Management Procedures for Data Collected via Intelligent Transportation Systems

Qiang Hong¹, Richard Wallace¹, Eric Dennis¹, Brian Reed², William Tansil³ and Matt Smith³
1. Center for Automotive Research, Ann Arbor 48108, USA
2. Geospatial & Applied Technologies Group, Parson Brinckerhoff, Baltimore 21201, USA
3. Michigan Department of Transportation, Lansing 30050, USA

Abstract: Addressing transportation planning, operation and investment challenges requires increasingly sophisticated data and information management strategies. ITS (intelligent transportation systems) and CV (connected vehicle) technologies represent a new approach to capturing and using needed transportation data in real time or near real time. In the case of Michigan, several ITS programs have been launched successfully, but independently of each other. The objective of this research is to evaluate and assess all important factors that will influence the collection, management and use of ITS data, and recommend strategies to develop integrated, dynamic and adaptive data management systems for state transportation agencies.

Key words: ITS, connected vehicle technology, transportation data management, state transportation agency.

1. Introduction

The benefits of being proactive in ITS (intelligent transportation systems) adoption are numerous. ITS can facilitate a safe and efficient transportation system, improve public safety, assist research programs, stimulate economic growth and improve the environment. However, a disadvantage in being at the forefront of technological advances is that technologies may be deployed before they are fully mature. In the case of Michigan, several ITS programs have been launched successfully, but independently of each other. ITS-generated data are similar to other transportation data collected by traditional methods, but they are collected continuously and are greater in quantity and temporal coverage. The growth rate, cost in storage, backup/retention and recovery management require a comprehensive data management plan across key stakeholders/critical member agencies to address cost effectiveness in managing growth and data integration. In other words, an efficient and cost effective data management strategy for dealing with the present needs and future growth of the ITS data will be critical to successful ITS programs.

The purpose of this research is to evaluate and assess all important factors that will influence the collection, management and use of ITS data, and recommend strategies to develop integrated, dynamic and adaptive data management systems for state transportation agencies. To effectively capitalize and leverage existing and future opportunities, the State of Michigan should develop a statewide master/strategic plan for database aggregation across ITS, GIS (geographic information systems) and transportation asset management subsystems and programs. The plan should be developed in conjunction with key stakeholders including MDOT (Michigan Department of Transportation), the Michigan DTMB (Department of Technology, Management and Budget) and other key stakeholders.

2. Related Literature

This research investigates a broad range of academic articles, public sector reports, industry standards and
other sources, in an effort to incorporate the most current and salient information into a foundational set of concepts from which to build broadly applicable and adaptable strategies of data management.

2.1 ITS Data

ITS data refer to data that are typically collected by and/or generated for ITS applications, such as ATMS (advanced traffic management systems), ATIS (advanced traveler information system), CVO (commercial vehicle operation), EMS (emergency management systems) and the emerging CV (connected vehicle) systems. Archived ITS-generated data are distinct from traditional data sources in three aspects: (1) ITS-generated data are temporally intensive—that is, collected in very short intervals; (2) ITS generated data meet some major data gaps that could not be met in the past due to resource limitations; (3) All ITS-generated data are on electronic media, thereby expediting data analysis and information dissemination. Common ITS data collection methods include:

- vehicle detection;
- video imaging;
- AVL (automatic vehicle location);
- in-vehicle transponders;
- license plate readers;
- cellular probes;
- GPS (global positioning system) probes;
- Bluetooth-based measuring systems;
- transit ITS;
- commercial vehicle operations and management;
- connected and automated vehicle technologies;
- telematics [1].

Many of these data collection methods have been used by public agencies charged with managing traffic flow, safety, carrying out planning activities and maintaining roadway networks. Some, such as sensing and telematics that will allow for wireless V2V (vehicle-to-vehicle) and V2I (vehicle-to-infrastructure) communications for purposes related to safety, security, convenience and maintenance, are relatively new and innovative.

2.2 Data Management Practices

Data management can be defined as the development, execution and oversight of architectures, policies, practices and procedures to manage the information life-cycle needs of an enterprise in an effective manner as it pertains to data collection, storage, security, data inventory, analysis, quality control, reporting and visualization [2]. There are two key areas to creating and sustaining effective data management systems. The first area centers on the technical challenges including development and maintenance of hardware and software and the specifications for data collection, analysis, archiving and reporting. The second area focuses on the institutional issues associated with data stewardship and data governance. Evaluation of data management programs may use criteria such as accuracy, timeliness, completeness, validity, coverage and accessibility [3].

Primary concerns in current research and development center on making sensor and mobile data sources accessible, usable and secure. Modern wired and wireless sensor networks, including ITS, must often handle real-time storage and data querying, possibly from multiple sources and formats [4]. Long-term data storage generally requires some form of archiving. Key actors during the process include data administrators, data users, data providers, government report systems, remote archives and financial institutions (when payments are needed for certain services) [5].

2.3 Primary Approaches to Data Storage and Query

Wireless sensor networks are generally connected to databases that manage the data. Processing time is critical in time-sensitive applications such as ITS. In these cases, query and response transactions must run in well-defined time intervals. For this reason, a RT-DBMS (real-time database management system) must be used to process transactions on time while
maintaining logical and temporal consistency of data [4]. The two main approaches to sensor database storage and query are the warehousing approach and the distributed approach, shown in Fig. 1. A hybrid approach is a third option.

2.3.1 Warehousing Approach

In a warehousing approach to data storage, data are stored in a central database and queries may be performed on it. This is the most simple and common arrangement, but is prone to time delay and bottlenecks on the central server [4], such as a TOC (traffic operations center) [6]. The limitations of central warehousing databases have led to the design of data indexing methods for more efficient spatiotemporal queries to sensor data. The STEWARD (Statewide Transportation Engineering Warehouse for Archived Regional Data) implemented by the Florida DOT in 2008 was one example of the central data repository system [7]. STEWARD demonstrated that data from regional TMC (traffic management centers) can be centrally archived in a practical manner and that a variety of useful reports and other products can be created. Capacities include:

• identifying detector malfunctions;
• providing calibration guidance for detectors;
• performing quality assessment on data;
• providing daily performance measures for systems and statewide performance measures;
• facilitating periodic reporting requirements;
• providing data for research and special studies.

2.3.2 Distributed Approach

In a distributed system, data are stored both at the central server and in the sensor devices themselves. The sensors generally have some processing capabilities and act as local databases. This approach has advantages over the warehousing approach because it processes queries faster, which can support long-running and instant queries, and uses less energy (by sending less information to a central server) [4]. In Ref. [8], the authors suggest storing massive unstructured data as BLOBs (binary large objects) with a versioning-oriented interface. Such a distributed approach would be necessary in peer-to-peer applications such as V2V communication. For peer-to-peer systems, devices will have to use a common language for query and response, such as TranQuyl. TranQuyl also can function in client-server mode for queries to the database [9].

Fig. 1 Primary approaches to data storage and query: (a) warehousing approach; (b) distributed approach. Source: Ousmane et al. [4].
2.3.3 Hybrid Approach: Mobile Sensor Network Data Management

Many real-world and conceptualized ITS systems employ a hybrid approach to data management and storage, e.g., a mobile sensor network consisting of IRSs (intelligent roadside stations), STAs (smart traffic agents) and a central database. The Encyclopedia of Database Systems defines MSN (mobile sensor network) data management as “a collection of centralized and distributed algorithms, architectures and systems to handle (store, process and analyze) the immense amount of spatial-temporal data that is cooperatively generated by collections of sensing devices that move in space over time” [10]. Advantages of MSNs include:

- dynamic network coverage;
- data routing repair;
- data nulling;
- staged data stream processing;
- user access points.

The most predominant data management challenges that have prevailed in the context of MSNs include:

- in-network storage;
- flexible and expressive query types;
- efficient query routing trees;
- purpose-driven data reduction;
- perimeter construction and swarm-like behavior;
- enforcement of security, privacy and trust.

2.4 Data Collection and Fusion Techniques

RT-DBMSs are essential to ITS ATMS [6]. Such applications may often have to combine unlike data from multiple sources to create information that is useful to end-users. This is a problem if data queries are subject to time constraints or if competing data provide an unclear or complex response. In such a case, data fusion techniques can be used to integrate data and create new data that is an improvement on raw data from an operational standpoint [11].

In traffic management and operation, data fusion may be useful for such applications as advanced traveler information, automatic incident detection, advanced driver assistance, transportation network control, traffic monitoring and traffic forecasting. These applications may improve transportation system safety and efficiency [12].

2.5 Data Distribution to Users

Distributing usable information to users is an essential component of an ATIS. One popular method of public information exchange is wiki-like platforms [6]. One complicating factor of data distribution and utilization is the temporal limitations of real-time data. For example, for a user plotting a quickest route considering traffic congestion, the most efficient route would need to anticipate cyclical congestion at the time the user would be anticipated to travel across a section of the network, rather than use only real-time data from the instant of the query response. Thus, a traffic status prediction method should be incorporated into time-sensitive applications such as real-time route planning within the ITS [6].

3. Analysis of Existing in Michigan Department of Transportation

MDOT currently has a series of transportation and ITS subsystems. These subsystems are briefly specified in following subsections.

3.1 ATMS (Advanced Traffic Management Systems)

ATMS platform has been recently adopted as a single statewide ITS control software package. The system provides control, management, and monitoring of travel time, DMS (dynamic message signs), vehicle congestion (through detectors), and CCTV (closed-circuit television) camera live-feeds. This management platform supports real-time operations and incident management through basic data integration from roadways with ITS devices already deployed, managed, and communications connections established. In conjunction with vehicle detectors along the roadway, travel times are projected using the
Management Procedures for Data Collected via Intelligent Transportation Systems

Data obtained via a third-party contract using vehicle probes.

3.2 RWIS (Road Weather Information System)

MDOT’s RWIS is an ITS subsystem expressly used to collect, analyze and report current roadway surface conditions (e.g., salinity, surface temperature, rate of cooling, etc.) and specific atmospheric conditions (e.g., ambient air temperature, precipitation, humidity, wind speed/direction, visibility, etc.). Through the effective use of RWIS data, operations staff can plan when to deploy maintenance fleets for targeted operations.

3.3 VIDAS (Vehicle-Based Information and Data Acquisition System)

VIDAS is an MDOT project intended to use visual observations to determine the accuracy and reliability of tested mobile data relating to road conditions. MDOT will utilize instrumented fleet vehicles for business operations travelling roadway infrastructure on a daily basis. With the instrumented vehicles, the datasets will be collected on a real-time basis and aggregated into a database with a presentation tier user application used to visualize and interact with the data. This system allows MDOT to evaluate the performance and accuracy of roadway data for use in pavement condition and management, as well as roadway surface conditions (cracking, heaving, etc.). As a final step in the project, MDOT will also compare visual observations to determine the accuracy of RWIS data.

3.4 Teletrac Data Server

Teletrac data server is a separate ITS subsystem utilized for fleet management and performance tracking. The system supports analysis of routes, miles travelled, maintenance cycles, regulatory compliance, and GPS tracking/routing. This system is a vendor solution and has a degree of overlay with OEM (original equipment manufacturer) fleet data server. Currently, this system has a limited deployment of less than 50 vehicles in total.

3.5 OEM Fleet Data Server

OEM fleet data server is a separate ITS subsystem utilized for fleet management and reporting. This system was developed in-house. The deployment scale and function were not reported well during interviews, so it is challenging to address the strategic differences between Teletrac and this solution.

3.6 HPMS (Highway Performance Monitoring System)

HPMS is a separate transportation data subsystem utilized to collect vehicle counts and classification. These data are used for reporting quantities and types back to the USDOT (US Department of Transportation) and establishing funding formulas for transportation funding. Additionally, many jurisdictions utilize these data in conjunction with real-time vehicle detectors to supplement volume and capacity data in support of congestion and travel time analysis. Lastly, many jurisdictions utilize these data to supplement and analyze locations to target enforcement of oversize or overweight vehicles from the vehicle classification types.

3.7 AMS (Asset Management System)

AMS is a separate subsystem utilized to analyze, manage, monitor, and report performance conditions on a wide variety of fixed assets within MDOT. The AMS is a key business differentiator, core business system, and suite of processes utilized within MDOT. AMS has provided MDOT a platform to establish, track, and report key metrics to legislators effectively to justify funding increases for the overall program.

3.8 TMS (Transportation Management System)

The TMS is an integrated and automated decision-support tool that has capability to identify condition, analyze usage patterns, and determine deficiencies of the transportation infrastructure. It is designed as a single management system with six components or subsystems. These systems include: bridge, congestion, intermodal, pavement, public
transportation and safety. This allows the TMS to include a common shared database, a common set of decision-support tools and functionality, and the use of a robust and consistent user interface. Data collected, processed, and maintained at the working levels are stored using an enterprise database management system. MDOT envisions TMS as one integrated management system, encompassing all functional subsystems, designed to be able to be expanded into other areas of MDOT’s asset and mobility management areas.

3.9 DUAP (Data Use Analysis and Processing) Program

The DUAP program was initiated in 2006 by MDOT to complement research initiatives from the USDOT, the VIIC (vehicle infrastructure integration consortium), and others to design and deploy the CV infrastructure, vehicle equipment, and initial applications. The intent of the program is to support performance management by enhancing agency-wide usage of connected vehicle and mobile data and increasing data sharing, availability, and awareness across the agency. The recently completed DUAP 2 ConOps (concept of operations) describes the advanced applications of connected vehicle data, concepts for the proposed systems, and operation scenarios [13]. The system will be designed to serve a multitude of purposes, as well as be expandable, scalable and modular. The physical DUAP 2 system will consist of several subsystems, such as data collection, data management, and data distribution systems (Fig. 2).

3.10 Implications of Existing Data Systems

After reviewing these data systems, striking resemblances of the summaries emerge, which highlight several systems do not account for holistic data integration between subsystems. Without an effective and efficient plan to merge the data into managed repositories, both asset management and transportation infrastructure operations become extremely difficult and ever more costly for agencies and tax payers. If State Departments of Transportation and other specific agencies recognize the amount of real-time data that can be leveraged and integrated, there is distinct opportunity to develop effective pavement management, asset management, and operation management practices using data that are currently accessible within the enterprise.

Secondarily, there is a larger and more dramatic opportunity to capture and obtain value from real-time datasets stored from mobile assets or connected vehicles. The DUAP program is the mechanism to bring these vehicle and dynamic datasets together in a meaningful way for MDOT. This program, while addressing vehicle and services datasets for the department, does not effectively address the asset management, spatial datasets or methods to effectively manage and process the dynamic vehicle data into an enterprise data warehouse. As these mobile assets become more of a reality, the data obtained from the vehicle increase and require integration with spatial, temporal, and asset data persisted in an enterprise data warehouse.

![DUAP 2 data flows](image-url)
From a GIS standpoint, most transportation datasets have a fixed asset perspective/focus (roadway, railway, signage, addresses, hydrographic features, bridges, etc.). Traditionally, the GIS world of data and management does not interact with or consume production data directly from real-time ITS subsystems. Far too often, each agency takes responsibility for its own datasets and does not create or leverage operational agreements to share data or create a master data repository with appropriate failover/resumption designs.

Currently, most of the existing ITS subsystems datasets are rather small and are not generating substantial data storage requirements (30 GB (gigabytes) or less annually). The DUAP subsystem and any integration efforts undertaken to aggregate datasets within a larger enterprise data warehouse will change the storage requirements rather substantially (1-3 TB (terabytes) on an annual basis). The key drivers in this storage increase are the statewide/regional stakeholder orthophotography, integrated data between subsystems and real-time data analytics/datasets from multiple vehicles (ongoing CV programs). While the real-time CV datasets offer huge opportunities and rich-information datasets, the ongoing trade-off will be costs associated with long-term storage, archival and data management practices.

System maintenance, database backups and the long-standing restricted system support from DTMB, put MDOT in a challenging position to deploy, maintain, and achieve value in industry-leading solutions. As a result, it may be more effective and efficient for MDOT to increase in-house staffing and IT (information technology) training to support new solutions being procured and deployed within the agency. This would allow for MDOT IT personnel to strategically support ITS subsystems, ITS communications network (center to field and devices) as a network segment, and servers/storage equipment.

Finally, as a key lesson learned from vendor deployed systems, there are several items that need to be in place prior to system solution procurement. First, data use agreements must include data warehousing and integration and publication distribution both internally and externally to MDOT without excess fees or license cost. Second, each vended solution must be turned-over to MDOT at acceptance testing with a full data dictionary and database entity relationship diagram of the database deployed for each ITS subsystem. These conditions will enable MDOT to effectively utilize the disparate ITS subsystem data, without requiring database schema reverse-engineering or substantial database analysis to determine the schema implemented by the vended solution.

4. Conclusions and Recommendations

In both private and public-sector activities, Michigan has established itself as a national leader in roadway, fixed asset management, GIS, and vehicle technology. To effectively capitalize and leverage each of these opportunities, MDOT should initially develop a statewide master/strategic plan for database aggregation across ITS subsystems and programs. The strategic plan or new data initiatives would facilitate and establish the framework for an enterprise data warehouse integrating GIS, AMS, and ITS datasets. Each information model (GIS, AMS, and ITS) should contain best practices for database schema design and integration, leveraging a core geospatially enabled and accurate base map (centerline and cadastral data layers), which are effectively maintained through the MGF (Michigan Geographical Framework). Statewide LiDAR and orthophoto datasets should be flown on an annual or biennial basis and integrated with the baseline GIS data model within the data warehouse.

In many application and system architectures, the interdependence and best value of information are derived from the integration across subsystems and datasets: this may often require subsystems that appear unrelated to be joined within a common enterprise data warehouse in support of the entire agency. A prime example of this best practice is the integration
geospatial/GIS, telemetry (vehicle to vehicle and vehicle to infrastructure), telematics (controller area network/vehicular data and connected vehicle), asset management, and highway performance data. By integrating some of these non-ITS specific subsystems, we gain the ability to perform business analytics and better spatial and temporal analysis across datasets geospatially tied to real-time positions. The potential integrated ITS subsystems are outlined in Fig. 3.

Once a migration plan is established from the funding and initiation of the strategic plan, the enterprise data warehouse hardware and software should be procured and setup. This would allow for a controlled scale proof-of-concept to be established and demonstrated from a use perspective with a series of dashboards, visualization, online mapping, and web-based user applications, both internal for MDOT personnel and external for public user subscribers. The data aggregation should be undertaken in a series of phases by ITS subsystem and performed based on prioritization of the datasets. Traditional approaches to such a DBMS (database management system) include Oracle or DB2 running on dedicated application hardware with separate storage platform architecture. Recent applications of NoSQL databases, however, increasingly have been targeted at uses involving big data and real-time web applications.

**Fig. 3 Integration of ITS subsystems.**
Data aggregation would be best served by integrating real-time (one per minute or one per 5 min intervals) from key ITS subsystems, with a retention period of 45 days being represented as current data. Once data moves past 45 days to 12 months, data should be represented and stored for trending and current analysis due to its recent occurrence. After 12 months, data should be moved 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. Once data are past five years old, it can be permanently archived to a hierarchical storage platform for periodic use and access.

Data security and privacy are large and distinct concerns of the public at-large, key stakeholders and funding providers. Each ITS subsystem should have a full legal review to determine the current rights and restrictions for use, data, database schema and data publication. If any of the license agreements are too restrictive to support data integration and eventual enterprise data warehouse, ongoing maintenance contracts should be re-negotiated to include more favorable terms for data use and publication.

Both the DTMB and MDOT have initiated projects that would allow for the design and implementation of an agency-wide DBMS. DTMB’s DMAG (data modernization assurance and governance) initiative is expected to update data management processes relevant to all State of Michigan Departments. MDOT’s DUAP 2 program aims to allow improved performance management and data sharing across MDOT by designing an enterprise data management system that will allow for migration of IT functions of all MDOT’s various business processes to a single, dynamic and scalable system. Implementation of DUAP 2, however, will require coordination with statewide IT services, such as MGF, managed by DTMB. A potential link between DUAP 2 efforts and DTMB coordination is the joint DTMB and MDOT TAMS (transportation asset management system) initiative. While the TAMS project is still developing, it is intended to “identify appropriate strategies for the optimal management of physical assets and support the development and implementation of performance-based operating systems for both highway maintenance and transportation services”. Such a project may include the development of a statewide master/strategic plan for database aggregation, as recommended by this study.

In summary, for the safety, mobility, environmental, social, personal, and business benefits to be realized from ITS, Michigan has a distinct leadership role and opportunity. By continuing to commit to this leadership role in national efforts to develop and deploy connected vehicle and ITS technology, Michigan will have a significant role in shaping the future of transportation infrastructure management and ITS information systems development. Through its leadership and implementation expertise, MDOT will be well positioned to reap the benefits of becoming the recognized home of this technology. These benefits could include early deployment and associated lives saved, as well as thousands of jobs within the automotive and transportation industries, and reduced capital costs for operations and ongoing maintenance of the transportation network.

Acknowledgments

This document is a product of the Center for Automotive Research under a state planning and research grant administered by the Michigan Department of Transportation.

References


Management Procedures for Data Collected via Intelligent Transportation Systems

Highway Research Program) report.


