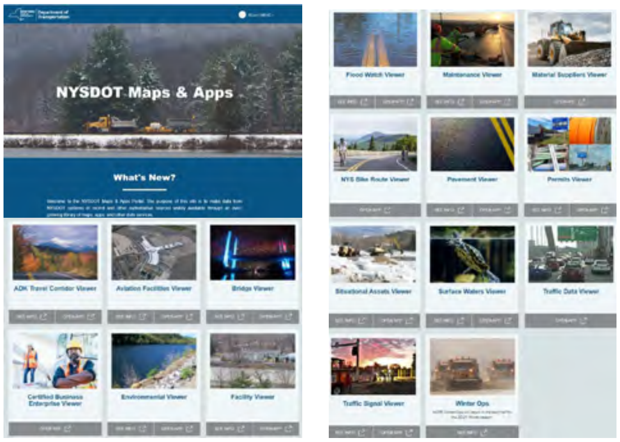
Winter Ops



Active Work Zones Collector

NEW YORK STATE OF OPPORTUNITY. | Department of Transportation

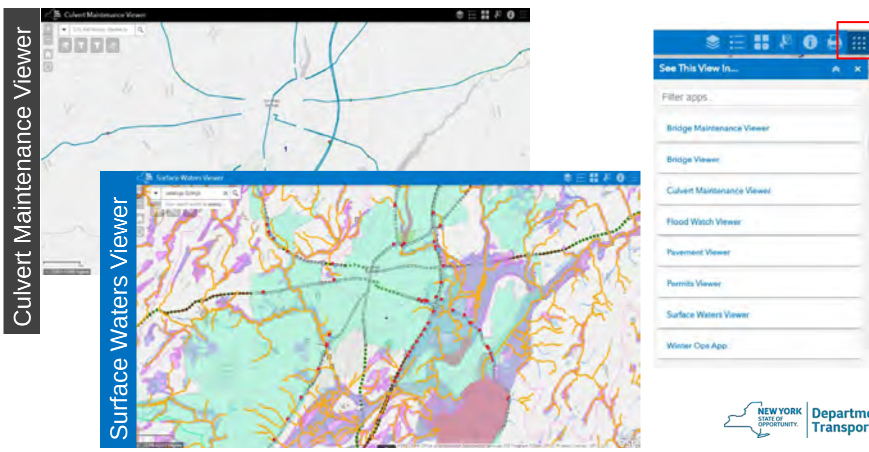
System of Engagement (SoE)



The purpose of this site is to make data from NYSDOT systems of record and other authoritative sources widely available through an ever-growing library of maps, apps, and other data services.

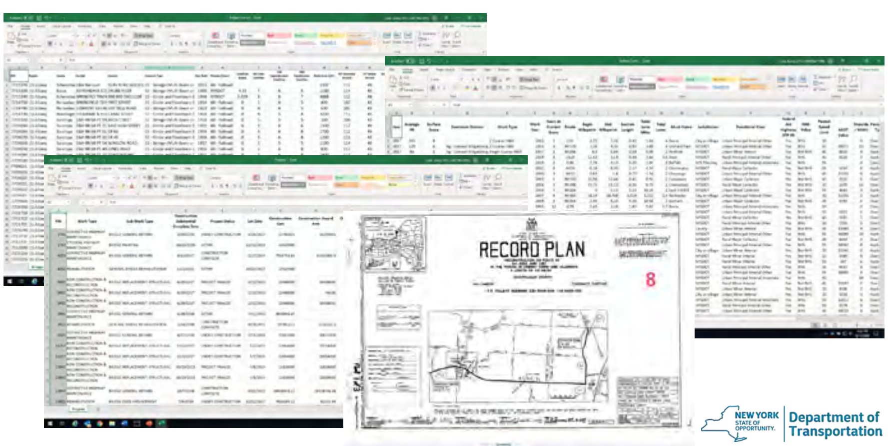
NEW YORK STATE OF OPPORTUNITY. Department of Transportation

NYSDOT Maps & Apps – "*The Hopper*"



NEW YORK STATE OF OPPORTUNITY. Department of Transportation

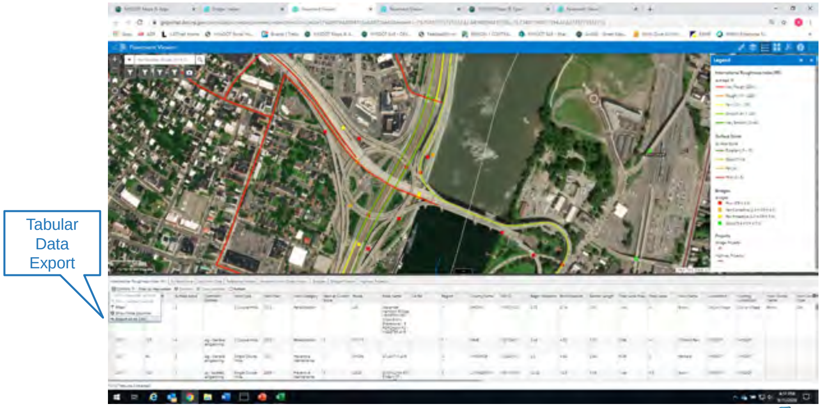
Paradigm Shift – From Tabular Asset Management Data



NEW YORK STATE OF OPPORTUNITY. Department of Transportation

13

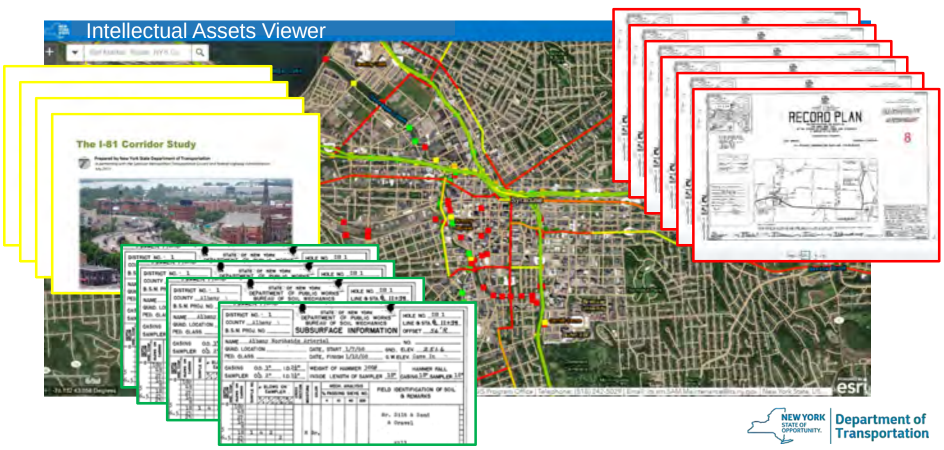
Paradigm Shift - To Geospatial Asset Management Data



NEW YORK STATE OF OPPORTUNITY. Department of Transportation

14

Unstructured Data on a Map (*Conceptual*)



NEW YORK STATE OF OPPORTUNITY. Department of Transportation

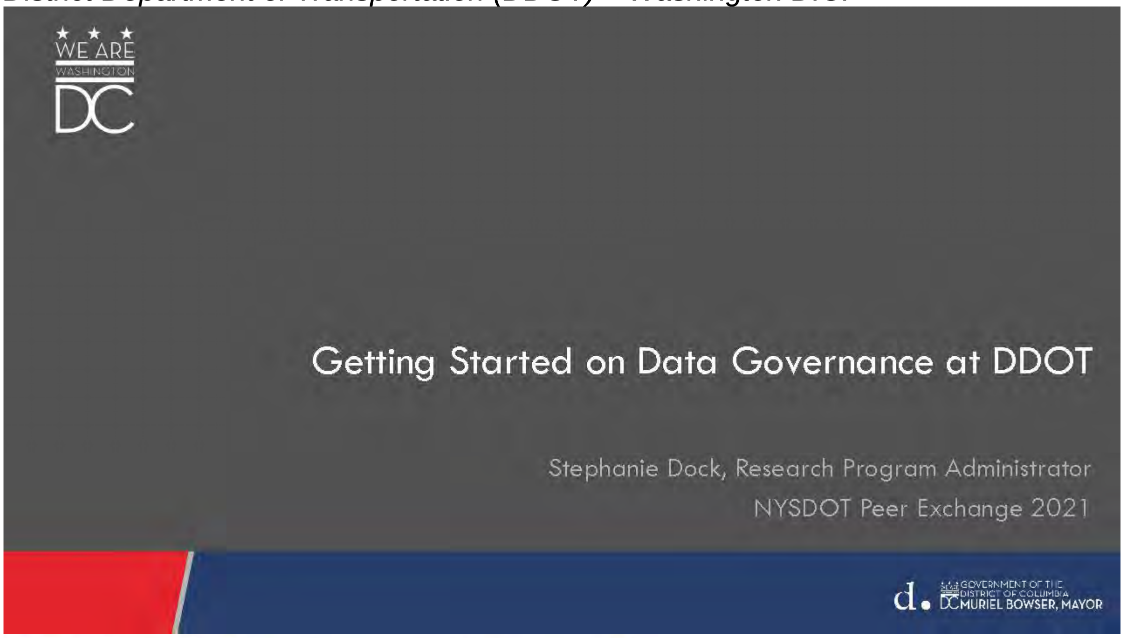
15

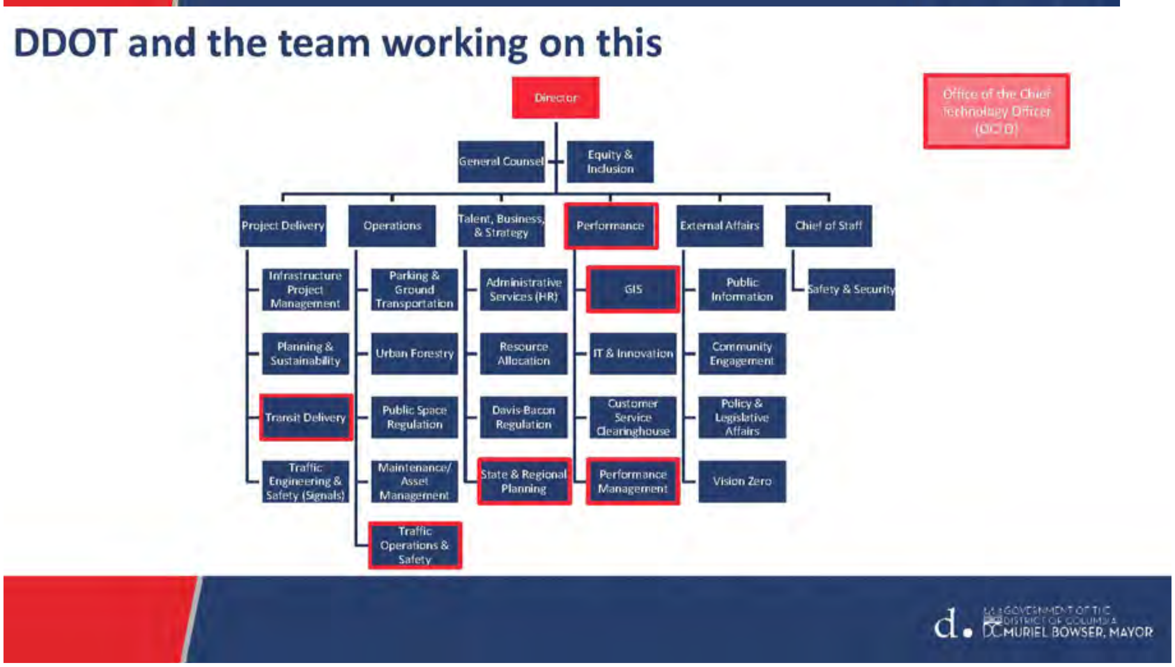
Summary

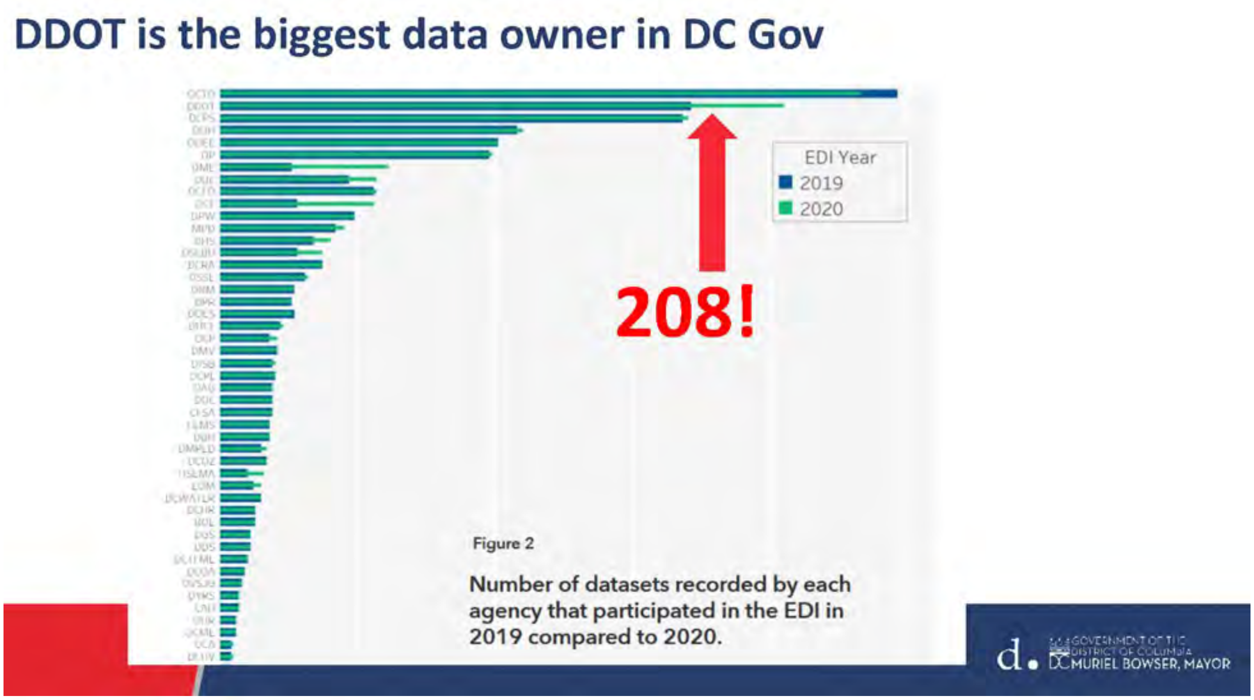
- SoE development is done through a cooperative governance structure
- SoE is a system of systems built on reusable web services
- Ability to give data back to end users in a meaningful way
- Unstructured Data – Geolocate and make available on maps
- Moving all asset management data from tabular to geospatial

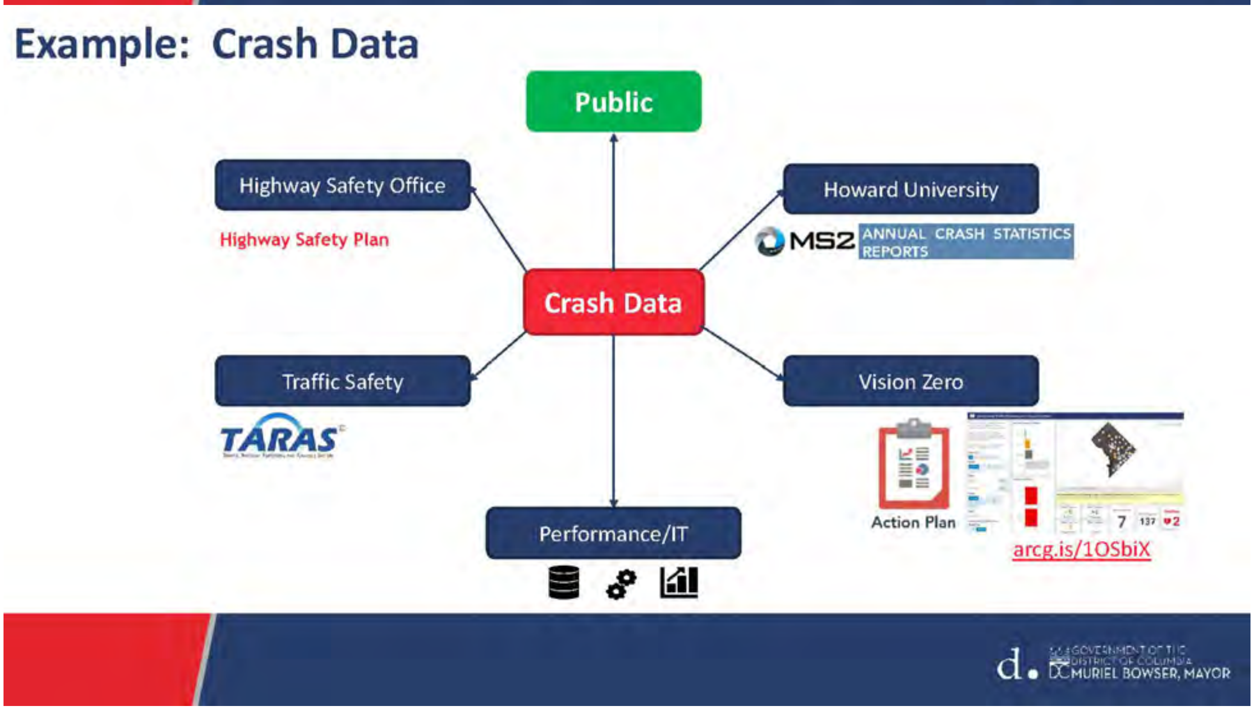
NEW YORK STATE OF OPPORTUNITY. Department of Transportation

District Department of Transportation (DDOT) – Washington D.C.

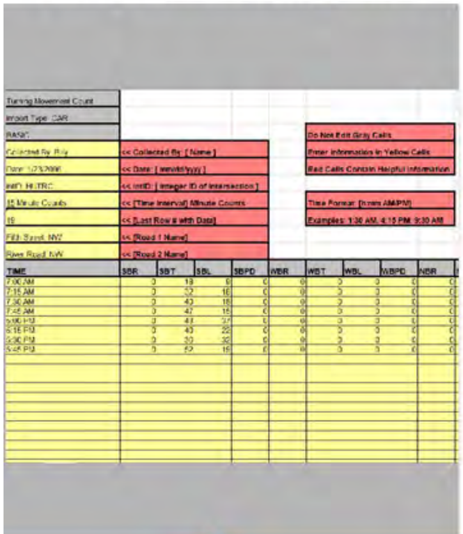





DDOT is the biggest data owner in DC Gov
EDI Year
2019
2020
208!
Figure 2
Number of datasets recorded by each agency that participated in the EDI in 2019 compared to 2020.
MURIEL BOWSER, MAYOR


Example: Crash Data
Public
Highway Safety Office
Highway Safety Plan
Howard University
MS2
ANNUAL CRASH STATISTICS REPORTS
Crash Data
Traffic Safety
TARAS
Vision Zero
Action Plan
arcg.is/1OSbiX
Performance/IT
MURIEL BOWSER, MAYOR

Example: Turning Movement Counts



Turning Movement Count	
Import Type: CAR	
BASIC	Do Not Edit Gray Cells
Collected By: [illegible]	<< Collected By: [Name]
Date: 1/23/2006	<< Date: [mm/dd/yyyy]
intID: [illegible]	<< intID: [integer ID of intersection]
15 Minute Counts	<< [Time Interval] Minute Counts
19	<< [Last Row # with Data]
Fifth Street NW	<< [Road 1 Name]
River Road NW	<< [Road 2 Name]

Do Not Edit Gray Cells	
Enter information in Yellow Cells	
Red Cells Contain Helpful Information	
Time Format: [h:mm AM/PM]	
Examples: 1:30 AM, 4:15 PM, 9:30 AM	

TIME	SBR	SBT	SBL	SBPD	WBR	WBT	WBL	WBPD	NBR
7:00 AM	0	18	6	0	0	0	0	0	0
7:15 AM	0	32	16	0	0	0	0	0	0
7:30 AM	0	43	18	0	0	0	0	0	0
7:45 AM	0	47	15	0	0	0	0	0	0
5:00 PM	0	43	37	0	0	0	0	0	0
5:15 PM	0	43	22	0	0	0	0	0	0
5:30 PM	0	30	32	0	0	0	0	0	0
5:45 PM	0	52	15	0	0	0	0	0	0

- Lots of teams are collecting this data
- No centralized repository
- Lots of (potential and real) duplication
- People cannot use data if they do not collect it themselves

d. GOVERNMENT OF THE DISTRICT OF COLUMBIA DC MURIEL BOWSER, MAYOR

Poor Governance Slows Us Down

COVID-19 Dashboard

- Had to ask individual programs about access to their data
- Several systems do not allow for automatic updates (e.g. bicycle counters)
- Data that is not readily available can be missed, reinforcing biases in our reporting

d. GOVERNMENT OF THE DISTRICT OF COLUMBIA DC MURIEL BOWSER, MAYOR

What is Data Governance?

- Data Governance is about establishing processes for effective data management through various standards and policies.
- Encompasses People, Process and Technology to achieve improved data accessibility, quality and accountability.
- It also involves setting strategies so that data and data systems can "talk" to each other.
- **Knowing your data and the users is the first step in the process.**

Benefits of Data Governance

- Less confusion
- Less duplication
- Retention of institutional knowledge
- Better **data-driven** decision-making
- Efficiency - less staff time spent hunting down data, leverage existing resources, know what data we need

Our Initial Goal: Accessibility

- Making data easy to find
- Can **search and discover** data

When a project kicks off, a manager will know what to do with their data from planning and collection through storage and sharing

Note: while data quality is important, this effort is not focused on fixing individual teams' data. We are looking at documentation first and foremost – and a place to add all those caveats about the data.

GOVERNMENT OF THE DISTRICT OF COLUMBIA
MURIEL BOWSER, MAYOR

Step 1: Document the Data

- DDOT Data Wiki houses information on all the datasets we could identify
  - Still have some work to do on one-off datasets (e.g. research projects)
- Intention is that users can maintain their own information
- Platform: Atlassian Confluence


Data Details Summary

Sidewalk Data	In Progress	DDOT	2016	GIS	ArcGIS Server	One-Time
Waze Traffic Data	Original	External	[illegible]	[illegible]	[illegible]	Every 2 mins
Private For-Hire Vehicle Trips (TNCs)		DFHV				Quarterly
Private For-Hire Vehicle Trips (TNCs)		Shared Streets	select weeks from Sept 2017 - June 2019			sporadic
[illegible]	Original	DDOT	2018-Current	CSV	[illegible] Shared Drive	Daily
Speed Limits	Original	DDOT GIS Team	Current	GIS	ArcGIS Server	Daily

GOVERNMENT OF THE DISTRICT OF COLUMBIA
MURIEL BOWSER, MAYOR

A Well-Managed Data Set

Pavement Condition Index

Created by [illegible], last modified on Sep 16, 2019

Data collection of pavement condition assessment of the streets within the District in both directions. This work is critical for the development of the annual Capital Improvement Program and reporting the annually required Federal Highway Performance Monitoring System. Information to be collected includes, distress data, ride quality, rutting, and cross-slope, curve, & longitudinal grade data on a block-by-block basis.

2019 data to be updated soon, contracted out to a vendor.

Data Details

Data Source	DDOT
Update Frequency	One-Time
Format	GIS
Date Range	2017
Location	ArcGIS Server
Data Condition	Final

Contacts

Data Owner	Edward Carpenter

Public Reporting

The data is not publicly reported.

Data Maturity Score

Accessibility	3
Storage	4
Integration	2
Relevance and Sufficiency	4
Quality	3
Collection Frequency	1
History	2
Privacy	1
Documentation	1
Utilization	2
Value	4

Data Access Details

The data is hosted on an ArcGIS server.

Server	[redacted]
Service/ Database	SQL Server
Username	[redacted]
Password	[redacted]

ArcGIS Pro needed to access the data. Request an account here.

Data Issues

- Last updates in 2017.

d. GOVERNMENT OF THE DISTRICT OF COLUMBIA DC MURIEL BOWSER, MAYOR

Data Initiatives

Goal	Internal: Improve the planning for bicycles and decision-making processes External: Better support bicycling as a mode of travel and communication to public stakeholders.		
Capabilities	Bike project planning; Public information on route planning; Provide bike lane/trails data		
People	Data Owner(s), Producer(s), Steward(s)	Active Transportation Branch staff, GIS team	
	Consumer(s)	DDOT Planners, DDOT Data Analysts, Regional Governmental Partners, DDOT Permit Analysts, General Public	
Processes	**MAJOR TASKS** • Identify (and fill) bicycle network gaps • Improve placement of bicycle corrals • Bicycle lane work plan, bike facility mileage reports • Determine maintenance responsibilities for trails • Manage bike count program	**ANALYSES** • Perform LTS analyses to determine connectivity • Perform network analyses on bicycle network • Connect to vehicle counting data for analysis • Capital Bikeshare location analysis • Bike data for Citywide Bikemap print product	
Technology & Data	• Bikeways and Trails data stored in GIS • Counter data stored in EcoCounter cloud • Manual counts data stored in Shared Drive & OpenData		
Challenges	**Issue**	**Problem Type**	**Data Investments**
	Bike data difficult to analyze alongside other DDOT data.	Quality	Integrate with LRS
	Communicate to external mapping providers (Google maps, OSM, etc)	Access	Create data-driven notifications workflow with PIO

d. DC MURIEL BOWSER, MAYOR

Where are we going?

- Common data templates and centralizing those datasets
  - E.g. turning movement count template
- Brainstorm on how we institutionalize this, so that we are not left with a stale resource
  - training for staff
  - linkages to project initiation and close out?
- Collaborate with data owners
  - How to improve aspects data quality, access, storage etc.
  - Getting involved in new projects for better data management techniques
- Hire a full-time staffer to support this!

d. GOVERNMENT OF THE DISTRICT OF COLUMBIA DC MURIEL BOWSER, MAYOR

Moving into Big Data Systems

- We have a data lake through our central IT department
  - On-premise solution built using the Cloudera instance of Hadoop
  - Had a workshop with Esri's big data expert to strengthen the linkages there
  - Still a work in progress
- We are big into LRS
  - Really fascinating work going on with our refreshed centerline (we know what *all* the lanes are!)
  - Pushing for things like Shared Streets referencing to allow for "conflate once"
- Research is very interested in datasets that benefit from the data lake structure
  - Private for-hire vehicle reporting (Uber, Lyft, Via)
  - Dockless micromobility data via the mobility data specification
  - Inrix TripsXD data we purchased several years ago

d. GOVERNMENT OF THE DISTRICT OF COLUMBIA DC MURIEL BOWSER, MAYOR


d.
District Department of Transportation
GOVERNMENT OF THE DISTRICT OF COLUMBIA
DC MURIEL BOWSER, MAYOR

AEM Corporation

Roadmap to Big Data

Steps to implement a data lake and cloud computing pilot to demonstrate the benefits for transportation applications

Kelley Pecheux, AEM Corporation
Benjamin Pecheux, AEM Corporation

NYSDOT Peer Exchange
February 3, 2021

Overview of Presentation

What is data-driven decision-making?

Progress in an activity is compelled by data not by intuition, personal experience, or political agenda. Data are now too large, fast, and change too quickly for the latter.

- Briefly revisit NCHRP 08-116 Guidebook (NCHRP Research Report 952)
- Dive into the 8-step Roadmap for Managing Data from Emerging Transportation Technologies
- Discuss big data concepts (e.g., data lake, cloud, distributed computing)
- Show how agencies can iteratively develop and demonstrate the benefits of a modern, big data approach to data management, analysis, and decision-making

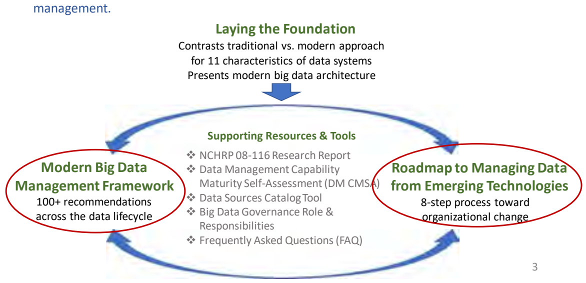
aem

2

The NCHRP 08-116 Guidebook Can Help Agencies Shift Toward the Modern Data Management Approach

NCHRP Report 952: Guidebook for Managing Data from Emerging Transportation Technologies

- Provides a *modern big data management framework* that introduces new concepts and methodologies, best practices, and 100+ recommendations for managing data in a modern, flexible, scalable, and sustainable way.
- Lays out a *roadmap* on how to begin to shift – technically, institutionally, and culturally – toward effectively managing data from emerging technologies.
- Provides *examples and references* of transportation agencies currently exploring or already navigating the implementation of big data, including their challenges and successes.
- Discusses *common misconceptions* within the transportation industry regarding big data management.



3

Roadmap to Managing Data from Emerging Transportation Technologies

- Step 1 – Develop an understanding of big data
- Step 2 – Identify a use case and an associated pilot project
- Step 3 – Secure buy-in from at least one person from leadership for the pilot project
- Step 4 – Establish an embryotic big data test environment/ playground
- Step 5 – Develop the big data project within the playground
- Step 6 – Demonstrate the value of the data to other business units
- Step 7 – Demonstrate the value of the data to executive leadership
- Step 8 – Establish a formal data storage and management environment

aem

Step 1 - Develop an Understanding of Big Data

- What is big data?
- Big data characteristics
- Big data concepts
- When to pursue big data
- Common misconceptions regarding big data
- Agency Case study – *The Importance of Understanding Big Data*
- Additional resources

BIG DATA CONCEPTS

- Data lake
- Cloud
- Distributed computing
- Distributed storage
- Nonrelational databases
- Common big data analytics techniques

aem

Step 2 - Identify a Use Case and an Associated Pilot Project

- Select a use case and pilot project that:
  - Aligns with business unit, leadership, and organizational goals
  - Includes drivers for change
  - Identifies big data source(s) of interest
- Engage others in the cause:
  - Internal to the business unit
  - Cross-business unit
  - Junior and mid-level staff
  - External partners
- Agency case study – *Portland Urban Data Lake Pilot Project (PUDL)*

Start small but think BIG.

DelDOT conducted a 30- to 60-day pilot project with a few selected datasets as a proof of concept for a new cloud-based solution. This pilot project duplicated critical data from on-premise storage to the cloud for more efficient access and use.

aem

Step 3 – Secure Buy-In from at Least One Person from Leadership for the Pilot Project

- Establish and communicate the value proposition for the pilot project:
  - Example projects
  - Value propositions
  - Questions to assist in developing the "pitch"
- Ways to create a sense of urgency and a fear of missing out (FOMO)
- De-risk the decision by identifying and communicating risks and other potential barriers up front
- Know how and when to make the pitch

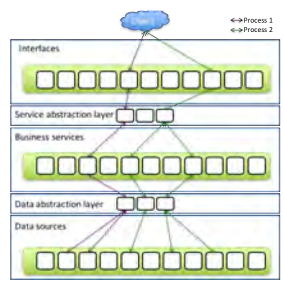
aem

7

Step 4 – Establish an Embryotic Big Data Test Environment or Playground

- Establish buy-in from IT
  - Potential challenges and barriers
  - Pros and cons of on-premise versus cloud storage
- Create the playground (see "Store" section of Framework)
  - Data storage layer
  - Data processing layer
- Agency case study – *On-Premise vs Cloud*

Understanding the concept and utilization of a cloud storage layer (i.e., data lake) is extremely important at this stage. Agencies that do not fully grasp the need for storing raw, unprocessed, data are far more prone to encounter mistakes or pitfalls as the architecture and use cases mature over time.



Cloud storage services have allowed DelDOT to more easily integrate their data sources (which was a challenge to do in-house), store all the data into a single repository so end-users can access them directly, and leverage cloud tools that they otherwise would not have access to.

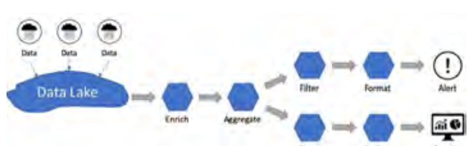
aem

8

Step 5 – Develop the Pilot Project Within the Big Data Test Environment/Playground

- Develop/ensure the availability of the right expertise (e.g., training/hiring in-house staff, trusted contractors and university partners, and big data experts/consultants)
- Apply a data science perspective
  - Collect raw data
  - Process and clean the data
  - Perform exploratory data analyses
  - Build data science pipelines
- Iteratively develop and improve the project and the associated outputs/data products

Reference the Modern Data Management Framework section of the Guidebook for more details.



Once established, this data pipeline will automatically pull in the appropriate raw data, transform it as necessary, and apply pre-defined and pre-established data analytics techniques to develop the end data product.

aem

9

Step 6 – Demonstrate Value of Data to Other Business Units

- Build support for the data and project across the organization, including others that may have an interest in the data, project, and data products for their own business areas
- Use the data to craft a compelling story of success using understandable and persuasive visualizations that tie the insights uncovered in the data to the ability to address an issue or solve a problem of the business unit
- Get others involved in sharing and using their data within the test environment – iteratively expand the use of the data to improved, enhanced, and new use cases
- Agency case study – *Iterative Growth and Success*

"Everyone who has interacted with our system has had something to say about it to someone."
– Kentucky Transportation Cabinet

aem

Step 7 – Demonstrate Value of Data to Executive Leadership

- Present the success stories/business case to executives
- Continue to build support, foster data sharing, and grow iteratively and incrementally
- Push for organization change/adoption of a formal big data environment
- Agency case study – *Buy-In from Executive Leadership*

aem

Step 8 – Establish a Formal Data Storage and Management Environment

- Establish a clear vision and goals
- Make data accessible yet secure
- Integrate at the data level
- Use data to make decisions
- Merge existing projects into the same data infrastructure
- Continue to seek input from other stakeholders
- Iterate on evolving data governance plans and procedures
- Seek continuous improvement
- Agency case study – *Continued Room for Growth*

A data-driven culture requires data openness, letting teams access data and consider new data-driven approaches.

aem

Next Steps for NCHRP 08-116

- AEM recently awarded an implementation project that will:
  - Identify 3-4 agencies interested in learning more about / implementing modern data management approaches to support decision-making
  - Developing and delivering training to these agencies
  - Providing tailored implementation support to participating agencies
  - Conducting a peer exchange amongst participating agencies
  - Conducting a national webinar to share experiences and lessons learned
- The project is to start in Feb/Mar 2021
- Provides opportunity for states to get free training and technical assistance

aem

13

Kelley Pecheux, Sr. Director, Transportation
kelley.pecheux@aemcorp.com
(703) 350-8487

Benjamin Pecheux, Director of Information Science
ben.pecheux@aemcorp.com
(703) 989-4776

Contact Information

aem

BIG DATA ANALYTICS AND ARTIFICIAL INTELLIGENCE FOR SMART CITY TRANSPORTATION (FOR TRADITIONAL, CONNECTED AND AUTONOMOUS VEHICLES)

Sanjay Ranka
Professor
Department of Computer and Information Science and Engineering
sranka@ufl.edu
352 514 4213

Work supported by
Florida/US Department of Transportation (Central Office, District 5)
National Science Foundation (Smart Cities, Cyber Physical Systems)

Acknowledgements

FDOT

Tom Byron
April Blackburn
Raj Ponnaluri
Trey Tillander
Fred Heery
Darryll Dockstader
John Krause
David Sherman
Jerry Scott
Ed Hutchinson

FDOT D5

Jeremy Dilmore
Tushar Patel
Clay Packard
Claudia Paskauskas
Charles Wetzel
Jorge Barrios
Chad Dickson

City of Gainesville

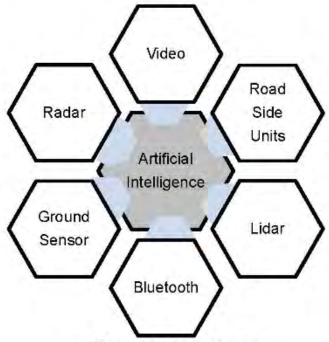
Emmanuel Posadas
Daniel Hoffman

National Science Foundation

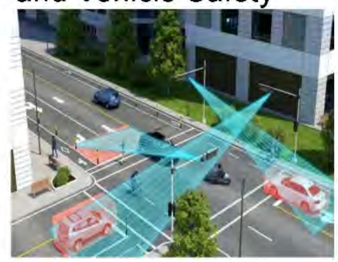
Jonathan Sprinkle
David Corman

Artificial Intelligence to Improve Safety and Operations

Improving Signal and Network Operations



Improving Pedestrian and Vehicle Safety

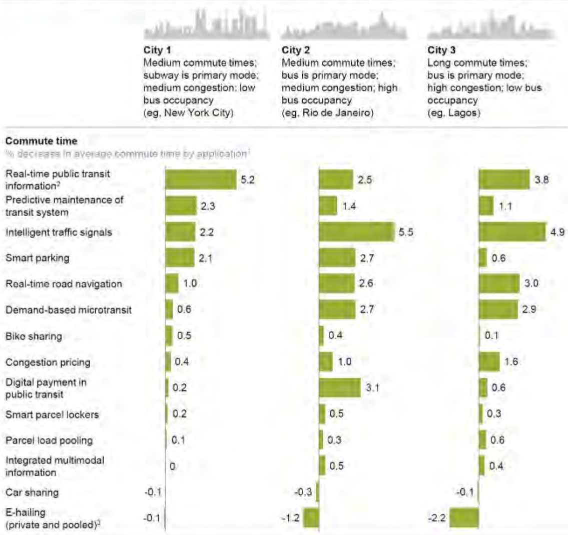


Generative Adversarial Networks
Reinforcement Learning
Long Short Term Memory Networks
Semantic Segmentation
Point Cloud Processing

Simulator Data

GPU Computing
Edge Computing
Cloud Computing
Large Scale IOT Data Management

Intelligent Traffic Signals: Demonstrated impact on mobility



Improvement of overall traffic flow through dynamic optimization of traffic lights and speed limits, leading to higher average speeds on roads and less frequent stop-and-go conditions.
Includes traffic light preemption technology, which gives priority to emergency vehicles, public buses, or both.

Source: Mckinsey Global Institute Report on Smart Cities: Digital Solutions for a more livable future, 2018

What is needed for successful application of AI

Transportation Infrastructure for Collecting Real Data and Testing New Methods

Simulator Infrastructure to generate large amounts of labeled data that is realistic as well as generating counterfactuals

PyTorch

TensorFlow   Keras

Software Infrastructure for Deep and Machine Learning

Hardware Infrastructure (Edge, GPU and Cloud)

Storage Infrastructure

Generative Adversarial Networks
Reinforcement Learning
Long Short Term Memory Networks
Semantic Segmentation
Point Cloud Processing

Algorithmic Infrastructure

Objectives: Smarter Intersections, Streets, Network and CAVs

Smarter intersections will be equipped with cameras

- Better Pedestrian and Bicyclists Safety by examining the conflict points of the vehicle/pedestrian trajectories.
- Better Demand Profiles by understanding real-traffic behavior for more effective signal timing
- Better SPaT messaging by understanding traffic behavior on the intersections.

Smarter Network by sensing data from multiple intersections will enable

- Better Incident Detection for alleviating traffic backups and secondary crashes.
- Better Signal Retiming for different corridors by time of day and day of the week.
- Better System-wide Network Utilization by utilizing a global view of the entire network.

Smarter Streets by use of transit buses and other vehicles with forward facing cameras

- Better Pedestrian Safety by detection of jaywalking/ “mid-block crossing” behaviors that are prevalent.
- Better Resource Management by understanding usage of street parking and signage.
- Better Lane and Street Sign Design by tracking indicator lights and lane-changing maneuvers.

Smarter Interactions with Connected and Autonomous Vehicles will lead to

- Improved Safety by effectively managing CAVs.
- Better Understanding of CAV Behavior and Interaction with Traditional Vehicles.
- Better Signal Timing and Trajectory Optimization for CAVS.
- More Accurate Trend Measurements of CAV Penetration.

Transportation Infrastructure: Trapezium

New Assets

Upgraded Linux based ‘ATC’ Controllers for ATSPM Data

Siemens DSRC Radios with MAP and SPaT Broadcast

Emergency Vehicle Pre-emption and vehicle OBUs (WIP)

Multiple Fisheye Cameras

60+ Onboard Units in Transit Buses and Public Vehicles

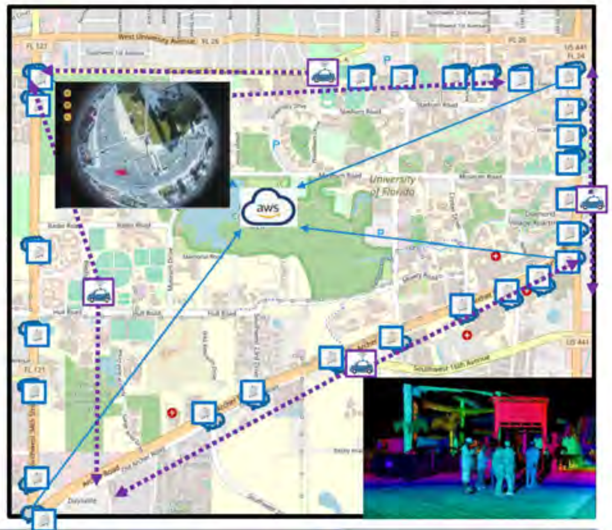
Other Assets

Connectivity: Fiber Optic Gigabit at each signalized intersection w/ 12 port Ethernet

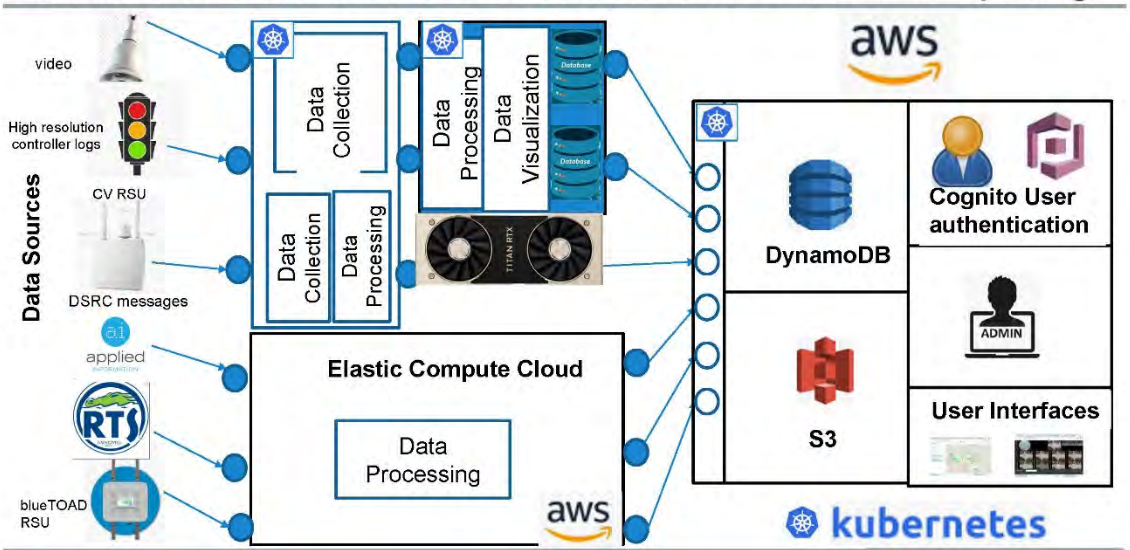
Basic Video Monitoring: Bosch PTZ CCTV at each Signal

Controllers: Linux Based ATCs

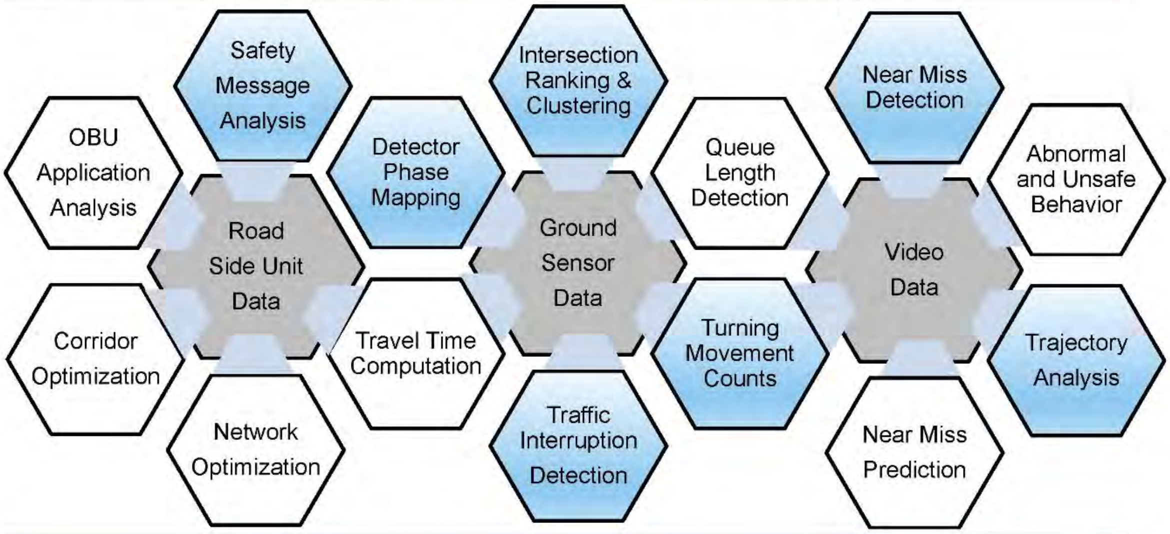
Travel Time: Segment Bluetooth sensor



Software and Hardware Infrastructure for Traffic Computing



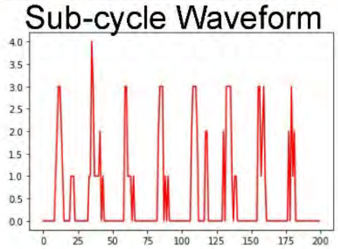
Application Infrastructure = Traffic Data + AI Algorithms

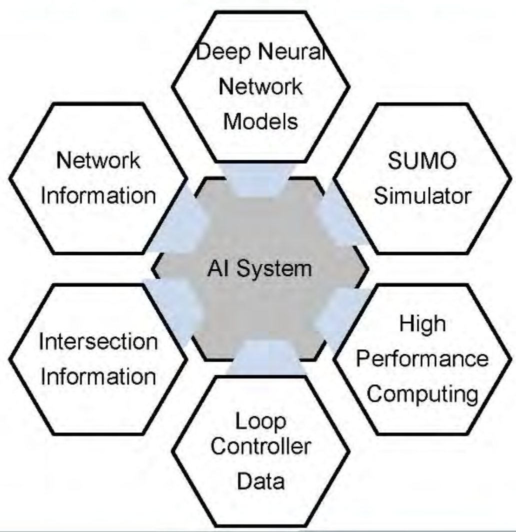


Multiple Sensors are Better than One (20 GBs per day)

AI based system for High Resolution Loop Controller Data

Sub-cycle Waveform





Loop Controller Data: Measures of Effectiveness

Split Failures (Measure of Demand): When the queue is not emptied till the end of green time for a cycle

These vehicles will wait for the next cycle

Arrivals on Red/Green (Measure of Signal Timing Effectiveness): Ratio of vehicles arriving on red versus green

These vehicles will arrive on a red traffic light

10 Terabyte of detector data – 300+ signals over 6 months.

	Split Failures High	Split Failures Low
Arrivals on Green/Red High	Capacity	Ok
Arrivals on Green/Red Low	Potential Timing: High Demand	Potential Timing: Low Demand

City Wide Traffic Intersection Analysis

High Resolution Controller Logs and Meta-data (Detector Mappings, locations)

Terabytes of data per year.

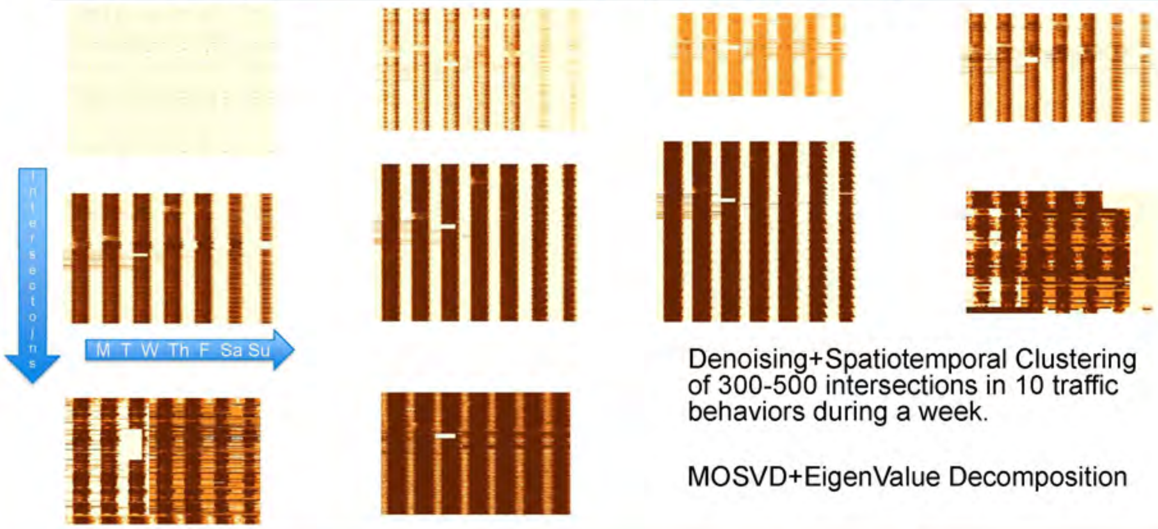
Winner, Big Data Challenge on Signalized Intersections, 2019

Compute relevant Measures of Effectiveness (Split Failures, AOR/AOG) → Machine Learning For Intersection Ranking → Machine Learning for Similar Behavior → Machine Learning for Change Detection → Machine Learning For Interruption Detection

ML techniques are uses to provide a traffic engineer key characteristics of traffic behavior in a city and potential changes over time.

Used for signal timing improvements and long terms planning

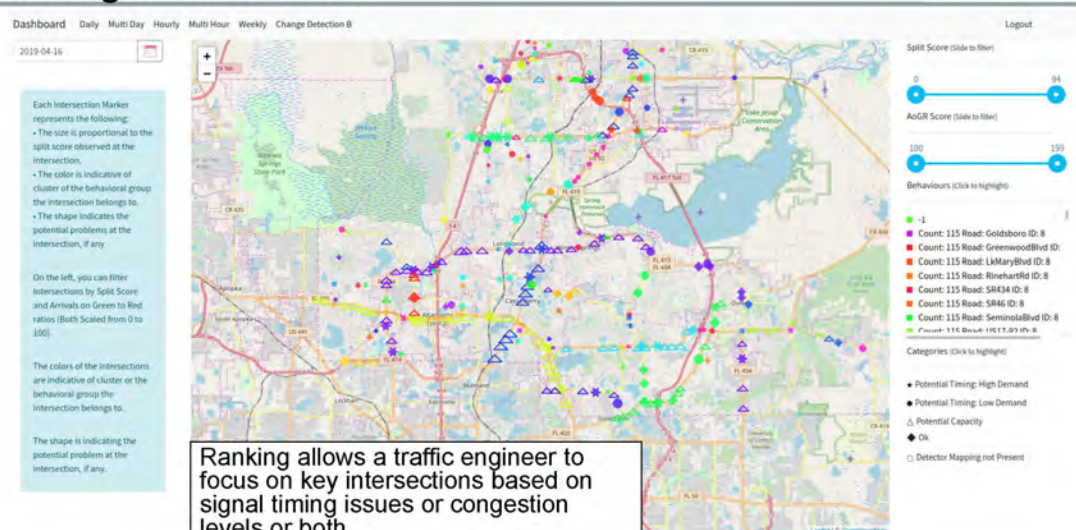
Representative Intersection Behaviors (Hourly)



Denoising+Spatiotemporal Clustering of 300-500 intersections in 10 traffic behaviors during a week.

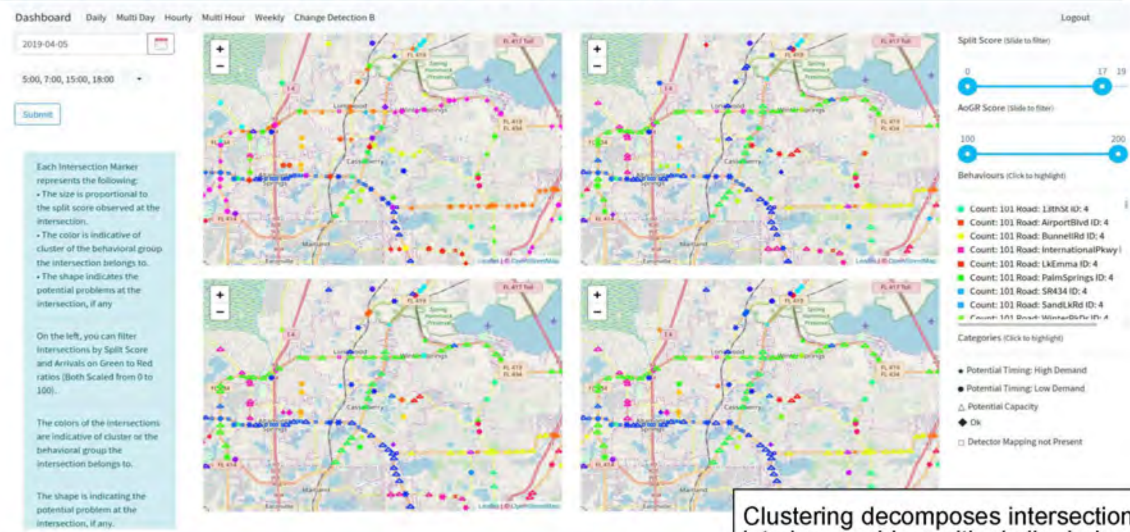
MOSVD+EigenValue Decomposition

Ranking Intersections



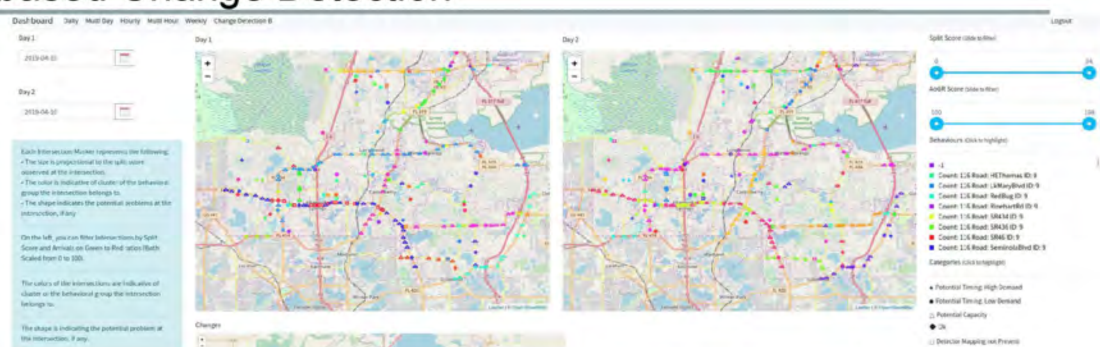
Ranking allows a traffic engineer to focus on key intersections based on signal timing issues or congestion levels or both

Multi Hour (Peak versus Non-Peak)

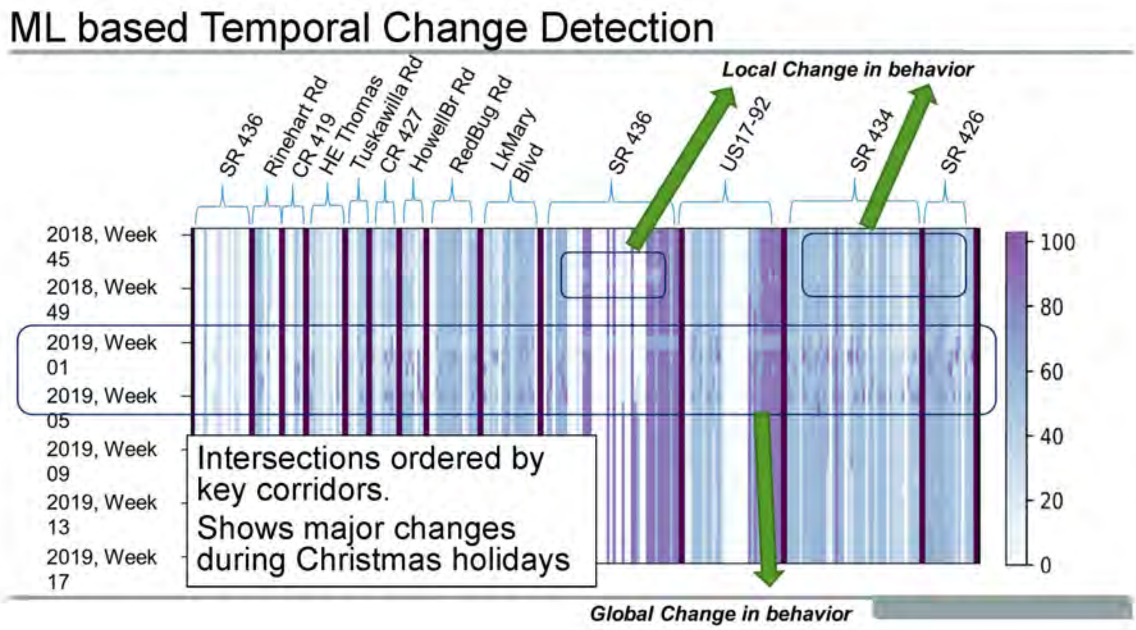


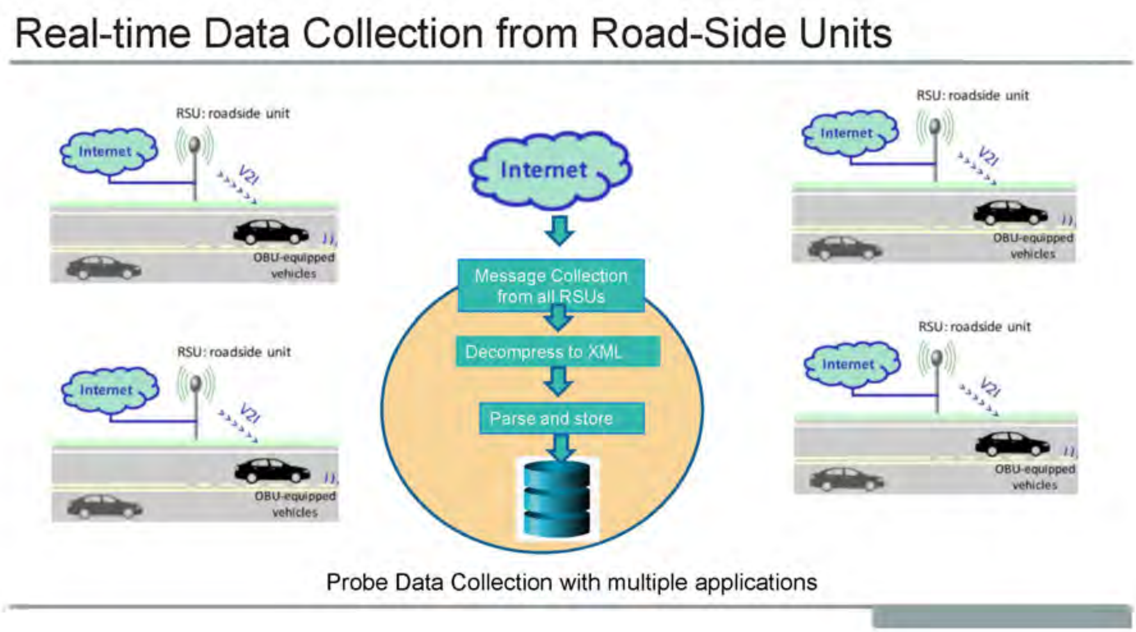
Clustering decomposes intersections into key corridors with similar behavior during multiple hours of the day

ML based Change Detection

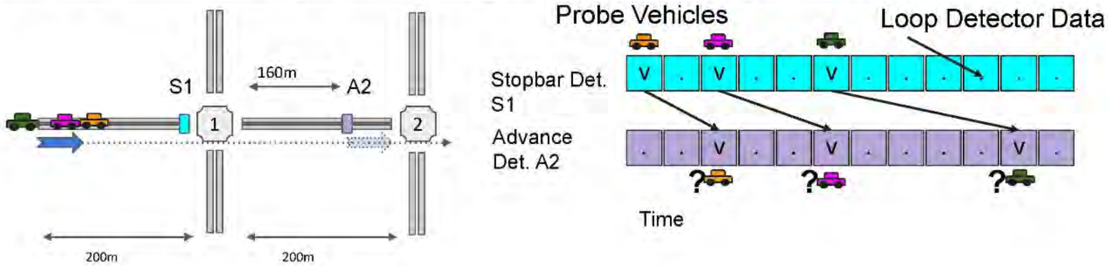


Statistically Significant Change Detection provides intersections where congestion levels have changed


ML based Temporal Change Detection
Local Change in behavior
SR 436
Rinehart Rd
CR 419
HE Thomas
Tuskawilla Rd
CR 427
HowellBr Rd
RedBug Rd
LkMary Blvd
SR 436
US17-92
SR 434
SR 426
2018, Week 45
2018, Week 49
2019, Week 01
2019, Week 05
2019, Week 09
2019, Week 13
2019, Week 17
Intersections ordered by key corridors.
Shows major changes during Christmas holidays
100
80
60
40
20
0
Global Change in behavior


Real-time Data Collection from Road-Side Units
RSU: roadside unit
Internet
V2I
OBU-equipped vehicles
Internet
Message Collection from all RSUs
Decompress to XML
Parse and store
Probe Data Collection with multiple applications

Probe Data: Sequence Alignment for Travel Time



Real-time travel time computation on a corridor is key to determine signal timing and congestion issues.

Travel Time is inversely proportional to velocity

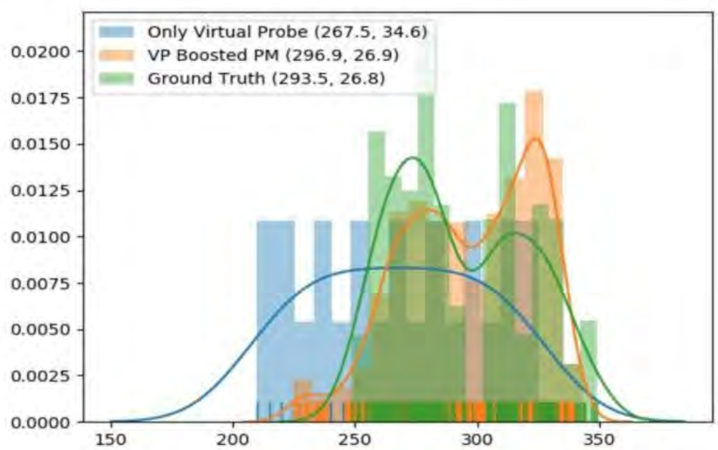
Our techniques combines genomic sequence matching algorithms with probe data to determine accurate travel time computations.

Results: Travel Time Distribution

Ground Truth (Green)
Multiple Virtual Probes (Blue)
Multiple Virtual Probes with Platoon Matching (Orange)

Simulated 8 Intersection Corridor

	Mean of Corridor Travel Times	Standard Deviation
Sequence Alignment	199.1s	50.9s
Ground Truth	211.6s	34.6s



Intersection Traffic Video Analysis

Generate rich spatial and temporal road user features using Convolution Neural Networks (CNNs) with real-time performance

Multi-Object Tracking

Object Detection + Track Handling and State Estimation + Data Association ⇒ **Trajectory analysis** +

Spatial feature

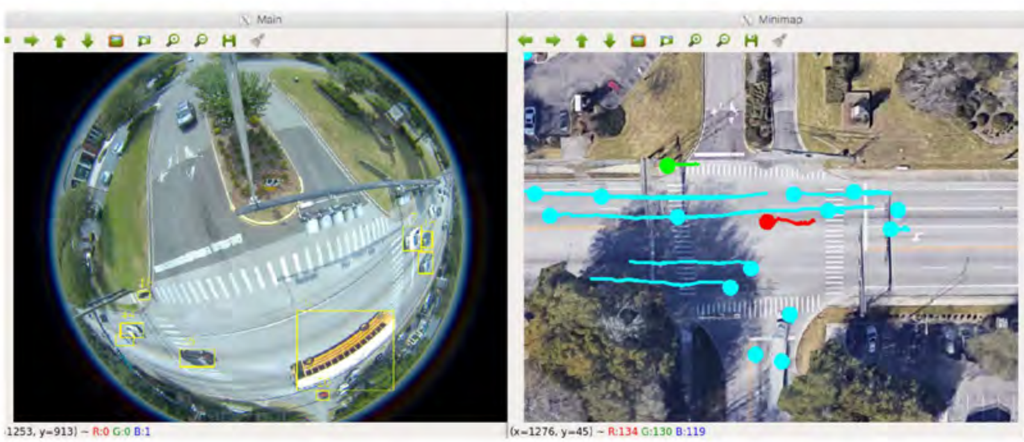
- Localization: position, size
- Recognition: road user type

Temporal feature

- motion, speed

40 fps

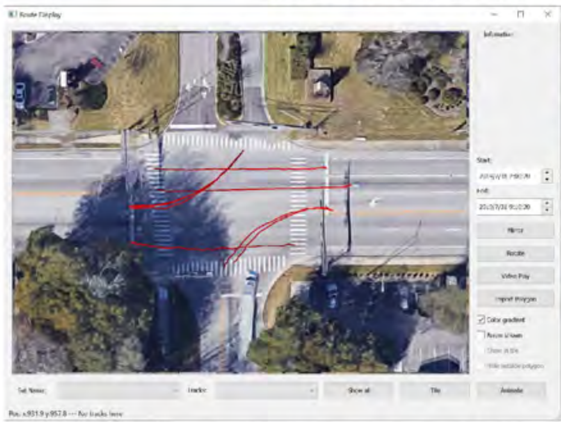
Real-time Detection & Tracking



Mini Map Trajectory Demo
Cyan: car **Green: pedestrian Yellow: bus**

Red Light Violations

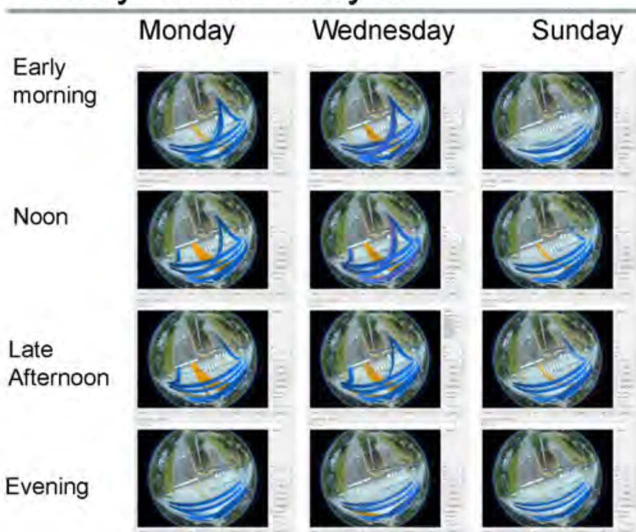
Mining 25000+ tracks per day per video per intersection



Tracks representing vehicle crossings on red light

Tracks changing lanes at an intersection

Weekly Trend Analysis

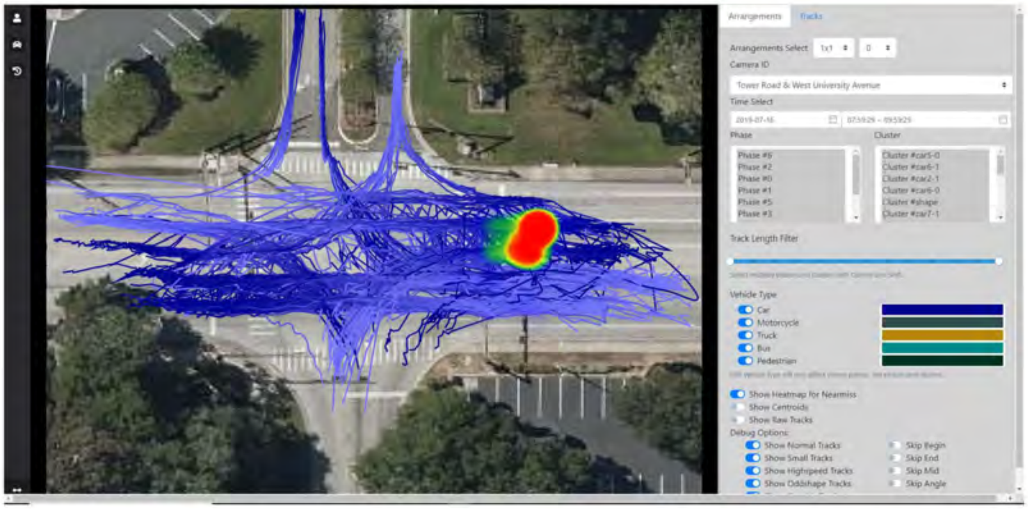


Higher Volume of orange tracks seem to be present during working hours, suggesting the presence of an office complex near the top of the image.

Late start and end on Monday as compared to Wednesday.

Higher northbound traffic in the morning and southbound traffic in the evening, suggesting presence of a residential area in the south.

TrafficVML: Heatmap (risk areas)

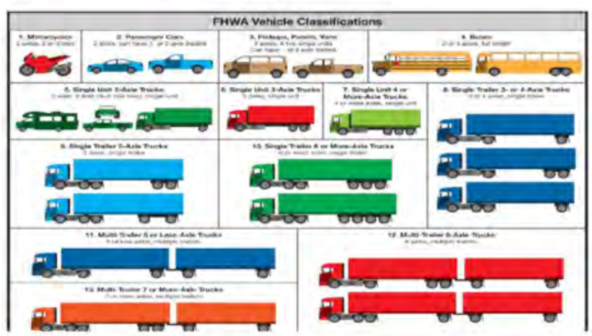


Video

Commodity Classification Using Machine Learning

Taxonomy + Commodity + Video + Machine Learning =

Configuration	Body type	Typical commodities	Typical industries
Five-axle tractor semitrailer, 3-S2 (59%)	Vans/reefers (63%)	• Palletized cargo • Refrigerated goods	• Retail • Produce
Six-axle tractor semitrailer, 3-S3 (19%)	Flat decks (16%)	• Equipment • Building supplies	• Construction • Manufacturing
Nine-axle turnpike double, 3-S2-4 (8%)	Hoppers (6%)	• Grain • Granular fertilizer	• Agriculture
Eight-axle B-train double, 3-S3-S2 (7%)	Tankers (4%)	• Petroleum products • Chemicals	• Petroleum • Chemical
	Dumps (6%)	• Aggregate • Grain • Refuse	• Construction • Agriculture
	Containers (2%)	• Palletized cargo • Freight of all kinds	• Retail



ML based Trailer Classification and Logo Recognition

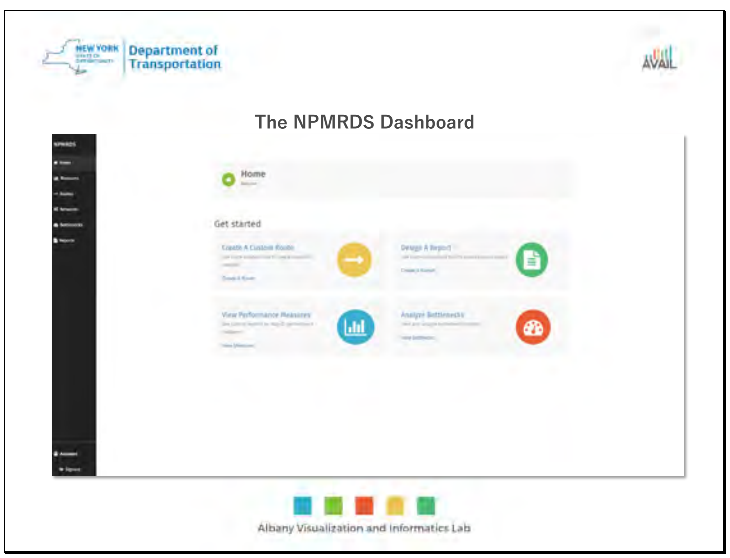
Logo Recognition – Text Detection

The developed algorithms can achieve a high recall (80+ %) with a competitive recognition accuracy.

Conclusions

- Rich transportation datasets – unique requirements for processing and storage
- Use of AI/ML for intelligent transportation in its infancy
- Demonstrated use of AI/ML to solve real-world applications
- Making headway towards Vision Zero Goals
- Collaboration with city and state transportation personnel bringing great dividends

State University at Albany Visualization and Informatics Lab (SUNY AVAIL)



Department of Transportation

NPMRDS = National Performance Management Research Dataset

- NPMRDS is an aggregated dataset made by the company HERE until Feb 2017, now aggregated by INRIX.
- Provided by FHWA at no cost to states.
- Based on passenger probe data obtained from a number of sources, including mobile phones, vehicles, and portable navigation devices, gathered in 5 minute intervals
  - NHS Highways are broken into segments, called “TMCs” based on navigational software company (Garmin/TomTom) needs.
- **Big Data: Terrabites in size**

Albany Visualization and Informatics Lab

Department of Transportation

Source of NPMRDS Probe Speed Data

- GPS
- Phone

Albany Visualization and Informatics Lab

Department of Transportation

Performance Measures Map-21 (PM3) Reporting

Albany Visualization and Informatics Lab

NEW YORK STATE OF OPPORTUNITY Department of Transportation

AVAIL

Multi-Geographic Resolution:

Segment || Route || Multi-Route Corridor || Network

Creation and editing tools make your geographies fully customizable

Albany Visualization and Informatics Lab

NEW YORK STATE OF OPPORTUNITY Department of Transportation

AVAIL

NPMRDS Tools for PM3 Reporting

Albany Visualization and Informatics Lab

NEW YORK STATE OF OPPORTUNITY Department of Transportation

AVAIL

Map-21 Performance Measure Scores by Segment Level of Travel Time Reliability in Syracuse

Albany Visualization and Informatics Lab

Department of Transportation

AVAIL

Leveraging NPMRDS for uses beyond PM3

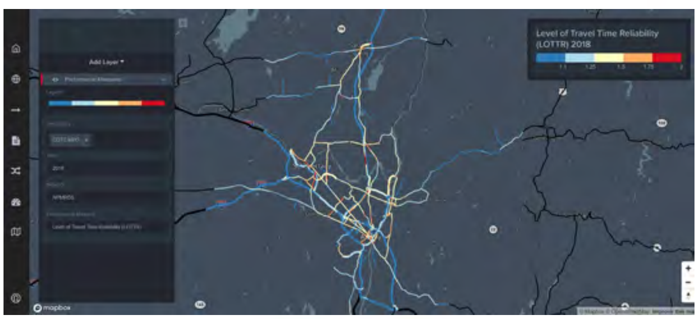
- Congestion Management Planning
- Corridor Analysis
- Network Analysis
- Bottleneck Identification
- Project Prioritization
- Post-Project Analysis
- Incident Post-mortem Analysis

Albany Visualization and Informatics Lab

Department of Transportation

AVAIL

Macro Tools for Congestion Management Planning



Albany Visualization and Informatics Lab

Department of Transportation

AVAIL

Performance Measure Scores by Segment CSV and Shapefile Downloads

Albany Visualization and Informatics Lab

Department of Transportation

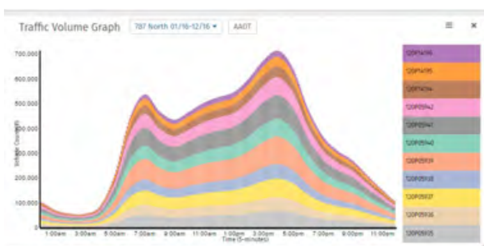
Data Integration

AVAIL has begun to integrate many other geo-spatial transportation datasets through conflation of the NPMRDS shapefile to the LRS and HPMS shapefiles

Albany Visualization and Informatics Lab

Department of Transportation

Data Integration



AADT and Short Counts

Albany Visualization and Informatics Lab

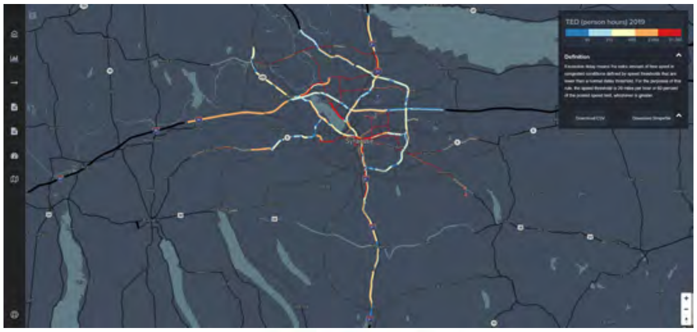
Department of Transportation

Expanded Data

Albany Visualization and Informatics Lab

Department of Transportation

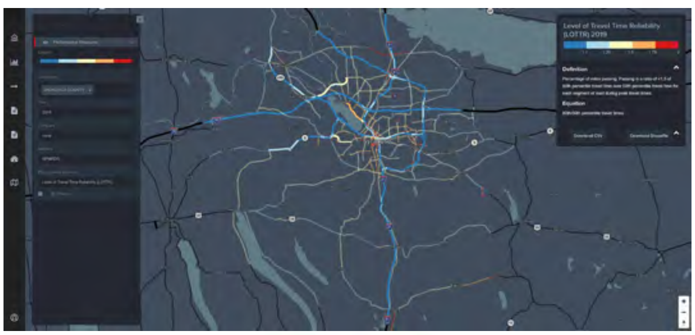
Onondaga County Without Expanded Network



Albany Visualization and Informatics Lab

Department of Transportation

Onondaga County With Expanded Network



Albany Visualization and Informatics Lab

Department of Transportation

Onondaga County Expanded Network

Before/After

Albany Visualization and Informatics Lab

Department of Transportation

Examples of NPMRDS Based Analysis

Albany Visualization and Informatics Lab

Department of Transportation

Corridor Analysis Tools
Penn Station Amtrak Track Work

Comparing traffic on the LIE entering Manhattan from June and July of 2017 to June and July of 2018

Albany Visualization and Informatics Lab

Department of Transportation

Corridor Analysis Tools
Penn Station Amtrak Track Work

Speeds and Travel Times were stable even though a significant number of public transportation commuters were forced to find other means of travel.

Department of Transportation

AVAIL

Corridor Analysis Tools
Penn Station Amtrak Track Work

Route Compare Component

ROUTE NAME	SPEED	TRAVEL TIME	HOURS OF DELAY	DATA QUALITY
July 2017	32.6 MPH	57.86 Minutes	14908749 Hours	97 Percent of Epochs Reporting
June 2017	31.05 MPH ▼ 4.75%	69.94 Minutes ▲ 3.07%	15107382 Hours ▲ 1.33%	96 Percent of Epochs Reporting ▼ 1.03%
July 2018	30.88 MPH ▼ 5.28%	69.54 Minutes ▲ 2.48%	15443572 Hours ▲ 3.59%	98 Percent of Epochs Reporting ▲ 1.03%
June 2018	31.31 MPH ▼ 3.96%	69.52 Minutes ▲ 2.45%	14749893 Hours ▼ 1.07%	98 Percent of Epochs Reporting ▲ 1.03%

Albany Visualization and Informatics Lab

Department of Transportation

AVAIL

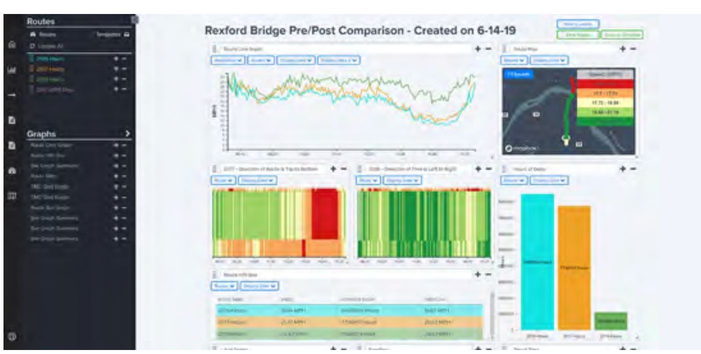
Rexford Bridge Pre/Post Construction

Albany Visualization and Informatics Lab

Department of Transportation

AVAIL

Corridor Analysis Tools
Rexford Bridge Open July 2017

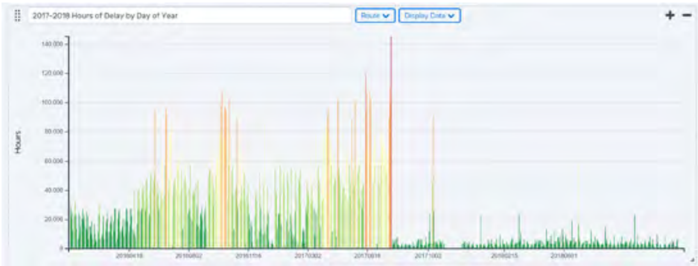


Hours of Delay have decreased dramatically since the opening of the new Rexford Bridge

Department of Transportation

AVAIL

Corridor Analysis Tools
Rexford Bridge Open July 2017



The corridor analysis tools provide a variety of data resolution tools. Here we see hours of delay for every day since January 1, 2016 on the X-Axis. Hours of delay decreased significantly after the new bridge was opened.

Albany Visualization and Informatics Lab

Department of Transportation

AVAIL

Corridor Analysis Tools
Rexford Bridge Open July 2017

The corridor analysis tools provide a variety of data resolution tools. Here we see hours of delay for every day since January 1, 2016 on the X-Axis. Hours of delay decreased significantly after the new bridge was opened.

Albany Visualization and Informatics Lab

Department of Transportation

AVAIL

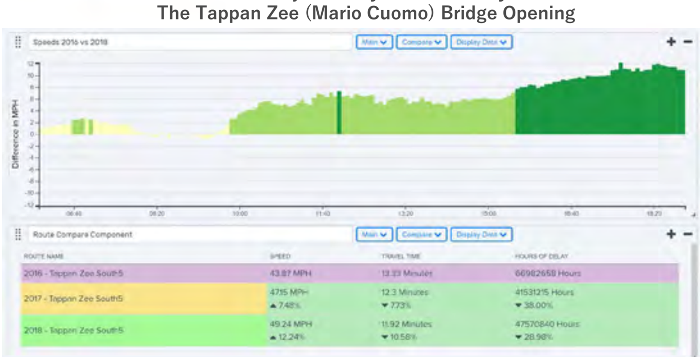
Post Project Analysis Case Study:
The Tappan Zee (Mario Cuomo) Bridge Opening

Here we see Speeds before and after the opening of the final span of the Mario Cuomo Bridge

Post Project Analysis Case Study:
The Tappan Zee (Mario Cuomo) Bridge Opening

Here we see Speeds before and after the opening of the final span of the Mario Cuomo Bridge

Post Project Analysis Case Study:
The Tappan Zee (Mario Cuomo) Bridge Opening



Albany Visualization and Informatics Lab

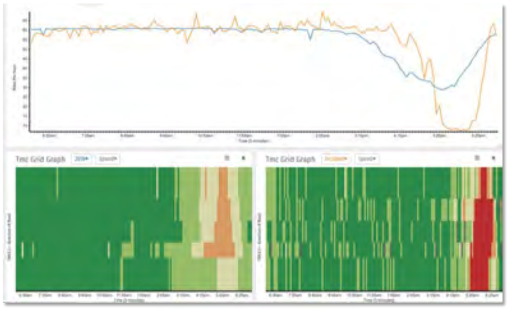
Corridor Analysis
I-87 Northbound to the Twin Bridges

Incident at the Twin Bridges

Albany Visualization and Informatics Lab

Department of Transportation

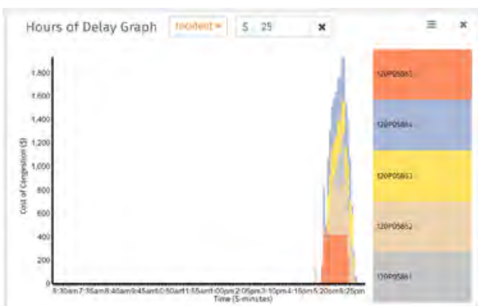
Incident Analysis



This case study shows an incident on I-87 Northbound at the Twin Bridges from November 21, 2016

Department of Transportation

Incident Analysis



Hours of Delay and Cost of Delay associated with the selected incident

Albany Visualization and Informatics Lab

2021 NYSDOT Data and Research Peer Exchange Webinar

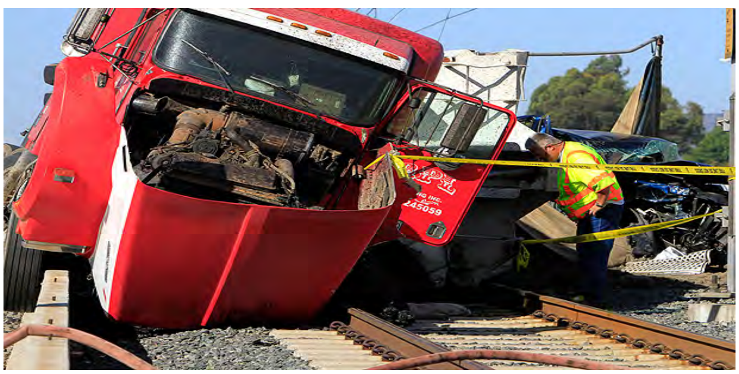
Open source and collaborative software development. Models for supporting development and integration of new functions and capabilities of evolving analytics.

TSMO at Caltrans

02/03/2021

Nick Compin Ph.D
Statewide TSMO, Connected Corridors and System Performance Measurement/Data Lead
Caltrans Headquarters' Division of Traffic Operations
nicholas.compin@dot.ca.gov
(916) 653-4575

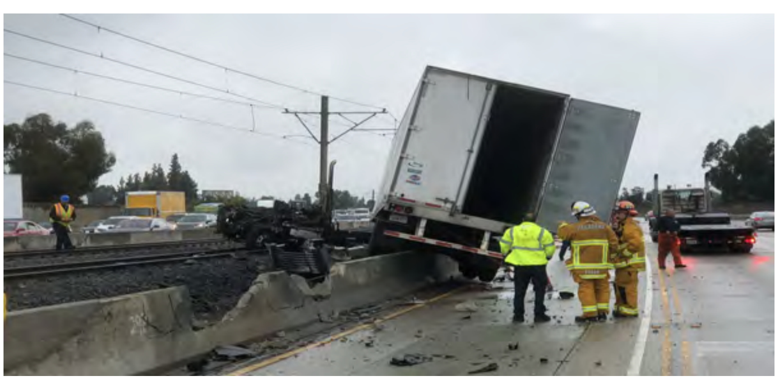
Recurrent Operations Challenge - I-210 Pasadena, CA



October 2010: Toyota Camry vs. Chevy Pickup vs. Big Rig

2

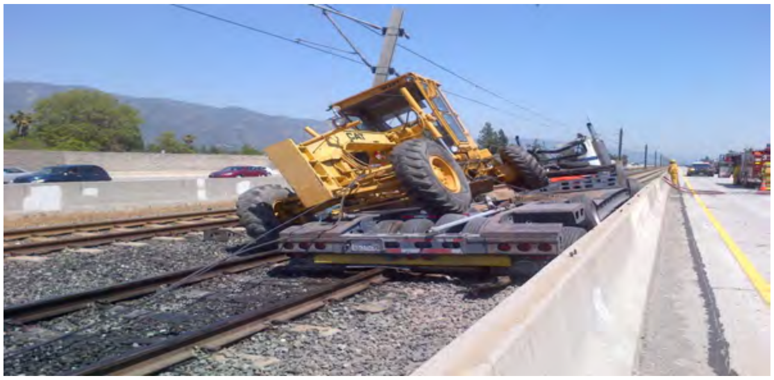
Recurrent Operations Challenge - I-210 Pasadena, CA



April 2014 Big Rig vs. Big Rig Source: Caltrans and Metro

3

Recurrent Operations Challenge - I-210 Pasadena, CA



April 2014 Big Rig vs. Big Rig Source: Caltrans and Metro

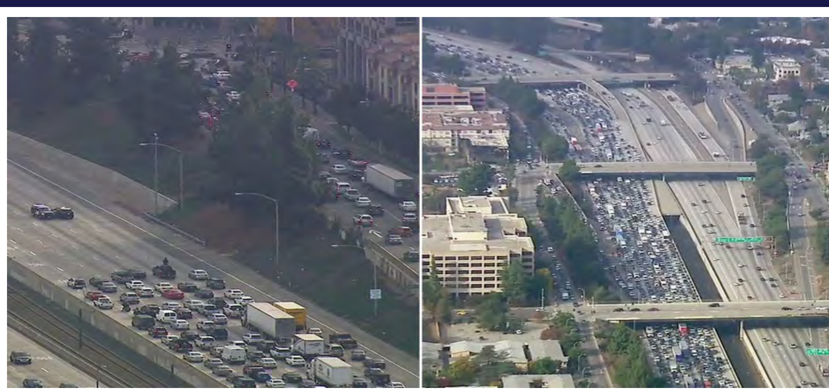
4

Recurrent Operations Challenge - I-210 Pasadena, CA

December 2016: Nissan Altima vs. Big Rig

5

Recurrent Operations Challenge - I-210 Pasadena, CA



Resulting Backup for Many Miles

6

Recurrent Operations Challenge - I-210 Pasadena, CA

Contraflow and Basically a Big, Semi-Organized Mess!

7

Opportunity: Freeway and Arterial Capacity

OPPORTUNITY: Use Existing Capacity More Efficiently

8

Opportunity: Transit (Bus and Rail)

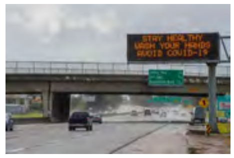
Foothill Transit's 1st Double Decker Electric Bus

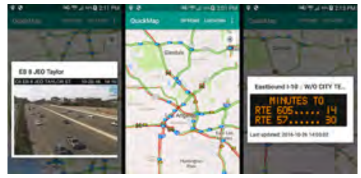
LA Metro Rail Map & Gold Line Train

OPPORTUNITY: Use Existing Transit Capacity More Efficiently

9

Opportunity: Traveler Information

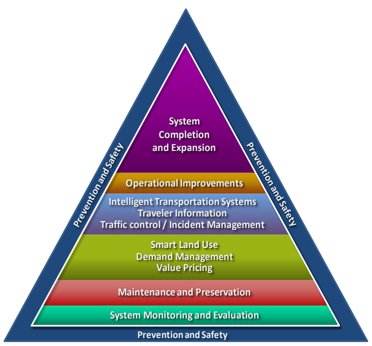




Real-Time Routing, Travel Time and/or Decision to Make a Trip or Not!

10

Mobility and TSMO at Caltrans



Caltrans Mobility Pyramid

Caltrans TSMO Umbrella

11

Caltrans Mobility Pyramid

Transportation Investments have more impact if built upon this foundation

NEED – INFORMATION/DATA SHARING

What Did We Start With?

A desire to manage the considerable existing transportation infrastructure in California to:

- ✓ Improve efficiency of the State's transportation system
- ✓ Improve the effectiveness of our decisions
- ✓ Ensure a multi-modal, multi-jurisdictional approach to TSMO

A desire to change Caltrans culture from our current state to a future state where TSMO is integral to Caltrans culture.

System Management	Current State	Future State
Systems Tools and Functions	Separated	Integrated
Data & Information	Historical	Real-Time
Decision & Business Process	Reactive	Proactive/Predictive
Resources	Static Assignment	Dynamic Assignment
Capital Process	Planning → Design → Operations	Planning & Design with Ops and Maint.

Caltrans and California - Setting

Caltrans

21,000+ employees

12 Districts

51,000 lane mile State Highway System

State of California

1,000 miles Mexico to Oregon
200 miles Pacific Ocean to Nevada Border
163,696 sq. miles

Population:

State of California – 40m
Los Angeles – 19m
San Francisco – 9.6m
San Diego – 3.2m
Sacramento – 2.6m
Fresno – 1.3m

Fundamental Guidance

1. Create a system management culture.
2. Follow a performance-based framework for all transportation system management (TSM) work activities and funding prioritization.
3. Establish a well-maintained and high-performing transportation management systems (TMS/ITS) infrastructure that supports real-time traffic management.
4. Cooperatively develop and implement real-time (active) traffic management to optimize flow, safety and aid regions and the State to meet greenhouse gas reduction (GHG) targets from transportation.
5. **Renew consensus on and adhere to critical statewide standards.**

Statewide Goals

- Enable existing transportation infrastructure and vehicles to work together in a highly coordinated manner
- Deliver improved transportation network performance (safety and mobility)
- Improve accountability
- Evolve Caltrans to "real-time" network operations and management
- Enhance regional, local, and private sector partnerships

Caltrans 16

Planning for Implementation

Connecting Goals and Actions

Caltrans

A Single Statewide System?

Traffic Operations Program Management Direction:

- **Control costs through economies of scale and provide uniform data from across the state for performance measurement**
  - **Replicable *** Common system must be easily replicable for uniform statewide use
  - **Scalable *** Common system must be flexible to meet the unique needs of the districts
  - **Affordable *** Cost sharing with local partners must be easy and defensible
  - **Maintainable *** Single statewide maintainer
  - ❖ **Design must meet current and future stakeholder needs and future needs of local partners across California.**

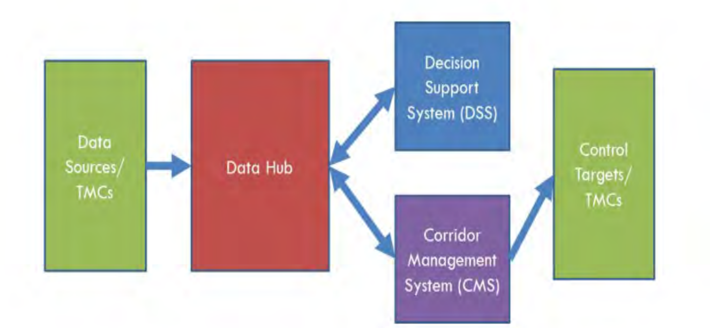
Caltrans 18

Caltrans and UCB PATH Collaboration

- Provide system engineering documents and process templates
- Provide guidance on workforce skills, training and organization
- Enable the consulting ecosystem
- Provide a ***cloud based*** reference implementation architecture with data quality
- Provide a ***cloud based*** replicable and scalable intelligent data hub and DSS (Decision Support System)

19

Major I-210 Pilot System Components



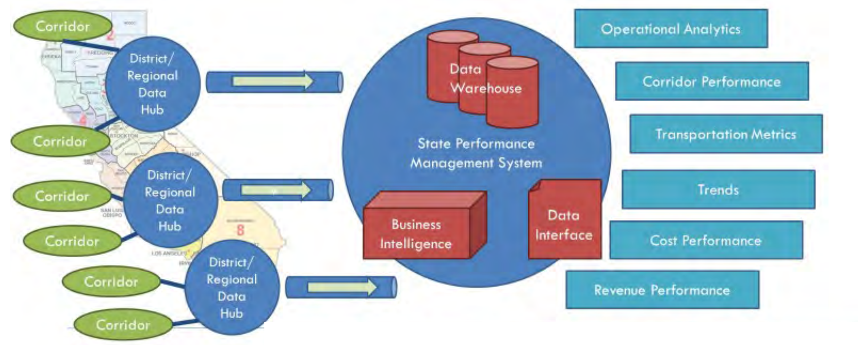
20

Leveraging Data Hub for the Future

1. **Integrated Corridor and Regional Management System** – Expand on the rules-based engine and performance management-based DSS.
2. **Integrated Regional Mobility (IRM) Transportation Portal** – Provide regional partners (public and private service providers) with real-time situational awareness by integrating all relevant transportation modes, systems, and data and sharing it.
3. **Statewide Data Warehouse** – Create a statewide data warehouse with analytics and data mining applications to monitor and analyze statewide operations and system performance.
4. **Performance Measurement and Business Intelligence** - Transform data from the Statewide Data Warehouse (SDW) and/or Regional Data Warehouses (RDW) , into meaningful and useful information that can be used to develop more effective strategic, tactical, and operational strategies and decisions.
5. **Enhance Regional and Statewide Traveler Information Systems** – Share information with 511 regional systems and traveler information service providers via an open data portal application.
6. **Pilot Vehicle to Infrastructure Technologies and Strategies** - Communicating directly with vehicles via connected vehicles applications, providing travelers with relevant information including roadway conditions and operational strategies developed by the regional management systems.

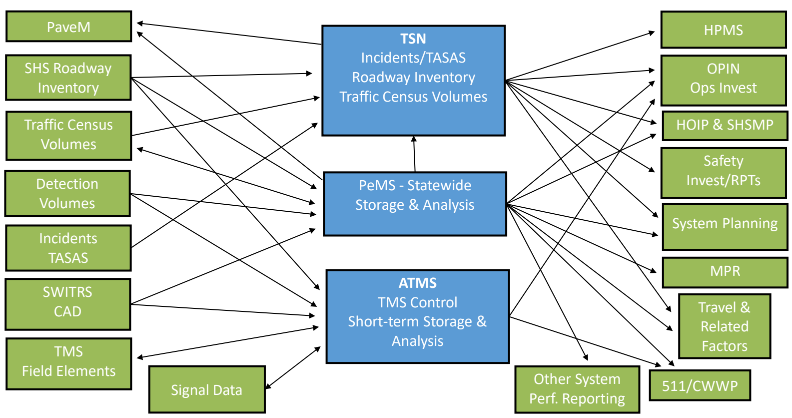
21

Regional & State Consolidation of Data - Potential



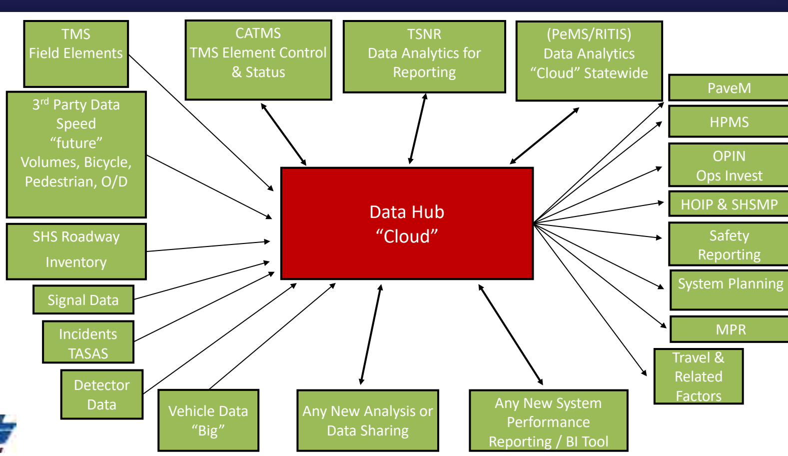
22

Existing State Data Pathways



23

New – More Efficient Data Exchange



24

24

Data Hub, C2C, DSS and I-210 Pilot Schedules

- Data Hub and C2C Interface – Ready and Tested
- Pilot Corridor Management System Interface – Ready and Tested
- Decision Support System (DSS) – Operations Testing Underway
- Rules Engine Cloud Implementation – Operations Testing Underway
- I-210 Pilot Launch – Summer 2021

PATH 25

Open Source and the Pool Fund Concepts

Benefits of "Open Source"

- ✓ No licenses – no ongoing payments forever….
- ✓ No sole-source contractors or vendor-specific solutions
- ✓ No procurement contracts
- ✓ Current software updates available and free
- ✓ Lower up-front costs (when using vendors with existing standard interfaces)
- ✓ Able to leverage community improvements/updates and contribute to community
- ✓ Low cost implementations of improvements made by the community
- ✓ Increased standardized vendor system integration reduces the cost of implementing ICM.
- ✓ Cost effectiveness increases as more agencies adopt the solution
- ✓ Reduces costs for the overall transportation industry when properly managed
- ✓ Permits new ideas and functions to be tested and implemented at a lower cost
- ✓ Open source can attract many people who want to help improve transportation and give them a way to be involved, potentially leading to better prepared workforce in the transportation sector.
- ✓ Standardization of open source interfaces and reference implementations that others can use is as important as having an open source software

26

THANK YOU ☺

Discussion!

27

New York State Department of Transportation (NYSDOT)

1

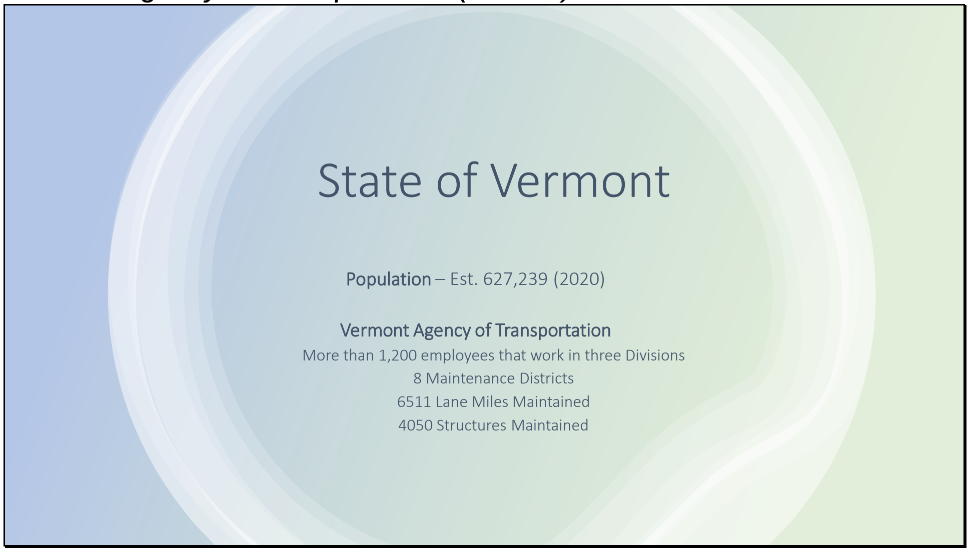
Virtual Peer Exchange

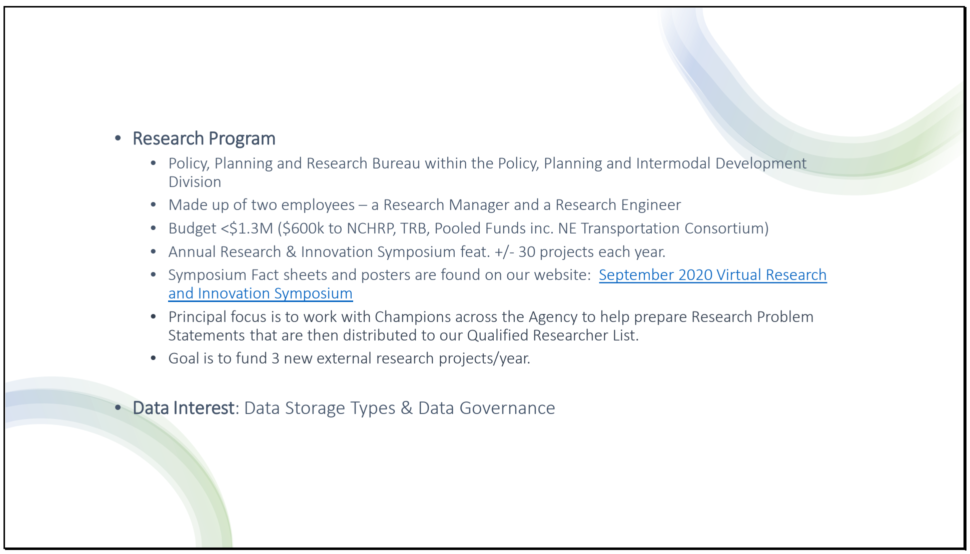
Use of Data Management and Analytics to Improve Decision Making and Research

Hosted by New York State Department of Transportation
February 2-4, 2021

NEW YORK STATE OF OPPORTUNITY. Department of Transportation

Vermont Agency of Transportation (VTrans)





District Department of Transportation (DDOT) – Washington D.C.

WE ARE WASHINGTON DC

NYSDOT Peer Exchange 2021
District Department of Transportation

Stephanie Dock, Research program Administrator

d. GOVERNMENT OF THE DISTRICT OF COLUMBIA DC MURIEL BOWSER, MAYOR

About DC and DDOT

- District of Columbia (Washington, DC; "the District")
  - Population: 705,749
- District Department of Transportation (DDOT)
  - 1,100 employees, no regions/districts
  - 1,400 lane miles, 236 bridges
  - We maintain a multimodal system: 1,400 miles of sidewalks, 150 miles of bike lanes and trails, Capital Bikeshare, dedicated bus lanes, 5 local bus routes, 2.4 miles of streetcar, and 1,600 signalized intersections
- Research Program
  - 2 employees (program admin, librarian) + support from Howard University
  - Housed within State & Regional Planning Division (which handles SPR Part 1 and Metropolitan Planning as well)
  - Report up to the Chief Administrative Officer (along with budget office and HR
  - Principal focus: urban, multimodal research
- Data Interest
  - Integrating data across modes and systems

d. GOVERNMENT OF THE DISTRICT OF COLUMBIA DC MURIEL BOWSER, MAYOR

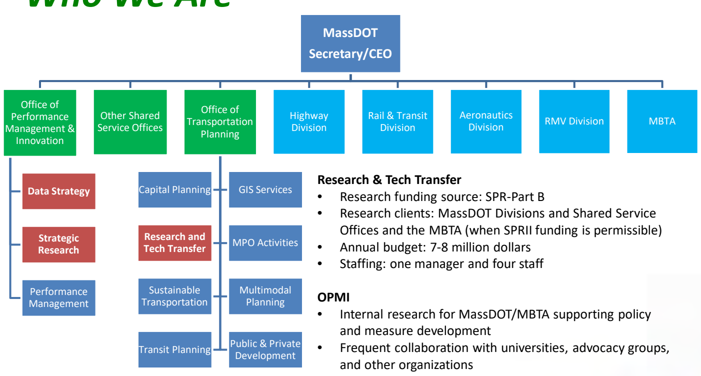
Massachusetts Department of Transportation (MassDOT)

NYDOT Research Peer Exchange

MassDOT SPRII Research Program & Performance Management

Feb 4, 2020

Who We Are



Research & Tech Transfer

- Research funding source: SPR-Part B
- Research clients: MassDOT Divisions and Shared Service Offices and the MBTA (when SPRII funding is permissible)
- Annual budget: 7-8 million dollars
- Staffing: one manager and four staff

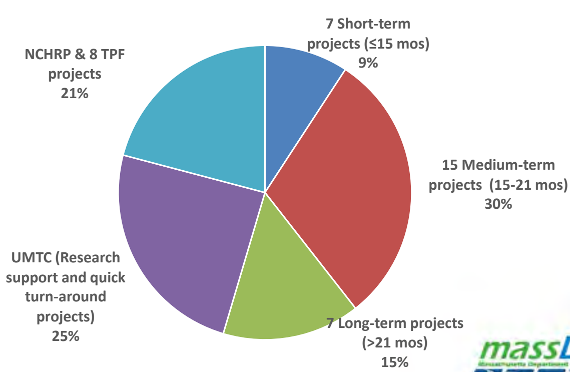
OPMI

- Internal research for MassDOT/MBTA supporting policy and measure development
- Frequent collaboration with universities, advocacy groups, and other organizations
- Data sharing and support: Open Data Portal, Data Blog

2

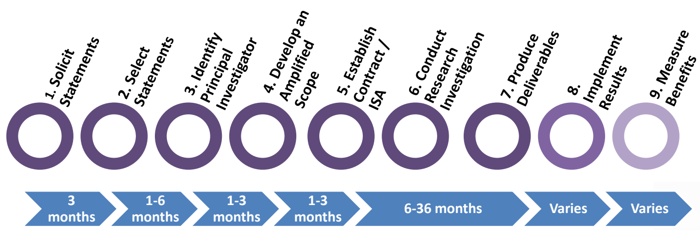
2021 SPR Research Funding by Project Type

~ $5 Million Contracted Research



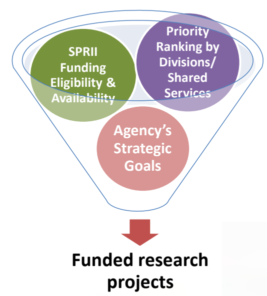
3

MassDOT SPRII Research Process



Research are Driven by Agency Needs

- Research solicitation is internal
- Project prioritization is based on
  - ✓ Agency needs
  - ✓ Implementation likelihood
  - ✓ Potential benefits
- Crucial role of Project Champions
  - ✓ Define technical direction and outputs
  - ✓ Provide access to data, materials, and internal knowledge
  - ✓ Evaluate the quality of deliverables
  - ✓ Facilitate stakeholder engagement
  - ✓ Implement the results



Examples of Research Projects Centering on Data Analytics and Integration

- Using Mobile LiDAR for Automated Asset Inventory and Condition Assessment
  - ✓ Pedestrian Infrastructure (sidewalks and ramps)
  - ✓ Pavement markings
  - ✓ Guardrails
- Multisource Data Fusion for Traffic Incident Detection
- Massachusetts Depth to Bedrock
- Measuring Accessibility to Improve Public Health

MassDOT Research and Tech Transfer:

https://www.mass.gov/research-and-technology-transfer

About OPMI:

We are a shared service in MBTA and MassDOT focusing on Performance Management, Data Strategy, & Strategic Research.
Contact: opmi@mbta.com

Find Our Work Online

Tracker, the annual report card of MassDOT

MBTA Back on Track, a public facing daily dashboard on MBTA metrics

Open Data Portal, a public repository of downloadable data

About OPMI:

Examples of Recent and Ongoing Work

- Transit App Data Infrastructure: are TNCs complementary or competitive with transit? What features of trips make Transit App users select TNC options over transit?
- MBTA Rider Census – FTA-mandated demographic data collection – storage and integration
- Recovering from the COVID-10 pandemic with a reasonable state of the transportation system matching our region's transportation needs
  - What do we need to know to supply the right amount of service, fare products, budget allocation, etc at the appropriate time? How to best organize existing and generated data?
  - How do we encourage the bicycling/walking behavior that has increased during the pandemic?
  - MassDOT/MBTA Employer panel: gather ongoing information about transportation needs and employer choices in the business community during the COVID-19 pandemic.

Contact Us:

Anna Gartsman: agartsman@MBTA.com

Laura Riegel: lriegel@MBTA.com

Lily Oliver: Hongyan.Oliver@dot.state.ma.us

Florida Department of Transportation (FDOT)

State of Florida

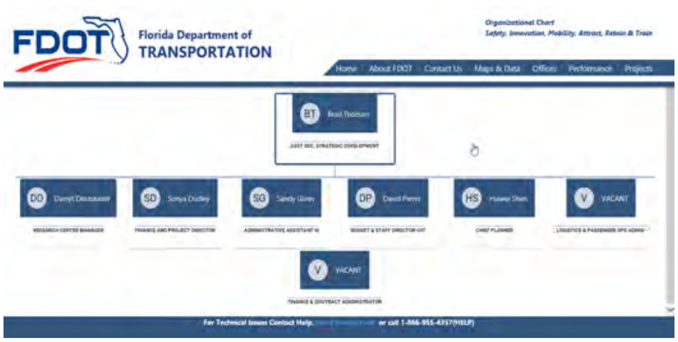
- **Population – 21.2M**
- **Florida Department of Transportation (FDOT)**
  - 5,357 employees
  - 7 Districts and the Turnpike Enterprise
  - 125 active research projects
- **Transportation System**
  - 123,104 centerline miles
  - 12,130 centerline miles on State Highway System
  - 7,007 bridges maintained by FDOT

Organizational Charts

- **DOT Executive Office**

Organizational Charts

- **Strategic Development - Research Center**



Program Focus Areas

- **Secretary's Vital Few**
  - Improve Safety
  - Enhance Mobility
  - Inspire Innovation
  - Foster Talent
- **Research Center**
  - Implementation and Impact
    - Project monitoring and support
    - Financial Achievability Model
  - UTC partnerships
  - Pilots and testbeds
    - I-STREET
- **Data Issues**
  - What's the role of research? Identifying data issues in proposed research, and coordinating between Project Manager/Team, Data Governance Administrator, and other stakeholders.

Florida Department of Transportation

Peer Exchange Takeaways

A record of all peer exchange takeaways is listed below, organized by general thematic category. These takeaways are drawn from all presentations and participant discussions occurring over the three days of the peer exchange as well as subsequent input received from participants.

Data Management Development

- New data provides many new opportunities. Be prepared for new use cases, and potential resistance. Be flexible.
- New, rich transportation datasets require new solutions for data management and storage.
- The nature of new data requires an updated approach.
- Transportation decision making will require a combination of traditional sources, big data, and crowdsourced data.
- The utility of new data opportunities is not always self-evident. There is a need to engage executive management on the importance and opportunities, but also be prepared to adjust.
- Improved data management practices will require champions, investment, and commitment for mainstreaming.
- It is critical to get IT on board with any adaptation to changing data management needs.
- Buy-in from IT is critical to establishing a big data test environment or "playground".
- Test projects must be tailored to the capabilities of the agency, available consultants, and partners.
- The value of improved data analytics must be demonstrated to support more sustained investment.
- Institutional change can be realized through iterative steps that demonstrate the value of big data approaches.
- Establishing a modern data environment requires a paradigm shift; it cannot be done incrementally.
- Collaboration between agencies (DOTs, MPOs), data providers, and the research community is critical.
- Incentivize consistency and common tools and platforms with partners.

Data Storage

- On-premise data solutions often exceed both the technical capacity and expertise of state DOTs and can be extremely costly. Cloud solutions or pay-as-you-go may provide better value.
- Cloud-based and third-party solutions typically lack integrated dashboard interfaces and can pose problems when there is a loss of power or data connectivity.
- Common data templates and centralized data datasets facilitate collaboration.
- Collect the data for a purpose, not merely for the sake of collecting data.
- Data silos tend to produce research silos.
- Data lake is a misunderstood concept; likely to encounter resistance from IT.

Data Governance

- Common data templates and data sets can help support data governance.
- Efforts to establish new data governance policies first require an understanding of data sources and data users.
- Data governance protocols are critical to connecting data sets in schema on read.

- There is a balancing act between the creative potential of data transparency and the security of restricting use to trained data users.

Data Analytics

- Scalability is critical for big data analyses.
- Making data useful to non-specialists within the agency is critical.
- Systems of Engagement allow data to be made useful continually in different use cases.
- Making tabular data geospatial helps to make connections between data and make data useful by “visualizing” data.
- Artificial Intelligence (AI) and Machine Learning (ML) have real-world transportation applications.
- Most data research activities at peer DOTs occur across a patchwork of program areas, IT departments, and outside academic and consultant partners.
- The potential time and cost benefits of open source are too great to ignore in developing agency data capacity.
- Open source approaches can attract unanticipated collaborators, enhancing agency engagement.

Personnel

- Data literacy or data analytics training can expand access to the data, leveraging its utility.
- Effective data projects need constant communication with users and a defined schema of how the data is analyzed. One way to manage: have a data expert in each program area. Need "many pair of eyes."
- Making data accessible helps with consistency and access and prevents agency over-reliance on a single expert staffer.
- Full-time staff would be very helpful for data analysis.
- Aside from hard IT skills, look for people with soft skills who can translate between the need for analysis and what data to use to get there.
- Having a third party manage data can be helpful but need to understand the data to understand what you are getting. Need trained staff to understand the data.

Research

- Successful data research programs need an ecosystem of support for continuity.
- Until internal agency data capacity can be developed, academic researchers can be productive partners in leveraging the utility of existing data resources.
- There is no shortcut to a robust data research program. Intentional, incremental changes will be required. A pilot program is an effective way to start.
- Partnerships are critical to the success of data research programs.
- Collaboration between academic researchers and city and state transportation personnel produces dividends.