Implementation Recommendations for Management Procedures for Data Collected Via CAV

MDOT
Michigan Department of Transportation

CAR
CENTER FOR AUTOMOTIVE RESEARCH

wsp

Sponsoring Organization:
Michigan Department of Transportation (MDOT)
425 Ottawa Street
P.O. Box 30050
Lansing, MI 48909

Performing Organization:
Center for Automotive Research (CAR)
3005 Boardwalk, Ste. 200
Ann Arbor, MI 48108

Implementation recommendations for management procedures for data collected via CAV

MDOT REQ. NO. 2193, Specialty Services: Connected and Automated Vehicle Technology
Task A3. Implementation recommendations for management procedures for data collected via CAV

Authors:
Zahra Bahrani Fard, CAR
Adela Spulber, CAR
Brian Reed, WSP

Managing Editor:
Richard Wallace

Additional Contributors:
Niles Annelin, MDOT
Valerie Sathe Brugeman, CAR
Scott Shogan, WSP

Abstract:
This report is an extension and update of a previous report prepared by CAR, WSP (former Parsons Brinckerhoff), and MDOT in February 2015, entitled *Management Procedures for Data Collected via Intelligent Transportation Systems*. The 2015 report provided an overview and broad discussions of state-of-the-industry and best practices, national ITS research programs and their implications, and existing MDOT plans and data systems. This report extends those findings to review the current MDOT data use cases and to discuss data management and analytics best practices. It also presents a series of recommendations to improve ITS and CAV data management at the state-level.

ACKNOWLEDGMENTS

This document is a product of the Center for Automotive Research and WSP under a *State Planning and Research Grant* administered by the Michigan Department of Transportation.

EXECUTIVE SUMMARY

The Michigan Department of Transportation (MDOT) is a leader amongst state transportation departments in the testing and deployment of intelligent transportation systems (ITS) and connected and automated vehicle (CAV) applications. As part of these efforts, MDOT asked a team led by the Center for Automotive Research (CAR) and assisted by WSP to evaluate the collection, management, and use of CAV data and to make recommendations to develop integrated, dynamic, and adaptive data management systems. This report builds on previous reports from 2013 and 2015 titled *Management Procedures for Data Collected via Intelligent Transportation Systems*. This research focuses on conveying the current state of ITS and CAV data systems and their connections. The goal of this report is to help MDOT increase the overall coherence of the Department's data systems, pursue data systems integration, and eliminate duplicative efforts. The report also aims to identify critical needs, best practices and strategies for better data management practices. The intended audience includes MDOT staff, program managers, and contractors responsible for data systems, as well as users of MDOT data systems from the Department of outside organizations.

The report first reviews six MDOT data systems: Advanced Traffic Management System (ATMS), Regional Integrated Transportation Information System (RITIS), Transportation Asset Management System (TAMS), Data Use Analysis and Processing (DUAP), Truck Parking Information Management System (TPIMS), and Road Weather Information System (RWIS). While MDOT had been proactive in ITS and CAV applications development and adoption, several programs and their underlying ITS and CAV data systems have been launched successfully but independently of each other. Current standards and best practices encourage broad interconnectivity and interoperability of these data systems, especially through use of national ITS data standards and enterprise data warehouse techniques. Therefore, the report recommends strategies to develop integrated, dynamic, and adaptive data management systems based on data management and analysis best practices, and MDOT representatives' insights about current data management practices.

To leverage existing and future opportunities effectively, MDOT should pursue efforts in developing an enterprise data management strategic plan that identifies and outlines best practices and priority systems and data aggregation elements across ITS and CAV data systems and programs. As part this this enterprise data management plan, MDOT should reach out to key stakeholders within the Agency (Planning, Operations, Engineering and Modelling) and identify common data elements and revise schema, data definitions and data values to be

consistent within each desperate system today and create a data transformation method by system. In addition, MDOT should work to develop a statewide strategic plan for data aggregation and data sharing between Agencies, that utilize or consume data between systems and decision support systems. In multiple critical use cases, the same data elements are defined differently within systems and across critical systems for MDOT. MDOT should develop and implement the plan in conjunction with the Michigan Department of Technology, Management & Budget (DTMB), and other select ITS and CAV data users within the state. While various institutional, political, and economic barriers might hinder the implementation of such a broad statewide effort, its potential benefits make it worth pursuing. Strategic database aggregation could multiply the value and utility of current and future ITS programs and datasets and reduce costs. In conjunction, with agency specific and statewide database aggregation and data de-duplication strategies MDOT could improve congestion, transportation network management, improve mobility, and obtain more relevant information in real time to effectively manage pavement throughout the state.

This would benefit multiple agencies, organizations, residents, and visitors in Michigan. By continuing to commit to a leadership role in national efforts to develop and deploy ITS technology and CAV applications, MDOT will have a significant role in shaping the future of transportation infrastructure management. Through its leadership and implementation expertise, MDOT will be well positioned to reap the benefits of becoming the recognized home of this technology.

TABLE OF CONTENTS

LIST OF FIGURES

1 INTRODUCTION

According to the National Cooperative Highway Research Program (NCHRP), data management is "*the development, execution, and oversight of architectures, policies, practices, and procedures to manage the information lifecycle needs of an enterprise ….as it pertains to data collection, storage, security, data inventory, analysis, quality control, reporting, and visualization*" (Cambridge Systematics Inc.; Boston StrategiesInternatiomal Inc.; Gordon Proctor & Associates; Markow, Michael 2010).

Although there is a growing interest in data management techniques, many state DOTs are facing challenges to improve their practices and achieve more success with less funding. This is due to a variety of factors including organizational procedures, growing complexities of data sources and structures, new data elements, and rapidly evolving data collection methods. A recent study suggests that key implementation steps should include organizational vision, communication, assignment of responsibilities and demonstration of value (Hall 2017). In addition, many organizations from government to Fortune 50 private companies recognize the challenges, hard and soft costs with data storage and inefficiencies of data in large volumes. With many traditional Transportation Systems, there remains proprietary data structures, duplicative data elements, conflicting data definitions for multiple common data elements/field values that further complicate information management systems, storage retention, and storage rates/costs. This fundamental business challenge is a driver to creating a data use and data aggregation plan to help MDOT achieve operational and long-term management success of the transportation systems (IT, devices, and roadway).

During recent years, MDOT has been actively engaged in developing advanced Connected and Automated Vehicle (CAV) and Intelligent Transportation Systems (ITS) data systems, as well as acquiring the systems developed by U.S Department of Transportation (U.S. DOT) and Federal Highway Administration (FHWA). These activities have exposed MDOT to an unprecedented amount of data, coming from various sources. The growing volume and complexity of data sources requires MDOT to update its data management practices regularly to capture the most recent advancements in data collection methods, vehicle technologies, and telecommunication approaches.

This research aims to help MDOT in mapping the current state of cross-units CAV/ITS efforts to provide a better overview of MDOT data systems and their connections. In essence, this report contributes to improving the overall coherence of MDOT data systems and to pursuing the long-term efforts to integrate data systems and eliminate duplicative efforts. The report also aims to understand the needs, best practices and strategies for better data management. The intended audience includes MDOT staff, program managers, and contractors responsible for data systems, as well as users of MDOT data systems from the Department of outside organizations.

This report serves to review and evaluate current MDOT ITS/CAV data management efforts and recommend strategies to develop integrated, dynamic, and adaptive data management systems. To achieve this goal, this report describes current CAV data use cases at MDOT, discusses data management and analysis best practices, and summarizes MDOT representatives' insights about current data management practices at MDOT. Finally, it presents a series of recommendations related to CAV data management.

2 OVERVIEW OF MDOT CAV AND ITS DATA SYSTEMS

In recent years, MDOT has acquired new data systems to manage the growing size and complexities of data sources and structures with new data elements, new data sources and continuing evolution of collection methods. The goal of these data systems is to support the decision-making process at reduced cost and with increased operational efficiency. These data systems collect inputs from multiple sources and, depending on the defined role, process them for certain applications. These systems are being utilized either as a data collector, data processor or data provider for other systems. Advanced data systems will provide the users within the Department with the most recent data and will reduce the chance of duplications. While MDOT has a substantial inventory of disparate systems to support the Agency, data collection, data analysis, program and asset management, and decision making for the Agency; many of them are contracted/supplied, several are "home grown", and nearly all of them have data schema that are not common/consistent. The systems identified below are identified by many Departments as critical data systems used within MDOT:

- **Advanced Traffic Management Systems (ATMS)**
- **Transportation Asset Management System (TMS)**
- **Truck Parking Information Management System (TPIMS)**
- **Regional Integrated Transportation Information System (RITIS)**
- **Data Use Analysis and Processing (DUAP) Program**
- **Road Weather Information Data System (RWIS)**

Detailed descriptions of main MDOT CAV/ITS data systems are provided in the following sections.

2.1 ADVANCED TRAFFIC MANAGEMENT SYSTEM (ATMS)

The ultimate mission of MDOT's Intelligent Transportation Systems Program is to assist the Department in improving the transportation system safety and operational performance using existing and innovative ITS technologies. To that end, DTMB contracted Delcan (a Parsons company) to design and develop the Advanced Traffic Management System (ATMS) in 2008, and the system management and support is provided separately or in conjunction with DTMB. The primary goal of the system is to provide real-time performance monitoring and decision-making applications for Traffic Operation Centers (TOCs) to assist them in mitigating traffic impacts in a timely manner. Handling and analyzing large amounts of travel time and probe data are only a few examples of ATMS benefits (see Figure 1). The Delcan contract has been in place for ten years and recently was renewed for another five years. The contract includes project management, design, development, implementation, software and hardware upgrade, and ongoing support and maintenance for ATMS.

Currently, the pilot system provides real-time performance monitoring through the travel time application assisted by probe data and through DMS. ATMS also has a built-in application, which automatically detects congestion due to incidents on the freeway and notifies the users by placing an icon on the interactive map using the event's latitude and longitude. Eventually, ATMS is projected to take over the roles of the current Cameleon ICX-360, LCAR and Call Tracker.

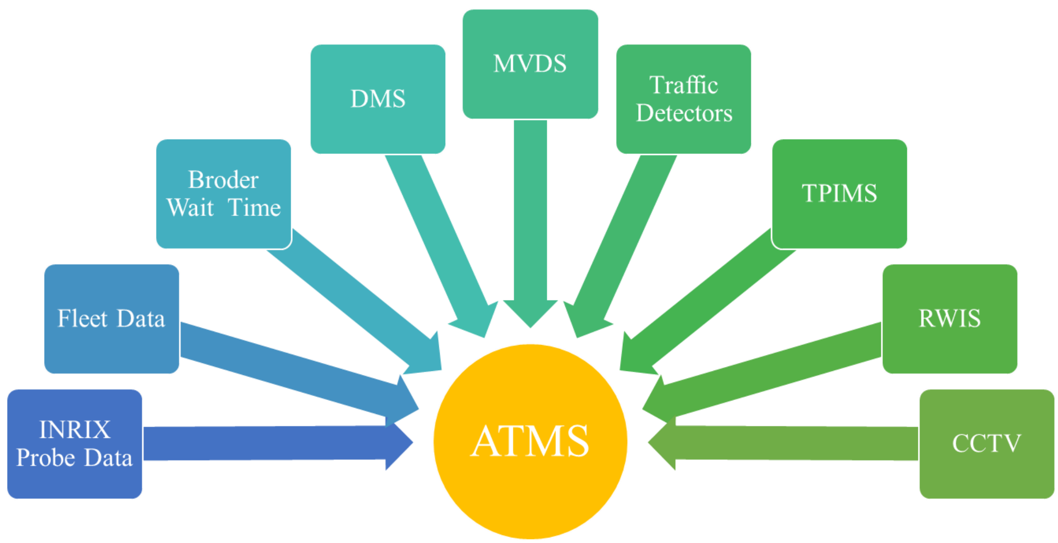


FIGURE 1. ADVANCED TRAFFIC MANAGEMENT SYSTEM MAIN DATA INPUTS

Upon full implementation, ATMS will be an advanced tool helping each regional TOCs in monitoring and managing traffic, recognizing and responding to incidents and delivering real-time information to motorists and coordinating within local and state agencies in an effective, timely, and efficient manner. This will be done through collecting data, sending automatic responses (or updates to traffic facing Dynamic Message Signs), providing video feeds to authorized systems from field-deployed Closed Circuit Television (CCTV) traffic monitoring cameras, updates in real-time to roadway incident/event maps, and allowing the key users to share and distribute information to the external systems/Agency partners. Also, by incorporating CAV data, in addition to ITS data, in the future, ATMS will help traffic management centers in the planning and traffic engineering processes. In its current stage of development, ATMS is being utilized by TOCs for data aggregation and analysis for incident, congestion management, and roadway network monitoring.

Some of the potential ATMS use cases are:

- **Manage traffic.** Potential applications could be:
  - create maintenance ticket,
  - determine toll rates,
  - manage alarms,
  - manage devices CCTV, DMS, ESS, gate, HAR, lane control signal, ramp meter, sub-centers, vehicle detector,
  - monitor AVL equipped vehicles, traffic information, truck roll-over sites, tunnel system,
  - obtain road/weather condition and traffic information.
- **Detect incidents and manage the support response.**
- **Publish traffic condition information**. For example:
  - define a message for conveying information to the monitoring public entities,
  - distribute video images; provide system reports,
  - publish detailed road/weather and traffic condition information, DMS, event information, speed limits, toll information, and travel times.
- **Manage infrastructure administration and operate infrastructure**.

ATMS needs certain improvements to make data extraction from and integration into the systems easier for users. Currently, working with ATMS to extract data is time-consuming and requires substantial efforts from the operator. In addition, MDOT should review the data use/sharing Commercial End User License Agreement and Delcan Contract Terms and Conditions to assess the viability of implementing new data fusion and aggregation for the enterprise. With future enhancements, MDOT internal users such as the Operation Field Services Division could get better quality data from ATMS for traffic incident management. Those enhancements would also enable MDOT units to integrate user-generated data like Waze into ATMS.

There are also a few procedural problems associated with ATMS. Right now, the system does not have any retention procedure scheduled yet, which means part of the stored data could no longer be useful for MDOT. MDOT and DTMB need to collaborate and develop a data retention policy based on MDOT's data use cases within ATMS. In addition, a clear definition on the future role of

ATMS (i.e., data mining or data collection) could assist MDOT users in finding the best use of the system in their work.

2.2 REGIONAL INTEGRATED TRANSPORTATION INFORMATION SYSTEM (RITIS)

The Regional Integrated Transportation Information System (RITIS) is an automated data fusion and dissemination system that provides an enhanced overall view of the transportation network (University of Maryland 2015; see Figure 2 for an overview of RITIS). The main goal of the system is real-time performance monitoring and prioritization assistance for mitigating traffic impacts and analyzing archived traffic data. The MDOT ITS Program Office currently has a three-year (2018-2021) agreement with INRIX and the University of Maryland to maintain, update, and improve the RITIS.

RITIS data comes from real-time incident, event, sensors and detectors, probe, weather, transit and ITS devices. Also, MDOT plans to incorporate data from the Advanced Traffic Management System and user-generated content from applications like Waze into the system. By taking in data from multiple agencies, RITIS is able to provide the most up-to-date and accurate information to its users (URC 2014).

Embedded tools like User Delay Cost Analysis, Bottleneck Analysis and Congestion Scan within RITIS enables users to use the data for research, forecasting, planning and operating the dynamic transportation system. Archived Data Analysis is a powerful toolset within the RITIS which allows users to query, analyze and perform after-action review on archived travel data. The output of the system could be in different format including reports, maps, chart summaries, email/text alerts or even raw travel and incident data. RITIS analytics tools include:

- Probe Data Analytics,
- Event Analytics Tools (from ATMS),
- Detector Data Analytics (from speed/volume sensors),
- Road Weather Data Analytics (RWIS Analytics),
- Virtual Weigh Station Analytics,
- Exploring and Visualizing Crashes,
- Hierarchical Data Explorers (TreeVersity),
- Transit Data Analytics (In Development),
- CHART Reporting.

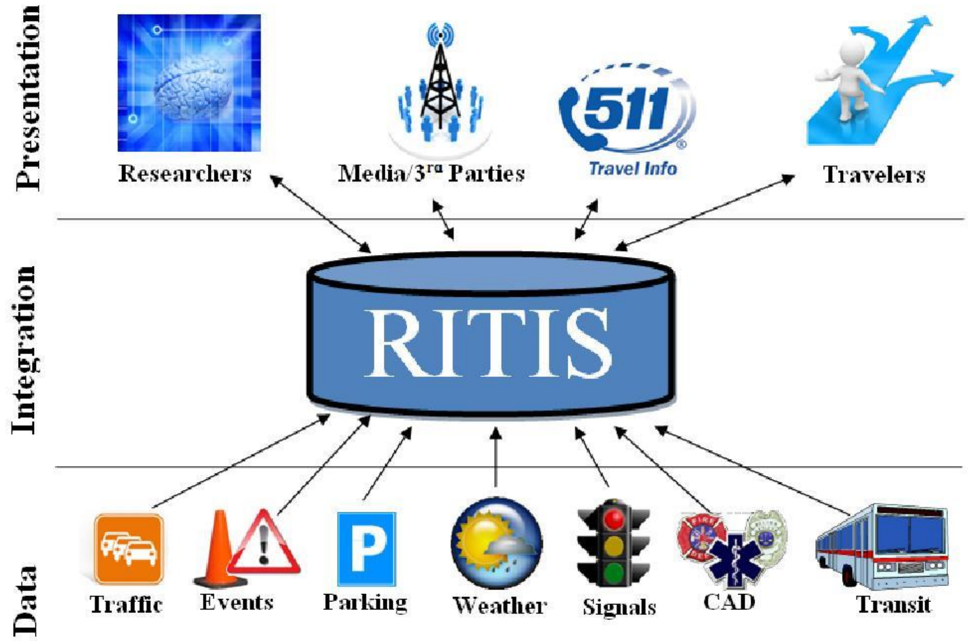


FIGURE 2. REGIONAL INTEGRATED TRANSPORTATION INFORMATION SYSTEM (RITIS) OVERVIEW

Source: (CATT Laboratory. University of Maryland, 2015)

Currently, users with governmental credentials are able to review, evaluate, and create charts and summaries of real-time information and archived travel data within their desired area.

MDOT Operation Field Services Division is the owner and manager of the RITIS system. This Division uses RITIS to respond to data requests from other units and to provide training on using the system for internal users. The Statewide Planning Division is among the main users of the RITIS system and uses travel data (time, speed, probe data, etc.) from the RITIS system to better validate the forecasting models as well as the RITIS Trend Maps as a basis for developing the forecasting models. MDOT TOCs are both data users and providers to RITIS. They use speed data, HERE data (to be replaced with INRIX data), user delay cost reports, and maps from RITIS for traffic operation management purposes and feed incident data into RITIS. Likewise, the Congestion and Mobility Unit uses RITIS to pull out performance measures and reports. Other units like Traffic Incident Management, are looking forward to integrating some of their main databases like Transportation Incident Management (TIM) into the RITIS data suite in the future, as well as to support other units in utilizing these databases in their projects.

2.3 TRANSPORTATION ASSET MANAGEMENT SYSTEM (TAMS)

In July 2015, the State of Michigan, through the Department of Technology Management & Budget (DTMB), and MDOT has Contracted the Data Transfer Solutions (DTS) company for an Enterprise Asset Management System (EAMS) or as used by MDOT, a Transportation Asset Management System (TAMS) Solution. This project includes the implementation of a Commercial Off-The-Shelf (COTS) solution, data conversion and migration, training, and maintenance and support; see Figure 3 for an overview of TAMS.

TAMS is a web-based, user-friendly, affordable and GIS-centric system aimed at supporting maintenance (work tickets/labor tracking), asset and inventory management. At the current stage of development, TAMS functionality is provided to MDOT's users and no other State agencies or MDOT contractors are involved with this initial procurement, although it is envisioned that TAMS functionality may be provided to MDOT partner agencies and MDOT contractors in the future.

MDOT seeks to implement a TAMS that will improve the business processes and support information systems that will enable management of transportation assets across all levels of the Department's organization, locations, and facilities. An effective TAMS will help MDOT maximize value of its physical assets, improve utilization and performance, reduce capital and operation costs, and extend asset life. The implemented TAMS will allow the optimal management of the MDOT's assets, both physical (e.g., buildings, equipment, and infrastructure,) and non-physical (e.g., accounting systems, design, right-of-way, and planning), to maximize their value and lifespan. The main components of TAMS are outlined below:

- **Road Network Management (RNM)**

The RNM will manage the location of highway assets, integrate roadway data from multiple referencing system networks, including coordinates and Linear Referencing Systems (LRS), visualize/analyze road network assets, ensure the TAMS is built in accordance with MDOT's existing LRS, and coordinate/synchronize road network updates with MDOT's Enterprise GIS and ultimately push updates to MiDRIVE (hosted and maintained by DTMB).

- **Asset Inventory Management (AIM)**

The AIM functionality will provide the ability to record, store and manage physical description, historical value, performance and other attributes of multiple asset types through a geo-referenced database with the ability to record physical description, historical value, performance, and more with an initial focus on lanes, signs, culverts, guardrails bridges and pump stations.

- **Highway Maintenance Management (HMM)**

The HMM functionality will allow MDOT to capture and report maintenance activities, quantities, location, costs, performance-based planning/budgeting, resource needs, annual work program, work order scheduling-tracking, performance monitoring, and customer service delivery.

- **Application Integration**

The Application Integration functionality will enable future integration with other existing and planned application systems and databases to be incorporated within the TAMS environment such as TMS, Management Information Database (MIDB), MIBridge, Property Damage Reclamation Process (PDRP), M5 Fleet Focus, FIN MSTAR, enterprise GIS (or SDE geodatabase), Global, and Milogin.

The above mentioned key functions will be running on five main underlying technologies detailed below.

- **VUEWorks**
- The VUEWorks software is a web-based system designed by DTS for enterprise asset and work management. The software is able to manage, analyze, integrate and share asset and work management information to the users within maintenance, planning and forecasting units to support their work. Inorder to be able to process geospatial data, the VUEWorks software is integrated with other components of the TAMS system (i.e., VUEWorks Mobile Solutions, ESRI Roads and Highways extension, the Linear Referencing System Framework as well as Transcend Road Highway Productivity Tools)
- **Mobilevue**

Mobilevue is a mobile asset and maintenance management solution that can be utilized on any platform (Toughbook, tablet, smartphone) in both connected and disconnected modes for mobile field asset data collection.

- **ESRI Road and Highways Extension & LRS**
  Per integration of ESRI Roads and Highways solution in to the Department TAMS, users will be able to visualize the related data in GIS maps or use them within reports. The tool is capable of analyzing the impacts of any

changes to the LRS Network using ESRI Roads and Highways on registered event layers from an external system.

- **Transcend Road Highway Productivity Tools**

The Road Highway Productivity Tools is a toolset consists of Segment Analyzer, Intersection Manager, HPMS Assistant, Report Engine, and LRM Convener for managing and analyzing transportation data.

- **Road Analyzer**

Road analyzer is a straightline diagramming tool that could be integrated with ESRI's R&H solution. The Road Analyzer tool facilitates access to roadway data by displaying the data in a stacked bar display format.

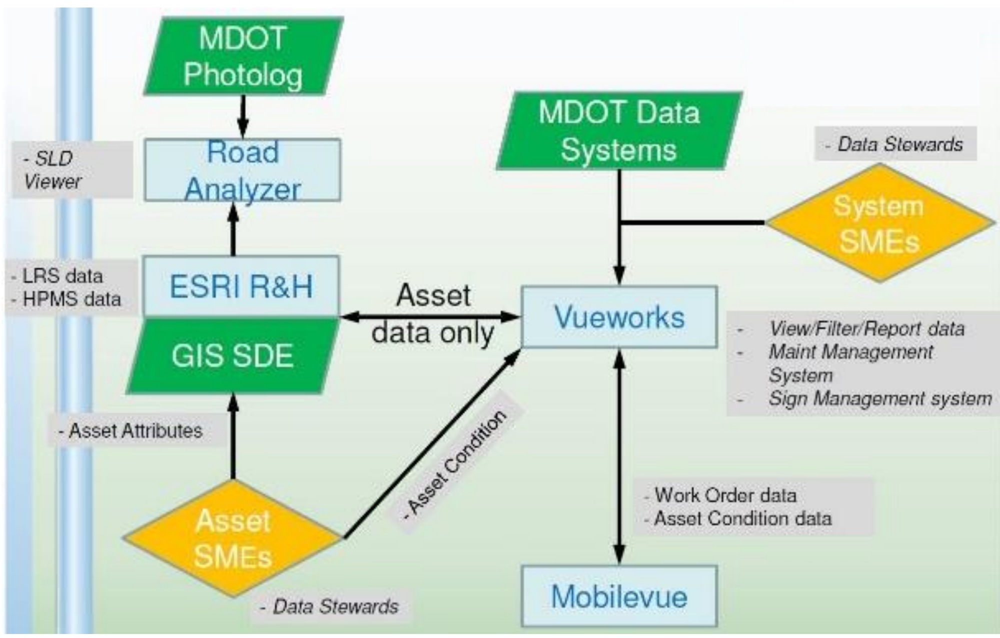


FIGURE 3. TRANSPORTATION ASSET MANAGEMENT SYSTEM SCHEMA

Source: (Droste, 2017)

Once the TAMS system becomes active, it is expected to be used by Central Office, Transportation Service Centers, and MDOT garages for the following applications:

- Asset Management Data integration,
- Add and view internal and external customers' requests,
- Manage service requests,
- Resource management,
- View, add and edit work orders,
- Asset Condition Tracking,

- Capture risks for strategic planning and prioritization of work activities,
- Calculate the value and description for all physical assets,
- Budget forecasting and analysis.

In addition to the above-mentioned use cases, users within the Statewide Planning Division are looking forward to adding inputs from their forecasting models to TAMS to collaborate with other units within the Department.

Like any other data systems, TAMS has its own limitations. For example, due to the lengthy process of developing the system and general setting of the off-the-shelf solutions, the ITS Program Office has had to develop its own ITS Asset Management Database to meet the immediate needs. This is the primary data system that ITS uses to manage linear assets, review programs and jobs, project scoping, and request funds, among other use cases.

2.4 DATA USE ANALYSIS AND PROCESSING (DUAP)

The DUAP program is a research and development effort aimed at increasing efficiency, improving performance management, and facilitating data sharing. By achieving these goals, leaders and decision makers within MDOT will be able to utilize reliable and near-time CAV data to support them in making short and long-term decisions. Currently, the DUAP system can support up to 1000 simultaneous users. The main target users within the Department are the Asset Management and Policy Division, the Design Division, the Construction Field Services Division and the Operations Field Services Division. In the future, DUAP will be the central system in which the ITS Program Office will integrate the ITS devices and mobile collection data into the system, and enable the ITS Program Office to showcase the use case of CAV data for other business areas within the Department.

DUAP I

First phase of the Data Use Analysis and Processing (DUAP) initiated in 2007 to investigate use and benefits of data from U.S. DOT's Vehicle-Infrastructure-Integration (VII) connected vehicle program. Later, due to lack of efficient connected vehicle data, DUAP Project was reconfigured to utilize data sources developed by MDOT directly, such as fixed vehicle detection station data, fleet data, and bridge infrastructure condition monitoring data. Through a five-year contract (2007-2012) with Mixon Hill, the program made some progress in developing a traffic monitoring and a pavement condition monitoring

DUAP II

Following the Phase I completion, the same contractor was awarded a new contract for DUAP 2 project in 2012. The goal of Phase II project was to transition MDOT data management practices to a system more capable of utilizing various data sources and distributing information across the department.

DUAP III

At the end of Phase II, MDOT contracted Mixon Hill for one year with an additional four-year extension (2017-2022) to develop DUAP Phase III. This contract is mostly focused on DUAP System hosting, software maintenance and enhancement, and new application development.

The currently developed DUAP system (see Figure 4) is capable of collecting, processing, storing, and analyzing data from a variety of sources and making it accessible to users across MDOT. The system supports three main CAV data applications within MDOT:

Pavement Application. This application supports users in pavement maintenance (e.g., performance-based maintenance and pavement defect detection), design (e.g., pavement warranties, pavement life cycle analysis) and asset management (e.g., pavement defect detection and surface condition).

Weather Application. The Weather Application support users in tasks related to maintenance (e.g., maintenance decision support and surface condition monitoring), operations (e.g., traveler information and real-time situational awareness), and traffic and safety (e.g., Wx-related crash monitoring).

Traffic Application. Through the traffic application, the DUAP system supports maintenance decisions like incident management, construction and real-time work zone monitoring, operation (e.g., congestion/queue detection and incident management), and the planning-related tasks.

To perform the above-mentioned analyses, the DUAP system collects data from various sources which include but are not limited to the following:

Traditional Systems

Maintenance Decision Support System (MDSS)

National Weather Service (NWS) Warnings/Radar

Lane Closures and Restrictions (LCAR)

Fixed ITS Devices

Advanced Traffic Management System

Closed-Circuit Television Cameras (CCTV)

Dynamic Message Signs (DMS)

Microwave Vehicle Detection System (MVDS)

Environmental Sensor Stations (ESS)

NWS Stations (ASOS/AWOS)

Mobile Source Data (Connected Vehicle)

Vehicle-Based Information and Data Acquisition System (VIDAS)

MDOT Automatic Vehicle Locator (AVL)

Connected Vehicle Data (Basic Safety Message (BSM))

Other Future Data

Construction

Maintenance

Pavement

Signal Controller (Phasing)

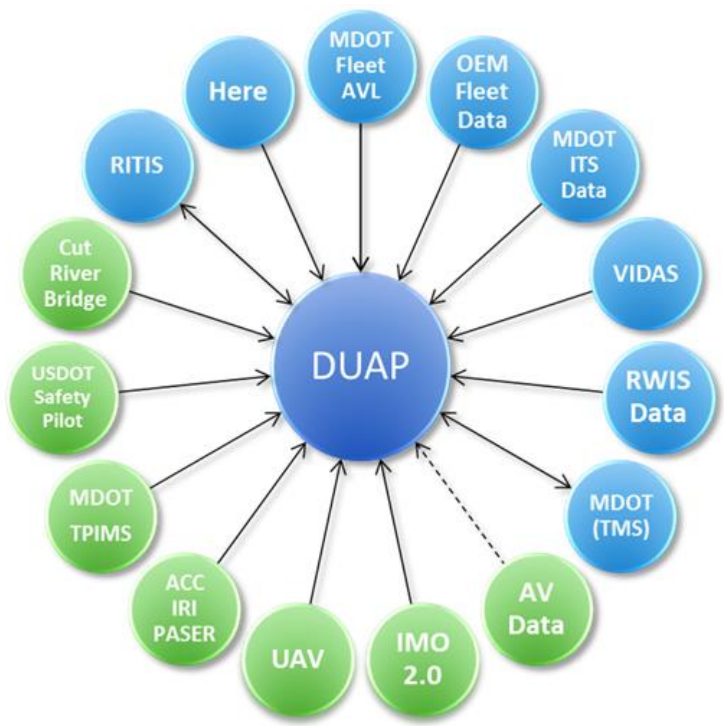


FIGURE 4. DUAP SYSTEM SCHEMA

Source: (Castle, 2015)

Due to the size and complexity of the system, the physical DUAP system is comprised of several subsystems that consist of many components and layers. Some of these subsystems are introduced below:

- **Data collection system**

As the primary component, the data collection system is detecting data that has been collected by an external intelligent transportation systems and submitted to the DUAP system. The system will process, analyze, check the quality of data, and sort them into appropriate databases.

- **Data management system (DMS)**

The DMS component is responsible for a round of in-depth quality checking, verification and validation of the data before sorting them into accessible locations for users.

- **Data distribution system**

Finally, the data distribution system works to respond to the dissemination requests from applications or systems. Any necessary business analytics are performed before the user consumes data through the applications being developed as part of the DUAP program.

Within the DUAP system, users can see the data on a map, or use the reporting or analyzing featured. Within the map feature, users can choose among the available data elements like connected vehicle data (VIDAS, MDOT AVL, and BSM) which contain information including but not limited to vehicle path, road condition, road defects, location and type of the defect, most recent photos of road condition, taken by the VIDAS vehicle camera every ten seconds.

Given DUAP's potential, MDOT is looking for more applications of the data available within the system. To meet this goal MDOT needs to start the communication among target users within the Department to identify the needs and educate internal users to utilize the DUAP system in their projects.

2.5 TRUCK PARKING INFORMATION MANAGEMENT SYSTEM (TPIMS)

The Truck Parking Information and Management System (TPIMS) project (see Figure 5) is a federally funded project under the Federal Highway Administration (FHWA) Truck Parking Facilities Discretionary Grants Program. Within MDOT, TPIMS is being used as a system that assesses truck parking availability at parking and private parking facilities along the I-94 corridor in Southwest Michigan. The system collect the parking availability data from multiple sources and deliver real-time availability information to truck drivers through different media including:

- Dynamic Truck Parking Signs (DTPS): The availability information displayed on the DTPS is sent in near real-time utilizing MDOT's ATMS application.
- Smartphone applications: Third-party smartphone applications disseminate truck parking information to subscribers.
- Connected vehicle infrastructure: Parking availability information coming from ATMS will be broadcasted to DSRC units to inform drivers ahead of rest areas.
- MDOT MiDrive website: ATMS publishes parking availability information on the MiDrive website and mobile application.
- Traveler information websites: Third party traveler information websites are being used as dissemination vessels for parking availability information to subscribers.

The system includes traffic sensors that collect truck entry and exit data, CCTV cameras which are installed at each rest area to verify sensor data and truck parking availability data at private facilities that are provided by a parking information service provider.

The system architecture includes components related to data collection, verification, processing, storage, communication, display and information dissemination as is shown on Figure 5. After sensors collect public facilities parking availability information, data is stored on local servers at the rest areas. Then the local servers process the sensor data and validate them based on the information coming from CCTVs. Verified information

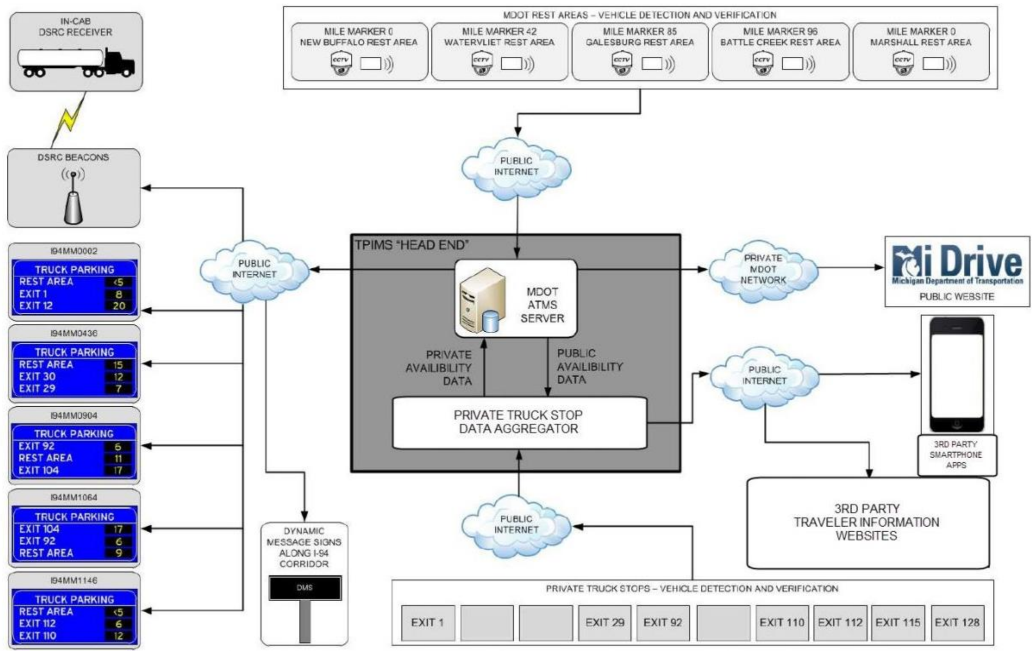


FIGURE 5. TRUCK PARKING INFORMATION SYSTEM ARCHITECTURE

Source: (MDOT, 2016)

2.6 ROAD WEATHER INFORMATION SYSTEM (RWIS)

RWIS is an ITS subsystem that consists of several Environmental Sensor Sites (ESSs or RWIS ESSs) connected to a central processing unit (CPU) (see Figure 6). These elements form an information system which is focused on collecting road-related weather information data to support roadway and maintenance decisions, especially during the winter season. The ESSs collect a variety of data including pavement temperature, atmospheric temperature, frost depth, snow depth, traffic, cameras, and/or visibility (Brinckerhoff & Iteris, 2013). Atmospheric, pavement surface, and sub-surface temperature, microwave vehicle detection sensors (MVDS), and CCTV camera are some of the sensing units used in ESS to collect data.

The overall goal of RWIS is to develop a system that improves the traveler experience through better communication and timely response to a weather event. By utilizing a group of optimally located ESSs, the RWIS was intended to help better identify inclement weather and adverse road conditions at critical locations in order to efficiently identify and prioritize maintenance needs and provide improved information to motorists. (same)

The mission of RWIS is to aid MDOT in state road and bridge infrastructure repair and maintenance decisions by collecting road weather and traffic data. The vision of the MDOT RWIS program is to have a robust system that provides stakeholders useful information about road weather conditions around the entire state of Michigan using an array of existing and next-generation technologies.

RWIS data is important in maintaining and improving travelers' safety and mobility especially during the unpredictable winter months in Michigan. Today, MDOT users, especially the Congestion and Mobility Unit and TOCs) have access to RWIS data through ATMS. However, the ability to disseminate road weather information automatically or forecasted data is not currently available within the existing system configuration and requires integration with ATMS.

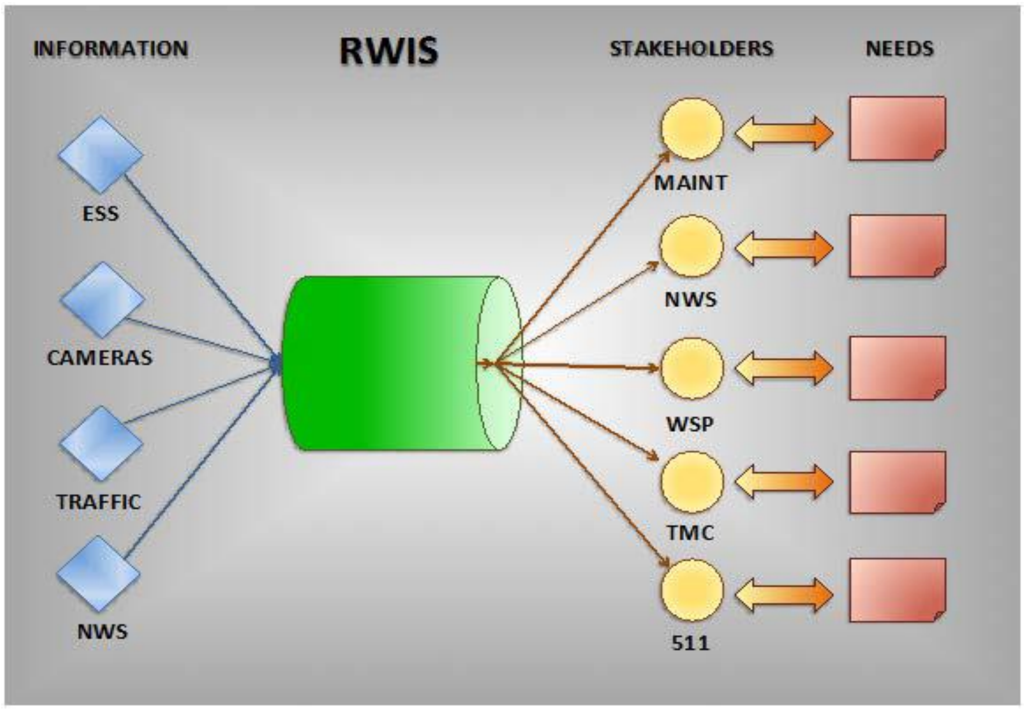


FIGURE 6. DIAGRAM OF THE COMPONENTS THAT COMPRISE RWIS

Source: (Brinckerhoff & Iteris, 2013)

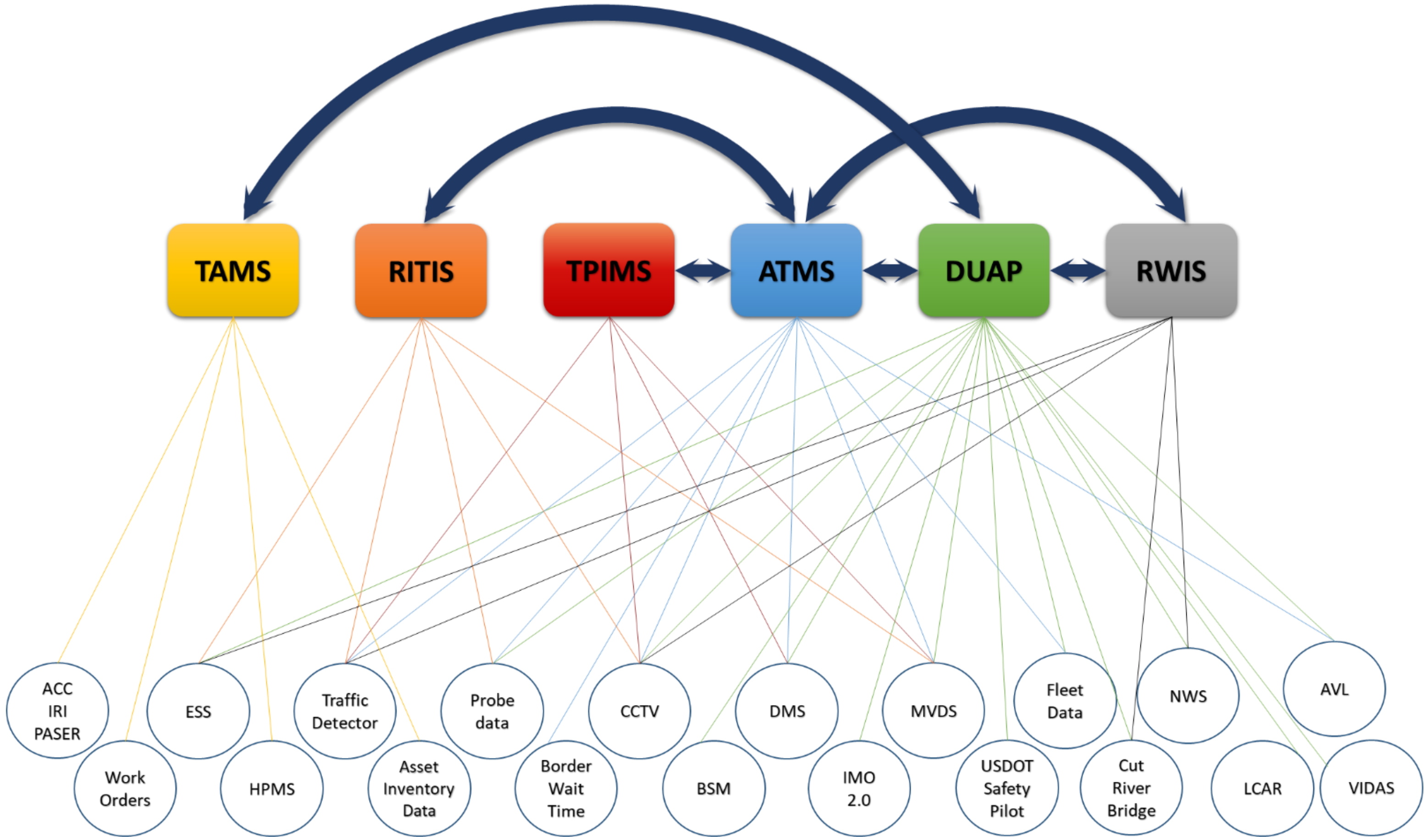
2.7 MDOT CAV DATA SCHEMA

Through years of implementing advanced CAV and ITS projects in Michigan, MDOT has collected enormous amount of data that could be used in asset management, maintenance, planning, construction and forecasting decisions. Understanding the dynamics of CAV data systems structures will help MDOT developing data management practices to better manage high demand data systems and find the best use cases for existing data.

Figure 7 shows six main MDOT CAV/ITS-related data systems and their main data inputs. Each system collects data from various sources and the list of data inputs for each system is very detailed. For that reason, the graph only shows some of the main CAV/ITS datasets to ease the understanding of MDOT CAV/ITS data schema.

Currently none of these systems are connected to each other but, with future improvements, the interaction between CAV/ITS data systems could enhance the collaboration between internal units and, reduce duplicative efforts, and facilitate data need identification.

FIGURE 7. MDOT CAV DATA SCHEMA



3 DATA MANAGEMENT AND ANALYTICS BEST PRACTICES

With growing application of big data in transportation systems management, there is a huge need for improving data management practices within the state DOTs. Based on the interviews with MDOT business unit key stakeholders, the CAR team figured out that MDOT business units are interested to learn more about advanced CAV data management practices. Below is an overview of next steps for MDOT to implement a master data management framework as data sources, storage and complexity continues to emerge within the Transportation.

The continued advancements of CAV datasets are beginning to fundamentally shift data analytics and forcing traditional IT and ITS systems to leverage and interact as part of Big Data "stores" to more effectively support the transportation network data analysis, performance measures, and new opportunities to improve Operations & Maintenance uses within the transportation market.

Business units within MDOT address a wide range of functions and data sources, meaning that they have important data located and stored in a multitude of places and file types (e.g., spreadsheets; Microsoft Access; database platform X, A, B, and C; and data coming from all kinds of field devices). In such cases, employees probably have important data in several different places – internal and external and not all the structures align, and in many cases have duplicates of data sources. In addition, it is more likely that with as many data copies within the Department there are multiple versions and likely that each have data modifications that do not exist in another version. Meaning that a data element defined in System 1 does not mean the same thing as the same data element defined in System 2, and the worst scenario is that data abbreviations/actual record data are not common across systems as well. What all this means is that a key recommendation is to create a data governance council, key data stewards from each MDOT divisions to meet, define and agree on how to define data elements, abbreviations, uses, data definition (per system/data source) and data schema for each key system. Over the next 5 to 10 years, as data volumes and data sources continue to increase a rapid rate it is critical that MDOT begin processes to identify data schema and manage data definitions by system as a twice a year update "iterative cycle"; this process would be the beginning of change management for data definitions/systems revisions over time.

As part of the data governance and management plan key expectations should be to define:

1. What, and where, all the data that's relevant to each business unit comes from (priority systems or datasets first per group)?

2. How many units need the same data in their projects?

3. Is the data accurate, current, clean and complete?

4. Can the data be aggregated readily as-is or will it need to be transformed before it gets warehoused or centralized (e.g., a data pull to core enterprise Big Data warehouse)?

5. How frequently is the data being updated or records being added; and how should users alert or find and flag data quality issues within the data (gaps, misalignment, or discrepancies by centralized and common date/time stamps across systems)?

6. How frequently does the data definition for each system need to be refreshed and maintained/revised with updates to an enterprise transformation index?

The biggest question that needs to be addressed here is: Is the data ready to support business analytics? An often-ignored truth is that before users can do exciting or valuable things with analytics, they need to be able to "do" data first; that is to ensure data management within each priority system.

Not every IT system, proprietary software or database system can be utilized to give an organization the best implementation framework for big data (e.g., not every IT solution is a best fit for big data management). Traditional data storage platforms like Microsoft SQL or Oracle are not well suited to centralize or warehouse data from multiple systems; especially without customized tuning and a specific work effort to design the data table structures of disparate data definitions of multiple single data sources/systems or even huge annualize licensing costs for an enterprise like Michigan Department of Transportation (MDOT). In the past five years alone, PostgreSQL and Hadoop open-source enterprise database management platforms have become extremely competitive alternatives and at no-cost or very low-cost entry points for organizations looking to better manage, analyze, and warehouse big data.

Although big data has received lots of attention recently, however there is a bit of misconception in the market place. Big data is a term that has been around for a while, but many organization's and people are confused about what it is. The concept of big data is continuing to evolve; as it remains a driving force behind many ongoing waves of digital transformation, including connected and automated vehicles, artificial intelligence, data science and analytics and of course the Internet of Things (IoT).

From Forrester Research (Forrester, 2016) on big data storage, there are several key growth areas that organizations need to understand to effectively store, manage, process, analyze and get value from the data they are collecting. These key growth areas are outlined below:

- The typical organization will grow their data storage by 50 percent just in the next twelve months.
- Database systems will grow by 97 percent.
- Server backups for disaster recovery and continuity will expand by 89 percent.

Big Data results in three basic business or organizational challenges: storing, processing and managing it efficiently. Scale-out architectures have been developed to store large amounts of data, and purpose-built appliances have improved the processing capability. The next frontier is learning how to manage Big Data throughout its entire lifecycle, and effectively aggregating data from multiple systems.

What many organizations don't recognize and that people don't know is that an overwhelming majority of Big Data is either duplicated data or synthesized data. In 2018, people generate data whenever they go online, carry their GPS-equipped smartphones, communicate with their friends through social media or chat applications, pass detectors, post a message on a Dynamic Message Sign (DMS), analyze Bridge Health/Ratings, and collect Light Detection and Ranging (LiDAR) for a transportation corridor or design project. It could be said we leave digital footprints with everything we do that involves a digital transaction, which is almost everything (everyday activities, devices, and "working" online).

On top of this, the amount of machine-generated data is rapidly growing too, from Internet of Things (IoT) devices, connected and automated vehicles, and drones/robots. Data is generated and shared when our "smart" devices communicate with each other or with business servers. Soon, CAVs will take to the streets, that utilize real-time, four-dimensional maps of their surroundings (and being sent Intersection Map messages via Dedicated Short Range

Communications (DSRC). As more vehicles and transportation infrastructure become equipped to send back Basic Safety Messages (BSMs) and probe vehicle data (PVD)_with or without detailed Controller Area Network (CAN) bus data interfacing to sensors and systems within vehicles_ to transportation agencies across Michigan, the amount of data and analytic possibilities exponentially increase.

These devices, data and systems become key inputs to manage, store, and analyze Big Data for the transportation network and more effectively manage incidents, improve mobility, and change how transportation is used.

Master data management (MDM) (Loshin, 2006) is a comprehensive method of enabling an enterprise to link all it's critical data to one file, called a master file, that provides a common point of reference. When properly done, master data management streamlines data sharing among personnel and departments. In addition, master data management can facilitate computing in multiple system architectures, platforms, data sets into a common Big Data Warehouse. MDM is used to aid in solving business issues and improving data quality through the effective and seamless integration of information with business processes. Unfortunately, a frequent mistake is to treat MDM as a technical issue only. While focusing on the technical or technological implementation within an organization enables a quick start its MDM initiative, it leaves most critical problems unattended, and dilutes benefits of a MDM program.

By using a technology-driven approach to an MDM program, it decreases business confidence, and makes it difficult to sustain the solution, thus causing an enterprise adoption of MDM program to fail early. However, a technical solution well integrated with business processes, along with a strong governance program, is a more effective approach beginning a MDM program. A business-driven approach can ensure the success of MDM program and enable a path for further expansion. Other issues for example with implementing a successful MDM program include data quality, consistent classifications/data definitions, and identification of data, and data-reconciliation issues.

Master data management of disparate data systems requires data transformation as the data extracted from each disparate source data system is transformed and loaded into the master data management system or Big Data Warehouse. To synchronize the disparate source master data, the managed master data extracted from the master data management hub is again transformed and loaded into the disparate source data system as the master data is updated. As with other Extract, Transform, Load (ETL-based) data movement, these processes are expensive and inefficient to develop and to maintain which greatly reduces the return on investment for the master data management product.

One of the most common reasons some large corporations experience massive issues with master data management is growth through new data acquisition or continuing to implement single vended or proprietary systems for managing devices/command and control functions without a method or means of implementing common data definitions from an MDM approach or program. Any organizations which merge, will typically create an entity with duplicate master data (since each likely had at least one master database of its own prior to the merger). Ideally, database administrators resolve this by deduplication of the master data during the data or system merge, frequently as part of the ETL-based data trigger. In practice, however, reconciling several master data systems can present difficulties because of the dependencies that existing applications have on the master database or Big Data Warehouse. Frequently, the two systems do not fully merge, but remain separate, with a special reconciliation process defined that ensures consistency between original source data repositories. As additional "disparate systems" are brought online is typically where size/scale issues begin to occur; the core issues become managing multiplies, more and more master databases appear, and data-reconciliation processes become complex and fragile. This results in unmanageable and unreliable data merge into the MDM program or Big Data Warehouse. Results of this trend, are typically found within organizations with 10, 15, or even as many as 100 separate, poorly integrated master databases, typically cause serious operational problems in the areas of operational efficiency, adequate performance measures, decision support, cost-effectiveness, data quality, and regulatory compliance.

4 CONCLUSION AND RECOMMENDATIONS

For this task, CAR/WSP team conducted a series of interviews across the Department to capture main stakeholders' insight about the current data management practices within the department.

MDOT and DTMB representatives interviewed for this report provided valuable information concerning use cases and applications for which CAV data would be beneficial and types of CAV data that are needed for those applications. Interviewees also identified gaps, issues, and challenges related to CAV data collection, analysis, storage, and management. The most important findings from the interviews are detailed in this chapter and they inform the implementation recommendations identified.

4.1 INTERVIEW FEEDBACK ON DATA GAPS AND DATA MANAGEMENT CHALLENGES

PRIORITY CAV DATA USE CASES AND NEEDED DATA ATTRIBUTES

During the interviews, MDOT and DTMB personnel mentioned how they would leverage CAV data. They also mentioned the applications or use cases for which integrating CAV data would make the most positive impact. The use cases are listed below by broad functional categories.

MDOT personnel interviewed for this project is interested in CAV data and believes that having access to more CAV data would be beneficial for the Agency. However, several MDOT representatives recognized they did not have sufficient knowledge about CAV data capabilities to fully understand the benefits of this kind of data and to define more opportunities to leverage this data. The interviewees mentioned several types of CAV data attributes that would be a priority for them. Overall, MDOT is interested in having access to rich probe data, both in terms of scale and timeframe. Access to data in real-time is also a priority.

Asset Management

- Pavement condition assessment (road profile, surface condition, defect type, location and severity) could be improved with CAV data. Real-time vehicle sensor data from a larger number of vehicles, beyond the MDOT fleet instrumented through the VIDAS project, would allow MDOT to have a more frequently updated and accurate assessment of pavement conditions, not just once every two years. This data should be shared across MDOT business units in a more real-time fashion to take more proactive actions, predictive maintenance, and better prioritize maintenance actions.

- MDOT personnel views this as the topmost priority because pavement is the most expensive MDOT asset.
- Data attributes that are a priority for MDOT:
  - Data from vehicle camera, accelerometer, suspension, LiDAR.

Incident and Work Zone Management

- Access to CAV data would potentially provide better and more frequent data sources to support incident detection and management. An Incident Response CV Application could help overcome traditional issues for work zones and incident management. For example, the stop traffic advisory and queue length or stop locations from CAV data could be linked. Rear-end crash alerts and other vehicle data could be used to adjust dynamic messaging sign in advance of work zone to reduce back of queue crashes and further adjust speeds within active work zones.
- CAV data could also be used to better track, analyze, and send update messages related to emergency vehicles and activities throughout the lifecycle of a roadway incident.
- Work zone design could be improved at the project level thanks to real-time analysis of CAV data.
- Data attributes that are a priority for MDOT:
  - Vehicle probe data and deep DSRC On-board Unit (OBU)/CAN data elements.
  - Vehicle speed, position, brake activation.
  - Vehicle incident location and timestamp (for secondary crash information).

Traffic Management

- Integrated corridor management and congestion management could benefit from integration with more CAV data. MDOT personnel mentioned queue detection, monitoring of queue progression, dwell time, identification of bottleneck locations, and speed reduction messages.

- CAV data could be used for signal optimization (potentially real-time), and signal priority (depending on vehicular arrival at a red light). This could include fusing signal performance measures with CAV to better assess adaptive, responsive signal control or even how signalized corridors are performing (arrivals on green, corridor progression, incident detection in lanes (or between intersection detectors), etc.)
- Vehicle data could also be used to analyze driver behavior and especially behavior modification regarding dynamic message signs for example.
- Lane merge assistance and platooning could be implemented and improved with leveraging future CAV datasets. Analysis of CAV data could improve zipper merge messaging from MDOT to vehicles on the roads, or perhaps implement more automated and efficient ramp metering applications (that are devoid of custom systems and field deployed hardware/communications devices).
- MDOT could also use CAV data for trip pattern analysis.
- This data could also be leveraged to calculate vehicle counts or traffic volumes in real-time.
- Data attributes that are a priority for MDOT:
    - Vehicle speed, position, path prediction, brake activation (for example, in conjunction with 'Detour' posting to compare actual traffic redistribution in real-time).

Road Weather Management

- Vehicle probe data could be used for weather and road condition monitoring, micro-weather detection and analysis.
- Probe data from a large number of users could improve accuracy of micro-simulations, weather forecasting and analysis, and support decision-making. The Agency could use this data on a daily basis to improve winter maintenance by optimizing plow location. For example, real-time and quality data could increase Agency responsiveness in the case of a lake effect snowfall.
- Data attributes that are a priority for MDOT:
    - Vehicle location, brake actuation, ABS activation, traction control, windshield wiper status (real-time).

Transportation Planning

- MDOT could supplement the coverage of data from existing sources, such as Inrix and Waze, with vehicle probe data and cell data to improve conditions assessment (traffic volumes, vehicle counts, and travel demand), and annual average daily travel (AADT) calculation. CAV data could be used to calculate travel time and speed data across federal aid system, which is currently limited to the HPMS system and could be expanded.
- CAV data could also improve vehicle counts, which are crucial for travel model validation and verification.
- CAV data could be used to perform vehicle classification numbers on a broad scale, not just at ITS device location as it is currently possible.
- CAV data could be leveraged to model scenarios of widespread use of automated vehicles and shared-use mobility services, as well as model their implications on road capacity, zoning, and parking, and ultimately inform policy changes.
- CAV data could also be used to improve transportation models and baselines, due to its scale and breadth of input compared to traditional site specific ITS deployments and data capture.
- This data could also serve as inputs for visioning impacts of CAV on the transportation infrastructure, as part of the development of the State Long-Range Transportation Plan.
- Data could also be leveraged for linear referencing systems (LRS) linkage across segments via TAMS.
- Data attributes that are a priority for MDOT:
    - Origin and destination, vehicle position and speed, path history and prediction.
    - Vehicle class.
    - Oversize/Overweight – vehicle attributes
    - Vehicle Identifier (license plate or VIN)
    - CAV use purpose (work commute, leisure, etc.).

RISKS AND ISSUES RELATED TO DATA MANAGEMENT

MDOT and DTMB representatives identified issues and risks related to data management. Overall, there are many missed opportunities to leverage and use data effectively. In addition, almost every person interviewed mentioned the need to break down silos within the Agency.

MDOT personnel that was interviewed identified data quality as the main issue. Some of the specific data quality issues related to various data systems are listed below:

- Whereas freeway data is generally good enough (despite periodic gaps), non-freeway data has many gaps and needs richer data sources, and rural roads data is very insufficient. However, data from Inrix (XD Traffic) will be added into RITIS in the near future, which will provide more granular data for non-freeway roads.
- Naming conventions between counties, cities, and villages differ. There are abbreviation inconsistencies across datasets and systems. Metadata quality and details are insufficient. These compatibility issues between localities and MDOT make comparisons and analyses difficult.
- AVL sensor failures, speed calibration, and equipment calibration issues still exist, but there have been improvements based on lessons learned, such as planned outages out of peak use.
- Data calibration issues in HPMS have also been mentioned.
- Data quality from ITS devices is insufficient. There are also issues with calibration of ITS data in data modeling.
- There are data management struggles concerning the guardrail inventory, which is not always up to date.
- The TIM database has inconsistencies due to its sources.
- The accuracy and location of work zone data and deployment is not real-time. In addition, the validation of lane closures and opening is a serious problem for the Work Zone group.
- The quality of data regarding vehicle occupancy, traffic volume, and travel times from MVDS devices is a frequent issue for TOCs. Operator key-in accuracy and timeliness continue to be an issue for data log, as mentioned by TOCs.

Other issues and challenges mentioned in the interviews:

- Data availability and coverage is a big risk for access and quality, and data acquired from private providers may not always match use cases.
- Other issues identified relates to data storage and retention policies, for example DTMB cloud share and deletion of data without warning. MDOT representatives also identified challenges related to data storage on the common drive and Sharepoint within the agency, especially version control management. MDOT personnel mentioned issues related to data loss or back-ups caused by the corruption or loss of a hard-drive or local PC.
- Data access issues on specific systems have been reported. For example, the JamLogic Software has access issues and software update restrictions by DTMB cause outages and lack of access to key information. Currently no good method exists that can handle centralized data access for traffic volumes and crashes. There are also issues with LRS versions.
- For TOC, outages of ATMS are a top operational risk. In addition, TOCs also mentioned issues with data changes over time between systems, which causes the systems to be unable to correlate events and time/date sometimes.
- Data security related to cloud service access was identified as an important challenge by MDOT personnel.
- MDOT has access to data on slow or stopped traffic and incidents from Waze through the Connected Citizens program, but found it difficult to allocate resources to understand how to use that data and integrated it into existing MDOT systems, such as RITIS and ATMS.
- Multiple systems duplicate data types and are likely to not have any functional matching date/time stamps.
- Traditionally many proprietary systems do not integrate well with other systems which could cause duplication in data analysis efforts within the department.
- Long-term and legacy data consolidation efforts need a fresh approach that leverage lessons learned from data warehousing and master data management methods and strategies.

- Storing data without applying appropriate data cleaning and classification processes could further hamper long-term data value and decision support system analysis.
- Duplicative data collection, analysis and data sources are a critical issue for MDOT (including the data definition inconsistencies between systems and data sources).
- A secondary challenge for MDOT and many Agencies is the communization of a consistent and accurate time source across systems, devices and data sources/probes – this inhibits consistency and has a downstream effect on data quality/verification efforts, as well as an impact in the transformation of data. Lastly, in the transform process with inconsistent time sequencing severely impacts the value and performance for an enterprise data aggregation system, big data analytics and nearly all MDM designs/implementations.
- MDOT business unites have taken good steps toward developing a more effective and efficient data sharing system, however better communication around these efforts is needed through the department.

CHALLENGES SPECIFIC TO CAV DATA MANAGEMENT

For the Operations Field Services Division, the rate of changes related to CAV data availability is the biggest challenge. In this context, understanding and assessing opportunities for existing and new data, as well as turning raw data in useful actionable information is difficult. Another challenge has been to leverage best practices and learn from other state and local transportation agencies activities related to CAV data management and applications.

Because CAV datasets are massive, it is difficult for MDOT personnel to identify issues, validate data, maintain confidence in these datasets, and use them with existing MDOT tools. There is also limited user knowledge on data processing options at MDOT. For these reasons, automated data processing tools and ETL methods should be undertaken by MDOT to improve data quality and the implementation of an enterprise data aggregation that rests aside from the core DUAP system.

MDOT representatives recognize that several datasets are currently underutilized and it would be beneficial to develop new functionalities to take advantage of that data and transform it into information that can improve MDOT processes. Linkages across data silos are still challenging, and data fusion is currently both a risk and an opportunity for the agency. The fact that there are currently several separate initiatives to collect asset and device data without a centralized repository represents a challenge. Several interviewees repeated the need for MDOT to have a data warehouse at an enterprise level. DUAP could accomplish this role, but the interviewees deemed access and effectiveness to DUAP too complicated and device data is often not sufficient or timely.

In the long term, storage of CAV data will become an issue as well.

Finally, a significant challenge in the future will be balancing between current needs, opportunities to fuse data within existing data systems, and adding new data altogether to systems.

4.2 UPDATE ON PREVIOUS RECOMMENDATIONS

The 2015 report recommended developing a statewide master/strategic plan for database aggregation across ITS subsystems and programs – integrating GIS, Asset Management and ITS datasets – as part of the Transportation Asset Management System (TAMS) project to develop the plan. However, data collected by ITS devices as well as its analysis and application would not be integrated in TAMS due to its real-time nature and storage volume; DUAP is a repository and aggregator of certain datasets within MDOT. MDOT's ITS deployments are still single-purposed and non-integrated to a large extent; MDOT should pursue efforts in identifying overlapping features and functions of MDOT ITS subsystems and assessing potential for consolidation of ITS subsystems and reduction in long-term costs, as recommended in 2015.

The previous report recommended developing a data use and integration policy in conjunction with IT, legal, and other key stakeholders. MDOT's Data Governance Council is currently leading that effort with the help of a consulting firm. In addition, as part of the TSMO program, a group of MDOT subject matter experts has been developing recommendations to better integrate ITS systems and CAV data. However, this initiative should be structured carefully, and connected to similar initiatives in a more deliberate way, with a focus on data implementation by involving more subject matter experts and data end users, as more CAV data applications are developed.

The 2015 report recommended engaging in an overhaul and renegotiation of services, support, and performance with IT and MDOT business units during project deployments and system integration. Those renegotiations have been occurring on a system-by-system basis. Having several data systems that are operating on different contractsand development timelines but are or should be functionally interconnected remains a challenge for MDOT.

Several MDOT units interviewed for this report stated the need to pursue work towards having an enterprise data warehouse and reducing duplicative efforts, as was recommended by the 2015 report. The first step would be to establish and demonstrate a controlled scale proof-of-concept with a series of dashboards, visualization, online mapping, and web-based user applications, both internal for MDOT personnel and external for public user subscribers. If that is deemed successful, MDOT should establish a migration plan, and then procure and setup the enterprise data warehouse hardware and software. For the enterprise database management software, it is recommended to opt for a solution that supports native high-availability clustering, ideally an open-source solution like Hadoop or MongoDB and leverage data from current enterprise systems (effectively placing the enterprise analysis, performance measures, decision support system over top of all existing database/data aggregation points and siloed systems in place currently). MDOT's Data Governance Council should evaluate these open-source tools on a bi-annual basis.

As part of this enterprise data warehouse, the 2015 report also recommended aggregating data in a series of phases by ITS subsystem, based on a prioritization of the datasets. For the database and hardware type, they use Oracle or DB2 running on dedicated application hardware with a separate storage platform/architecture is recommended for this load condition. Also, for uses and applications involving big data and real-time web applications, NoSQL databases could be considered. The 2015 report recommended the following data integration and retention process:

- Phase 1: Integrate real-time data (one per minute or one per five-minute intervals) from key ITS subsystems, with a retention period of 45 days being represented as current data.
- Phase 2: After 45 days, store data for trending and current analysis due to recent occurrence.
- Phase 3: After 12 months, move data on an annual basis to secondary storage disks to allow for rapid access but represented as archival. This

would allow for longer-term business analytics and metrics analysis/trending.

- Phase 4: Once data is past three years old, archive it to a hierarchical storage platform for periodic use and access for up to five years and then evaluate methods to identify and control automated disposition of data that is no longer used or used very infrequently.

During the interviews conducted for this project, MDOT personnel mentioned it is important to review the existing separate initiatives dedicated to collecting asset and device data and pursue the discussion on a centralized repository, such as DUAP. Agents have also suggested that data collected with LiDAR could be integrated in DUAP. More broadly, MDOT personnel has pointed out that in the future enterprise data warehouse, it is crucial to point out which datasets have been verified and are of good quality.

The 2015 report made a series of recommendations related to data security and privacy. MDOT personnel interviewed for the present report recognized that policies on data confidentiality and personally identifiable information (PII) still need to be defined across systems by MDOT. However, most of these issues fall under the responsibility of DTMB. The 2015 report recommended that for any implementation of an enterprise data warehouse, MDOT should:

- Implement security practices with their network domain (Lightweight Directory Access Protocol (LDAP), Virtual Private Network (VPN), Public Key Infrastructure (PKI) systems), with regimented database and web application patch management practice.
- Maintain both hardware/software security devices (e.g., load balancers, web application firewalls, firewalls, Demilitarized Zone (DMZ), etc.).
- Use encryption methods such as Secured Socket Layers (SSL) with active wildcards certificates across the application-tier on the front-end applications.
- Remove vehicle and passengers' private data or minimally encrypt these key data when managing real-time CV data.
- Focus on the integration of data that does not have privacy or legal publication/distribution issues and secure it using a minimalist approach.
- Establish user groups to identify the value of the user data and MDOT applications that make use of it, thereby helping to reduce the risk of negative public perception on redistribution of user data. Currently, when a third party makes a data request, it goes to the Attorney General first and

then to MDOT. The Department has a single individual responsible for reviewing MDOT data requests and it is important that MDOT builds more in-house expertise on that front.

Finally, the previous report suggested to give each ITS subsystem a full legal review to determine the current rights and restrictions for use, data, database schema, and data publication. MDOT performs this review on a system-by-system basis at the time of contracting.

4.3 RECOMMENDATIONS FOR DATA MANAGEMENT PROCEDURES, POLICIES AND REQUIREMENTS

PROCEDURES

Based on the interviews conducted, MDOT personnel, especially from the Bureaus of Transportation Planning and Field Services, need more training on:

- What CAV data is currently available or will be available soon, how to access it, what are data gaps, and how to fill them.
- How to analyze and use CAV data accurately (data mining skills) – especially traffic data, vehicle counts and class data.
- What are DUAP's current and future functionalities and how can these be useful to MDOT units.
- Universal view of all MDOT data systems and datasets, leveraging findings from the present report.
- Design and implement a top tier data aggregation system over top core existing data systems as a "fusion, data analytics and enterprise aggregation point" to commonize critical data from each underlying critical data system (DUAP, TAMS, TIM, RITIS, ATMS) to support key workflows and decisions support functions within the Department (Modelling, Planning, Design, Operations and TOCs). Big data models and transportation analytics in this aggregation system need to be implemented for high-speed and automated analytics, ETL processes and real-time data pipelines to improve the storage performance (input and output) from the enterprise-class MongoDB or Hadoop.

The ITS team could lead these types of initiatives and could, for example, organize bi-annual training and brainstorming sessions on CAV data management, and develop training documentation for MDOT personnel, especially to units within the Bureau of Transportation Planning and the Bureau of Field Services, including Transportation Operation Centers.

In addition to that, there is a need for more cross-division and cross-unit discussion on the potential benefits of CAV data, the opportunities and specific use cases, and assessment of the real added value of CAV data to improving existing MDOT data systems.

MDOT interviewees mentioned that the benefits of collecting probe vehicle data on weather and pavement condition and how that can be used to improve metrics for pavement conditions are not yet well understood within the Agency. One of the 2017-2020 projects within the MDOT contract "Specialty Services: Connected and Automated Vehicle Technology (Req. No. 2193) is "Implementation Recommendations For Its/Cav Data Support Of Asset Management." Among the goals of this project should be disseminating research on the benefits of vehicle data on weather and pavement condition within the Agency.

MDOT's Data Governance Council, chaired by the Chief Data Steward, needs to continue and reinforce its work concerning:

- Maintaining a thorough inventory of MDOT data systems: data uses, gaps, sources, and quality.
- Develop tools and procedures to better document and share knowledge about all MDOT data systems and how to best use existing data and new CAV data to support operations.
- Reviewing the distribution of responsibilities and dependencies between MDOT units related to data management.
- Identifying and implementing steps to eliminate specific silos between MDOT units that hinder data management, procurement, and sharing.
- Integrating CAV data with existing data using current methods and tools, as well as developing new methods and tools.
- Continuing to improve data sharing across MDOT business units.
- Define data management practices and communicate across business units.

It is important to include DTMB in Data Governance Council meetings to improve DTMB's understanding of CAV data use cases and requirements from MDOT, and to make the liaison with the Enterprise Information Management initiative led by DTMB.

POLICIES

MDOT should accelerate work towards developing an "Enterprise Data Aggregation and Management Procedures Plan" across the agency and in coordination with partner Agencies.

Based on interview feedback, data retention policies should be reviewed and defined more effectively for each data system. MDOT needs to continue identifying legacy systems limitations for data use, export and reporting and to develop improvements. For example, MDOT personnel mentioned that some datasets older than ten years are unlikely to be used in the future. Because their storage has an impact on server space and performance, these datasets could be archived or deleted. In addition, data storage policies need to be revised regarding how frequently to add storage. Private cloud data storage would be a substantial benefit for the Department, including data fusion, analysis and decision support systems portal access.

MDOT Divisions have made progress in terms of data integration, but improvements can still be made. For example, more design and analysis efforts are needed to tie key data together on the TIM database and Incident data to drive the time/space diagram and key analysis outputs.

More broadly, MDOT should revise data definitions to make sure they are consistent across data systems and business units. All data systems should have detailed data glossaries developed by data consumers and producers, especially regarding CAV and ITS data.

REQUIREMENTS

Based on MDOT input, vehicle probe data is most susceptible to quality issues and therefore needs a thorough review prior to use and analysis.

Based on interview feedback and data systems analysis, the Bureau of Transportation Planning could make the following improvements:

- Processing and handling of Census data should be more efficient. This could include custom developed GIS tools to leverage and automate processing of

Census data into enterprise MDOT geodatabase format by using current Engineering Consultant contracts.
- The time to validate employment data purchased from vendors reduced. Currently, data use/licensing restrictions hinder that.
- Data validation should be done first within MDOT and then by Road Commissions, Metropolitan Planning Organizations, and other agencies.

Based on interview feedback and data systems analysis, the Bureau of Field Services could make the following improvements:

- The Maintenance Decision Support System (MDSS), developed through a pooled fund study, could use more specific MDOT upgrades and could be integrated with GIS.
- Within the Property Damage and Reclamation Process (PDRP) there should be a requirement to add latitudinal and longitudinal data or geolocation data as part of crash reports to be able to tie a specific asset as part of the PDRP. The PDRP could also be improved for mobile use.
- There is an opportunity to convert the Local Area Payment System (LAPS) to web-function or mobile use.
- ATMS needs to have better information and be more efficient for operator input. For example, the call tracker extract to ATMS is problematic.
- The Traffic Incident Management (TIM) input and output workflows need to improve in efficiency.
- The quarterly Safety and Security export spreadsheet could be fused with the Maintenance/TIM database for data sharing and distribution.
- Access to data should be provided in usable formats, not .pdf exports.
- AVL could be integrated with GIS.
- Reducing data variability between TOCs should be a priority.

5 REFERENCES

Brinckerhoff, P., & Iteris. (2013). *Road Weather Information System Evaluation.* MDOT.

Cambridge Systematics Inc.; Boston StrategiesInternatiomal Inc.; Gordon Proctor & Associates; Markow, Michael. (2010). *Target-Setting Methods and Data Management to Support Performance-Based Resource Allocation by Transportation Agencies* (Vol. Report 666). Washington, D.C: (NCHRP), National Cooperative Highway Research Program.

Castle, C. (2015). Michigan DOT Connected Vehicle Data Use. Presentation to V2I Deployment Coalition_TWG3 (partnership).

CATT Laboratory. University of Maryland. (2015). *RITIS Platform;Features & Applications Overview.* Retrieved from http://www.cattlab.umd.edu/files/RITIS%20Overview%20Book-2-2-15%20FINAL.pdf

Droste, J. (2017). MDOT Transportation Asset Management System (TAMS). *MAASTO Conference.* Cleveland.

Forrester. (2016). *Big Data Management Forecast 2016 to 2021*. Retrieved from Forrester: https://www.forrester.com/report/Forrester+Data+Big+Data+Management+Solutions+Forecast+2016+To+2021+Global/-/E-RES135913

Hall, J. (2017). Data Governance at State Departments of Transportation. *MWAIS*(24).

Loshin, D. (2006). *Defining Master Data*. Retrieved from B-eye-Network: http://www.b-eye-network.com/view/2918

MDOT. (2016). Truck Parking Information Management System Request for Proposal Form.

6 APPENDIX A: INTERVIEW QUESTIONNAIRE AND LIST OF INTERVIEWEES

MDOT TASK A3

Implementation Recommendations for Management Procedures for Data Collected via CAV

Interview Questions

1. Which systems within MDOT do you or your group use frequently – by sharing or using data between? How frequently do you use each system?
2. Which systems within MDOT do you or your group send data to, and are those systems physically located within MDOT?
3. What key systems do you use today and what is their role? Which users access the "key systems", are they vended/contracted or maintained internally, and what are they mainly used for?
4. What other types of data or systems do you use from within MDOT? What are the inputs and outputs of the systems you "predominantly use"?
5. To improve your job, efficiency, data quality, analysis, or reduce the time needed to complete tasks please describe what additional data (data fields/attributes) would be helpful in the systems you use now?
6. What other types of data does your department use from external sources (outside MDOT)? Which Agencies/Departments do you interact most frequently (how frequently)?
7. With the introduction of Connected and Automated Vehicle (CAV) data for transportation Agencies – how would you leverage or best use that data "source"?
8. What opportunities or data uses do you see from access to a rich/deep CAV datasets? If you had access to CAV data how frequently would you use the data?

9. What CAV data attributes would be a priority for you?
10. Please list some specific issues related to data management (e.g. retention polies, storage availability, limited access, data quality, timing, redundant data, system performance, uptime/network availability, etc.) How do these data issues impact your unit?
11. What are your top five systems risks (outages, breaches, corruption, performance, etc.) and which systems are most at risk?
12. If you could make changes to any of the top risk systems, what changes would you propose?
13. Which systems are most and least susceptible to data quality issues?
14. Are there other specific issues related to data management (e.g. limited access, timing, interoperability or compatibility)? How do these data issues impact your unit?
15. What uses, analysis opportunities or data needs are not being met by existing systems?
16. What are the biggest challenges for CAV/ITS data management (e.g., rapid advancements in CAV/ITS technologies, real-time data collection and storage, data standards, data querying and user access, IT and hardware costs, institutional issues, enforcement of security, privacy, and trust)?
17. Do you have any other suggestions to improve data management at MDOT or within your group?
18. Which systems do you or your group / department maintain?
19. If a new business need or system was identified, who would have to be involved to implement the new system and get it functional? What is the typical timeframe (avg. time required in months) to setup a new system?
20. If a new business need or dataset (change in database) was identified, who would have to be involved to implement the new changes? What is the typical timeframe (avg. time required in months) to make system changes for systems you or your group maintain?

21. For the systems, you or your group maintain is system documentation available in terms supplying Concept of Operations or User Guides to our team?
22. Do all systems share a common Database Management System (DBMS) type? If not, please identify and describe the DBMS type, version, and vended/maintained by third party per system.
23. For each key ITS/CAV and associated database, what is the storage size per system (e.g., 5 GB average, 10 GB average, 30 GB average, 100 GB average)?
24. How often is each system you maintain or use updated with new data (e.g., hourly, daily, weekly, monthly, annual)?
25. For systems that you or your group maintains how frequently is storage space added by system; please identify which systems need the most data changes and type of maintenance?
26. What CAV/ITS data systems are the most important for your work?
27. Do you have any other suggestions to improve CAV/ITS data management?

List of Interviewees

- Niles Annelin – Statewide Planning Division, Bureau of Transportation Planning, MDOT
- Dan Belcher – Design Services Unit, Design Division, Bureau of Development, MDOT
- Chris Brookes – Traffic Incident Management Unit, Operations Field Services Division, Bureau of Field Services, MDOT
- Wendy Burton – Data Inventory and Integration Division, Bureau of Transportation Planning, MDOT
- Mark Burrows - DTMB
- Collin Castle – ITS Program Office, Operations Field Services Division, Bureau of Field Services, MDOT
- Monica Coulter – IT Architect, DTMB

- Tim Croze – Maintenance Services Unit, Maintenance and Operations Section, Operations Field Services Division, Bureau of Field Services
- Jason Firman – Congestion and Mobility Unit, Operations Field Services Division, Bureau of Field Services, MDOT
- Sarah Gill – Southeast Michigan TOC, Operations Field Services Division, Bureau of Field Services, MDOT
- Dave Jannett, MDOT
- Annjanette Kremer – Traffic Incident Management Unit, Operations Field Services Division, Bureau of Field Services, MDOT
- Melissa Longworth – Maintenance Operations Section, Operations Field Services Division, Bureau of Field Services, MDOT
- Kevin McKnight – Statewide Planning Division, Bureau of Transportation Planning, MDOT
- Erica Moore – MDOT
- Hillary Owen – Statewide TOC, Operations Field Services Division, Bureau of Field Services, MDOT
- Suzette Peplinski – West Michigan TOC, Operations Field Services Division, Bureau of Field Services, MDOT
- Brad Sharlow –Urban Travel Analysis Unit, Statewide Planning Division, Bureau of Transportation Planning, MDOT