NCHRP 08-36, Task 129

Scoping Study to Establish Standards and Guidance for Data for Transportation Planning and Traffic Operation Purposes

Requested by:

American Association of State Highway and Transportation Officials (AASHTO)
Standing Committee on Planning

Prepared by:

Cambridge Systematics, Inc.

Texas Transportation Institute

October 2016

The information contained in this report was prepared as part of NCHRP Project 08-36, Task 129, National Cooperative Highway Research Program (NCHRP).

AASHTO
THE VOICE OF TRANSPORTATION

ACKNOWLEDGMENT

This study was conducted for the AASHTO Standing Committee on Planning, with funding provided through the National Cooperative Highway Research Program (NCHRP) Project 08-36, Research for the AASHTO Standing Committee on Planning. The NCHRP is supported by annual voluntary contributions from the state Departments of Transportation. Project 08-36 is intended to fund quick response studies on behalf of the Standing Committee on Planning. The report was prepared by Anita Vandervalk (Cambridge Systematics, Inc. (CS)), Rich Margiotta (CS), Michalis Xyntarakis (CS) and Shawn Turner (Texas Transportation Institute). The project was managed by Lawrence D. Goldstein, NCHRP Senior Program Officer.

DISCLAIMER

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Executive Summary

Planning and operating transportation systems involve the exchange of large volumes of data. The lack of common data formats has been a limiting factor for transportation agencies and all practitioners involved in data analysis and reporting. Well-designed data standards can be a viable solution to this problem since they improve the efficiency of data-driven processes and can support innovation.

NCHRP 8-36 Task 129 examined the feasibility of developing standards for transportation planning and traffic operations. This was accomplished through three tasks as follows: literature review, examination of feasibility of data standards and development of a Research Statement.

The report revealed that it is difficult to predict standard adoption. There are many well-designed and technically superior standards that have failed and become marketing case studies. Based on the research, a business case and clear incentives for a critical mass of supportive stakeholders is required for market adoption.

Standards are most successful if they:

- Have a clear business purpose;
- Are clear in application, specificity and versioning;
- Are developed with broad outreach and buy-in;
- Are well defined and simple;
- Are open standards;
- Are forward looking; and
- Involve a national or worldwide community.

There are also many challenges associated with standard development. These include:

- Reluctant data vendors;
- Dynamic data content - Standard may be outdated soon after it is released;
- Complexity of data to be standardized;
- Standardization process takes too long to complete;
- Standardization process does not take into account a critical mass of would be users or decision makers;
- There are significant disincentives or conflicts of interest;
- Limited outreach - Agencies may be unaware of the benefits of adoption; and
- Inadequate resources to overcome the barrier of entry.

The report concluded that transportation planning and operations standards are feasible and desired. Five specific areas or bundles of standards were identified to be ripe for standardization. They are travel time, demand, incident and work zones, network and transit. This study included the development of a Research Problem Statement. The objective of the project resulting from the Problem Statement will be to prioritize transportation planning and traffic operations standard areas and develop standards and/or guidance to be used and adopted by the transportation community.

Introduction

Planning and operating transportation systems involve the exchange of large volumes of data. The lack of common data formats to facilitate the exchange of data across different business platforms has been a limiting factor for transportation agencies, vendors, contractors, and other groups. Well-designed Data standards can be a viable solution to this problem since they improve the efficiency of data-driven processes and introduce innovation.

The U.S. has a long history of standardization that started in 1901 with the establishment of the National Bureau of Standards. Since then, thousands of standards have been published and have been widely adopted. Initially, standards were only related to the design of infrastructure and the flow of products from the producers to factories and the consumers. However, as more and more of the society depends on information flows and Big Data are generated by a variety of sources, standardizing data flows becomes increasingly important. Since the early standardization efforts, US standardization strategies call for voluntary consensus rather than Government intervention. A convening role for the government is reserved only when there is justification that Government intervention is warranted.

Although there have been many successful voluntary consensus standards for Intelligent Transportation Systems, there have been few attempts to standardize data sources used in Transportation Operations or Planning for performance measurement, modeling, or data analytics. In this document, the study team presents a number of transportation data standards and examines some of the reasons behind their success or failure. FHWA's Highway Performance Monitoring System (HPMS) is a prime example of standard procedures that are federally mandated and whose success depended on the magnitude of the issuing authority. Publishing a standard, however, by no means ensures success. TransXML, for instance, is an NCHRP-funded and well-designed transportation data standard that has seen limited success since its introduction in 2006. In contrast, without any elaborate design, the General Transit Feed Specification has become the de facto standard for storing transit networks and schedule information.

It started not as a standard but as a simple data exchange format between TriMet and Google and, in just a few years, the General Transit Feed Specification became so popular that it eliminated any other competing formats. The Open Matrix Format is an example of an independent effort to develop a format to exchange rectangular scientific data such as trip tables. It was introduced in 2014 by a group of dedicated transportation professionals and it is gradually gaining acceptance, even though it has no formal Government involvement. As a final example, the various standards that have been developed but failed to be universally popular for network data are examined. Network data are increasingly important since they carry information both about infrastructure location and characteristics as well as human activity.

It is hard to predict standard adoption. There are many well-designed and technically superior standards that that have failed and become marketing case studies. A business case with clear incentives for a critical mass of stakeholders is required for market adoption. A standard not only needs to solve a current problem efficiently but also needs to make sure that the solution can overcome the barrier of adoption by providing incentives to stakeholders to be active proponents. In this document, the project team analyzed a number of factors behind the success or failure of the standards mentioned in the previous paragraph to

support future decisions. Rules for identifying data sources for standardization are presented as well as the different roles Government can play in the standardization process.

Internet of Things refers to the network of electronic objects such as smartphones or GPS devices with embedded sensors that surround us daily. Mobility information from these sensors forms Big Data streams which, after filtering and processing, can become extremely valuable data for transportation practitioners. Recent examples of Big Data that have reached the transportation practice are the National Performance Management Research Dataset and cellphone derived mobility tables. Despite their potential, Big Data are not always easy to integrate and analyze by transportation practitioners. Recent Big Data standardization efforts by the National Institute of Standards and Technology and the International Standards Organization can lay the groundwork for easier data integration.

The purpose of this study was to examine the feasibility of the development of standards for transportation planning and operations.

This report is organized by major tasks as follows:

Chapter 1 – Literature Review
Chapter 2 – Feasibility of Data Standards
Chapter 3 – Research Problem Statement Development
Chapter 4 – Proposed Research Statement

CHAPTER 1

Literature Review

Data Standards

Standardization in the US started in the early twentieth century after considerable debate over the government's role in the economy. The National Bureau of Standards (NBS) was established in 1901 and renamed National Institute of Standards and Technology (NIST) in 1988. Technological innovations such as electricity and the increased mobility of the American population highlighted the importance of standards. One story from 1904, three years after the establishment of NBS is revealing:

An accidental fire in a dry goods store in downtown Baltimore left the city burning for 30 hours, reducing 1,500 buildings on 70 blocks to rubble, killing five people, and leaving thousands unemployed. The Fire chief arrived only 11 minutes after the incident was reported and immediately issued urgent calls for help to all available firefighters in the region. Before the fire had been put away 30 hours later more than one thousand firefighters had arrived from Washington, DC, Philadelphia, Wilmington, Frederick, Westminster, York, Altoona, and even New York. However, most of them were unable to connect the fire hoses they brought to the city's fire hydrants even though there was enough water to flood the entire business district according to the fire chief. Shortly after the disaster, an investigation by the National Bureau of Standards revealed more than 600 sizes and variations in fire-hose couplings across the country. In 1905, the National Fire Protection Association established a national standard diameter, known as the Baltimore standard that remains the national standard to this day. However, as late as 1964 incompatible types of hydrants could be found in adjacent counties.

According to Aneesh Chopra, who served as the first Chief Technology Officer of the United States, the fire may not have gone for so long, or done so much damage, if compatibility had been a priority in the fire equipment industry. Instead, the market incentives led manufacturers to design entirely proprietary systems, including different couplings, for each vendor. After all, a city that purchased a particular system would be entirely dependent on that system's manufacturer for any improvements or upgrades.

When standardization first started in the early 20th century, it focused on nuts, bolts, and the flow of products from farms to factories and consumers. If one is to replace the words *flow of products* with *flow of data* in the following quote from Paul Agnew, the longtime Secretary of the American Standards Association, it is evident that little has changed:

In the flow of products from farm, forest, mine, and sea through processing and fabricating plants, and though wholesale and retail markets to the ultimate consumer, most difficulties are met at the transition points – points at which the product passes from department to department within a company, or is sold by one company to another or to an individual. The main function of standards is to facilitate the flow of products through these transition points. Standards are thus both facilitators and integrators. In smoothing out points of difficulty, or bottlenecks" they provide the evolutionary adjustments, which are necessary for industry to keep pace with technical advances.

Standards and the Standardization Process

According to International Standards Organization (ISO), "A standard is a document that provides requirements, specifications, guidelines or characteristics that can be used consistently to ensure that materials, products, processes and services are fit for their purpose". According to Circular A-119 issued in 1998 by the White House Office of Management and Budget (OMB) a standard is:

(1) Common and repeated use of rules, conditions, guidelines or characteristics for products or related processes and production methods, and related management systems practices.

(2) The definition of terms; classification of components; delineation of procedures; specification of dimensions, materials, performance, designs, or operations; measurement of quality and quantity in describing materials, processes, products, systems, services, or practices; test methods and sampling procedures; or descriptions of fit and measurements of size or strength.

Standards have a measurable positive impact on economic growth and technology transfer. However, according to the OMB, "your agency should also recognize that use of standards, if improperly conducted, can suppress free and fair competition, impede innovation and technical progress; exclude safer or less expensive products; or otherwise adversely affect trade, commerce, health, or safety."

Some of the most important objectives of standardization are the establishment of compatibility and interoperability and the removal of technical barriers though harmonization. According to World Wide Web Consortium (W3C), Standards can be classified in the following categories:

Formal or de jure standards have a legal basis and can be made mandatory. They are normative documents from formal standards bodies and have passed through a full and open consensus process. They are implemented on a national level and there is strong mandate to apply them. Considerable time (up to 4 years) may be needed for completing the full approval process.

Technical or industry specifications are based on consensus among members of standards bodies, consortia or trade organizations and do not have a formal character or legal basis. They are recommendations and require less time to produce (1-3 years). When widely accepted and used in practice by relevant market players; they can become de facto standards.

Guidelines, conformance, or reference implementations. They have an informative character and are usually produced in a relatively short timeframe since they do not require any consensus.

According to World Wide Web Consortium (W3C), although not all standardization processes follow exactly the same steps in exactly the same order, a number of commonalities can be identified that characterize a typical standardization process:

First, a market need for a new standard or standardization activity has to be identified and recognized among a sufficient number of members of a standards organization.

Subsequently, a set of requirements has to be drafted, underlying the actual technical specification work (usually referred to as commercial, user or functional requirements).

Based on consensus reached among the organization's members on these requirements, a specification is drafted by a group of technical experts.

Once the draft specification is finalized, a formal approval process is conducted; this may be limited to the organization and its members, but may also invite a wider audience, e.g. to broaden the support for, or impact of, the future standard.

After its approval, arrangements are made for testing or (self-) certification by the industry, in order to guarantee interoperability between different implementations; this may also encompass developing reference implementations or implementation guidelines.

Finally, a maintenance or periodic review process is embedded in the organizations procedures to ensure the standard will remain in sync with market requirements.

Government's Role in Standardization

The Federal Government can play various roles related to standard development, adoption, or usage. In order of increasing involvement, these roles can be the following:

- User;
- Participant in standard setting bodies;
- Facilitator by providing advice and sources of funding;
- Advocate;
- Technical advisor/leader;
- Convener: when the government takes the leadership and actively engages with the industry and though a voluntary consensus activity specifies a new standard; and
- Specifier: when the Government mandates or strongly encourages Government-unique standards to participating parties.

Government role usually falls in the first five categories and does not involve a convening role. In fact, according to a memorandum by the Office of Management and Budget published in 2012:

Most standards used in the U.S. are created with little or no government involvement. This approach – reliance on private sector leadership, supplemented by Federal Government contributions to discrete standardization processes - remains the primary strategy for government engagement in standards development.

However, in the same memorandum it is stated that in limited policy areas, active engagement or a convening role by the Federal Government may be needed to accelerate standards development. In the latter case, the memorandum invites government action to be open and transparent to every stakeholder. Furthermore, it urges government at the outset of the standardization process to:

- Analyze the reasons that have led to a standard's gap, using both quantitative and qualitative methods;
- Define the standardization goals as precisely as possible; and
- Identify the challenges involved and the reasons why government inaction would be the least effective path forward.

Standards shape and are shaped by competition. When the government adopts the convener or specifier role described above, the outcome should be timely, cost-efficient, and competition-sustaining standards.

Principles of Data Standardization

Developing a data standard should start with a business case that lists the incentives and disincentives of adoption, the level of effort to develop a standard specification, and the level of coordination required to gain adoption by a critical mass of stakeholders. Therefore, designing data interoperability must start with a practical understanding of what data needs to be shared, and how sharing of the data will benefit the transportation community. Stakeholders with hands-on experience of the challenges involved and the inefficiencies in the existing data processes can provide valuable advice on possible standardization areas.

A data stream that is a good candidate for standardization has the potential to save time and money for a large number of stakeholders, or facilitate improved access to information by:

- Allowing information produced in one process to be used in other(s);
- Allowing agencies to analyze the same set of data in several different software tools or to share input data across different agencies;
- Eliminating or reducing the need for agencies to build custom procedures for the same common problem/application domain; and
- Enabling reuse of the same information for multiple purposes.

The benefit of a proposed standard increases as:

- The number of different users increases;
- The number of different uses increases; and
- The efficiency of dealing with complex information increases.

Poor candidates for standardization include the following:

- There is wide variation in data content that is specific to each application. A common denominator may still exist but exceptions may be so frequent that they endanger the universal applicability of the standard.
- There are disincentives, or even conflicts of interest, for the producers of the data to collaborate on a common specification.
- Information is shared across a small number of agencies using data of limited usage. Even though a standard can be defined, it will have limited use and benefit.
- The structure of data content is highly dynamic. By the time the standard is developed it may already be outdated. New datasets from cellphones or other sensors may fall into this category.
- The data is so simple that there is little benefit to standardization.
- Using the standard requires resources and knowledge that is not widely available by the majority of the stakeholders.

Transportation-Related Data Standards Organizations

Transportation data is fundamentally geographic in nature. Although there has been considerable work on the development of data standards focused on the geospatial representation of transportation information, very often it is spatial data that still represent the biggest obstacles for data integration. In this section, standards organizations focusing on transportation-related data standards are being presented.

The Object Management Group (OMG) is an international, open -membership, not-for-profit technology standards consortium. The standards published by OMG support interoperation between enterprise applications and data companies such as Oracle, IBM, and Unisys. OMG has published many standards, including the following:

The **Unified Modeling Language (UML)** is a standard on how to visualize the design of a system or a business process. UML is being used extensively in software engineering and in the definitions of standards published by ISO or other bodies. For example, UML was the language used to define TransXML, a transportation standard reviewed later in this document.

A variety of meta-data standards, such as the **Common Warehouse Meta-data (CWM)**, and business modeling standards, such as Business Process Model and Notation (BPMN). Both standards can be part of the data management or governance process in transportation.

Global Spatial Data Infrastructure (GSDI) is an organization promoting international cooperation and collaboration in support of local, national and international spatial data infrastructure developments. Spatial data infrastructure is defined as the framework of fundamental spatial datasets, metadata, and interoperability standards that enable integration and compatibility. The Open Geospatial Consortium is a not-for-profit, international organization that prepares consensus-based geospatial standards. Many of its standards are widely used and several transportation standards, such as HPMS, build upon them.

The **Federal Geographic Data Committee (FGDC)** is an interagency committee that promotes the coordinated development, use, sharing, and dissemination of geospatial data on a national basis. The FGDC develops geospatial data standards in consultation and cooperation with State, local, and tribal governments, as well as the private sector and the academic community. FGDC has developed many standards related to the transportation field, such as the Geographic Information Framework Data Standard, which addresses the representation of transportation networks. For example, Figure 1 shows the standardized representation of the physical roadway based on the FGDC-STD-014.7c-2008 standard. Other standards published by FGDC include:

- Content Standard for Digital Geospatial Metadata;
- Spatial Data Transfer Standard; and
- Geospatial Positioning Accuracy Standards.

More than just publishing standards, FGDC has produced several directives related to the standards review process, review criteria, and maintenance of standards than can provide valuable guidance for future transportation standardization processes.

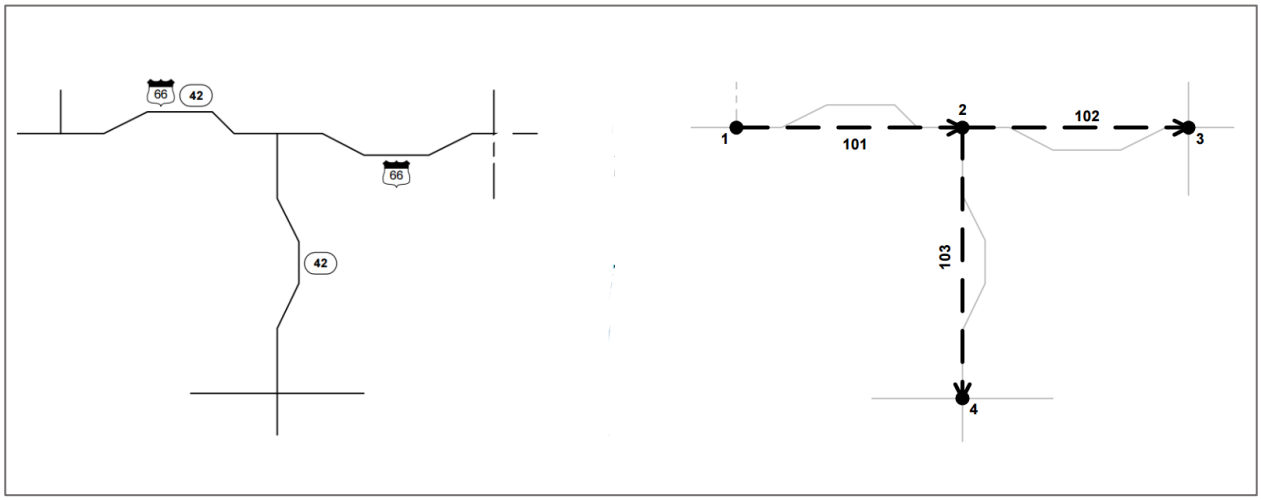


Figure 1: Physical Road Infrastructure (Left) and Standardized Network Representation (Right)

Infrastructure for Spatial Information in the European Community (INSPIRE) is a directive that aims to create a common spatial data infrastructure in the European Union (EU). The directive aims to enable the efficient sharing of spatial information among public sector organizations. One of the areas of focus is Transport Networks and Coordinate Reference Systems. For example, in terms of transportation and mobility, a recent report documents the general methodology enabling a mobile network operator to process cellphone records to estimate population density for public policy projects. The population estimates are made over a geographical grid that ensures interoperability with other data sources. The general principles followed by INSPIRE follow (based on their website):

- Data should be collected only once and kept where it can be maintained most effectively.
- It should be possible to combine seamless spatial information from different sources across Europe and share it with many users and applications.
- It should be possible for information collected at one level/scale to be shared with all levels/scales; detailed for thorough investigations, general for strategic purposes.
- Geographic information needed for good governance at all levels should be readily and transparently available.
- It should be easy to find what geographic information is available, how it can be used to meet a particular need, and under which conditions it can be acquired and used.

In order to ensure that the spatial data infrastructures of the member states are compatible, the INSPIRE Directive requires that additional legislation, or common implementing rules, are adopted for a number of specific areas, such as metadata. The types of data standardized by INSPIRE or the rules adopted to ensure uniformity and interoperability can serve as a source of ideas, a point of comparison, or reference to this project.

The **Organization for the Advancement of Structured Information Standards** (OASIS) is a nonprofit consortium that drives the development, convergence and adoption of open standards for the

global information society. OASIS has published standards for the Internet of Things, cloud computing, and recently on big data. ISO has working groups that have published standards in the areas listed in Table 1.

Table 1. ISO Workings Groups and Standards Related to Transportation

Working Group	Standardization Area
WG3	Public transport
WG4	Traffic and traveler Information
WG5	Traffic control
WG7/8	Geographic road data base: road traffic data
WG11	Route guidance and navigation systems
TC241	Road traffic safety management systems
ISO/IEC JTC 1/SC 32	Data management and interchange, including database languages, metadata management, model specification

Other standards bodies include:

- American National Standards Institute (ANSI)
- National Electronic Manufacturers Association (NEMA)
- American Society for Testing and Materials (ASTM)
- Institute of Transportation Engineers (ITE)
- Institute of Electrical and Electronics Engineers (IEEE)
- Society of Automotive Engineers (SAE)
- American Association of State Highway and Transportation Officials (AASHTO)

Standardization Efforts in Planning and Operations

If Intelligent Transportation Systems (ITS) are excluded, there have not been many standardization efforts in transportation planning and operations. Even though a large number of standards have been developed for ITS, many of them are related to hardware rather than data analytics and performance measurement. The open, voluntary, and consensus-based process by which ITS standards are being developed has demonstrated several successes and requires further investigation. In this chapter, standardization efforts on data sources that can be used for performance measurement and analytics are researched. The reviewed standards cover the entire spectrum, from those that were designed and mandated by the government to those that were developed on a voluntary basis by interested stakeholders, without any government involvement.

In this chapter, data for motorized traffic are presented. Bicycle and pedestrian movements is an emerging source of mobility data that is not accounted for. Currently, such datasets are expressed as a collection of trajectories in simple tabular formats containing spatiotemporal information.

Highway Performance Monitoring System (HPMS)

The HPMS is a national level highway information system that includes data on the extent, condition, performance, use, and operating characteristics of the nation's highways. The HPMS started in 1978 as a continuing database and has been modified many times since. HPMS has several significant uses and applications, the most important of which are:

- Serving as the basis for apportionment of Federal-aid Highway Program funds to the States.
- Being considered as the primary input mechanism for the biennial Condition and Performance Reports to Congress. In turn, these reports are used for the development and evaluation of FHWA's legislative, program and budget planning.
- Serving as a primary resource for assessing highway performance under FHWA's strategic planning process. HPMS is also the primary source for measuring safety (in terms of fatalities and injuries), pavement roughness, and changes in congestion levels.
- Being utilized by States, MPOs, local agencies, and other stakeholders in assessing highway condition, system performance, air quality trends, and future investment requirements. Many air quality analyses rely on HPMS data.
- Providing information and statistics for FHWA's annual Highway Statistics media publications.

There are three sources of data used in the HPMS database. These include:

- Data that are to be developed and supplied by the States;
- Data that are obtained by FHWA from other non-State sources; and
- Data that are generated or calculated by FHWA.

Most of the raw data are provided by the states to FHWA, with the states being responsible for data quality and integrity. Some data elements must be reported to their full extent (i.e. system-wide), while others are reported through a random selection of roadway sections. The data items that need to be reported include:

- Roadway inventory information;
- Route designation;
- Traffic information;
- Roadway geometry characteristics; and

- Pavement data.

Personal interviews with HPMS administrators have highlighted the following elements of the program:

- The importance of a well thought-out design from the beginning, as opposed one involving later modifications and updates;
- The importance of establishing a mechanism that provides constant guidance and training to all available parties; and
- The diverging views and needs that different states have - without FHWA's involvement, states will most likely follow different and incompatible approaches.

Many of the transportation data have a geospatial component that is defined over a network. Obvious examples are traffic, travel time, and incident information. Currently, different entities, such as city DOTs, MPOs, county DOTs, private companies, and federal agencies, have different reference networks for the data that they collect or use that are incompatible. As a result, data integration is a laborious and time-consuming process. The All Road Network Linear Referenced Data (ARNOLD) is an effort by FHWA to develop a master network that contains all the roadways and an advanced linear referencing system. It is envisioned that networks deriving from the ARNOLD network and maintaining the original linear referencing system will be compatible with each other.

TransXML

The National Cooperative Highway Research Program (NCHRP) developed as part of the Project 20-64 XML Schemas for Exchange of Transportation Data, a set of eXtensible Markup Language (XML) schemas for transportation applications. XML is a text format that is both human and machine-readable and is very similar to the HTML format used by websites. XML's major advantage over tabular formats is its flexibility to support nesting of fields to represent hierarchical relationships between data elements.

The new standard, TransXML, whose initial phase was developed in 2006, was targeted at providing broadly acceptable public domain XML schemas for exchange of many kinds of transportation data. The vision behind TransXML was to develop an extensible format that can support the exchange of transportation data from more than one domain. The broad domains that were initially standardized are the following:

- Survey/Roadway Design;
- Area Features Schema;
- Geometric Roadway Design;
- Design Project;
- Transportation Construction/ Materials;
- Bid Package;
- Construction Progress;
- Materials Sampling and Testing
- Project Construction Status;
- Highway Bridge Structures;
- Bridge Design and Analysis;
- Transportation Safety;
- Crash Report; and
- Highway Information Safety Analysis.

For storing network data and for all the initial domains, a Linear Referencing (LR) schema was also developed based on ISO 19133.

Figure 2 shows the initial four domains and potential focus areas for TransXML. Building on the flexibility of the XML format and its popularity, TransXML is intended to be an extensible specification that will incorporate more schemas to address the entire spectrum of the transportation industry.

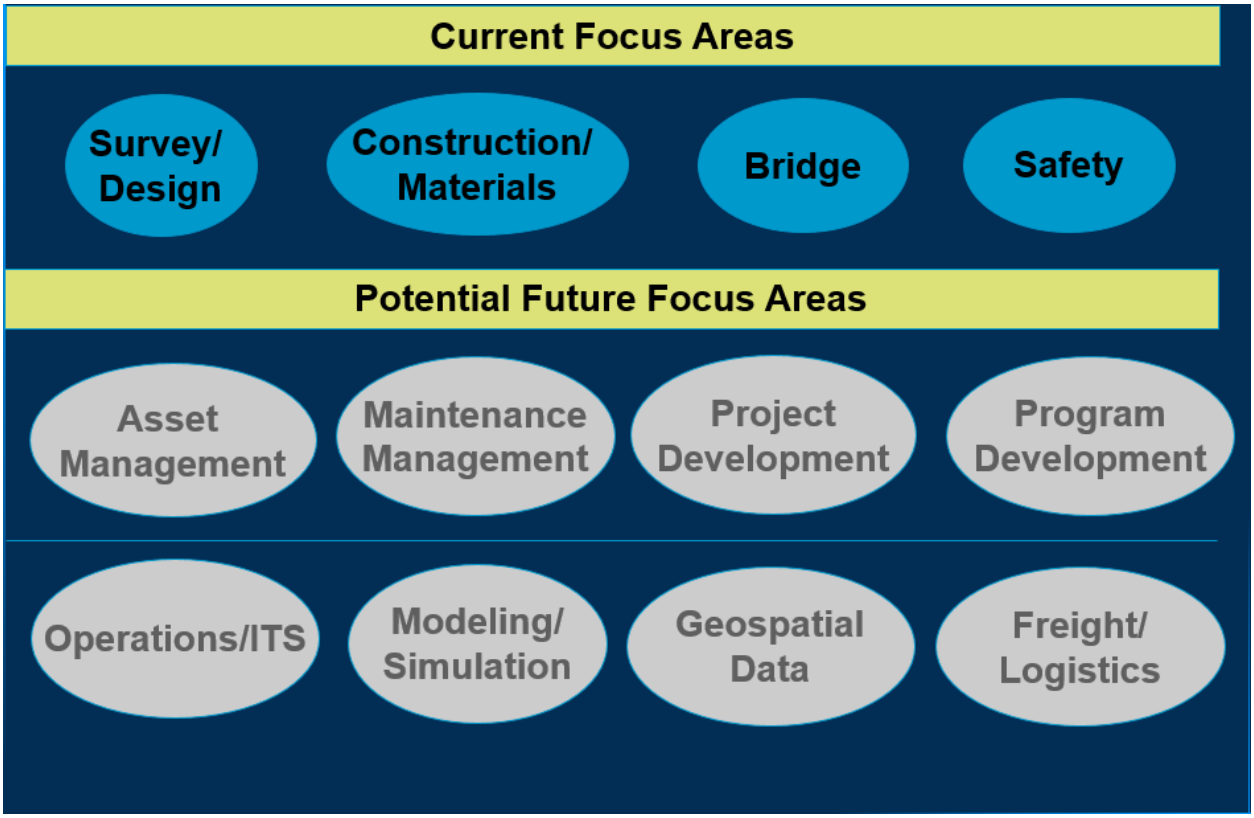


Figure 2: Initial (2008) and Potential Focus Business Areas of Implementing TransXML

Eight years after the publication of the standard, a workshop entitled The Future of TransXML was conducted that focused on past adoption and potential future development and implementation of TransXML. The workshop had many objectives, including:

- Gaining a better understanding of the activities critical to furthering development, adoption, and maintenance of TransXML;
- Exploring partnership opportunities and alternative approaches; and
- Identifying a possible TransXML stewardship framework and process for developing, maintaining, and updating TransXML schema.

In the workshop proceedings, it is mentioned "a survey conducted in 2011 as part of a second NCHRP project found that many state departments of transportation (DOTs) and other agencies were unaware of TransXML and identified only a few examples of TransXML schemas in use."

Furthermore, survey respondents were asked to identify major barriers that limited the adoption of common data formats, with the following results:

- Lack of awareness (70%);
- Too costly to retrofit (56%); and
- Incompatibility with existing formats in use (46%).

General Transit Feed Specification (GTFS)

GTFS is one of the most impactful and popular transportation standards. It started as a simple data exchange format between TriMet and Google in 2005. Since then, it has become the de facto standard for all schedule-based transit information that is delivered online, or though smartphone apps, to millions of travelers that ask for transit information between an origin and destination. GTFS has also entered into Transportation Planning and Operations, with the Federal Transit Administration (FTA) starting the STOPS project that uses detailed descriptions of an area's current transit services from GTFS. Popular transportation planning software, such as TransCad, have built-in support for it, while a growing number of public agencies, such as the New York Metropolitan Transportation Council (NYMTC) and the Baltimore Metropolitan Council (BMC), are increasingly adopting it as the most error-free and up-to-date information about transit level of service.

GTFS solved a clear business problem: the fragmentation and incompatibility of transit data formats among transit agencies. In 2005, when GTFS started, each transit agency had its own transit format and relied on in-house or proprietary tools. In addition, some agencies had started building their own transit or multimodal web-based route planners for the general traveler using different technologies and interfaces. Overall, the fragmentation of the field made finding transit directions in unfamiliar cities a challenge, as transit agencies had little incentive to collaborate on a common data standard. Furthermore, some of them saw transit schedule data as their asset and had already reached agreements with private companies to provide route-planning services.

GTFS was designed by transportation engineers as the simplest solution to store and exchange transit schedule information between TriMet and Google. Google had agreed to introduce a transit route planner service in Portland Oregon as a limited pilot service. Google's initiative was not a high-level management decision but a hobby project supported by a number of dedicated employees. The standard is a list of tables, each of which stores a collection of similar elements such as stops, trips, route shapes, etc.

Figure 3 shows all the tables that are part of the standard. Arrows between the tables correspond to foreign key value relationships, similar to how relational databases are defined.

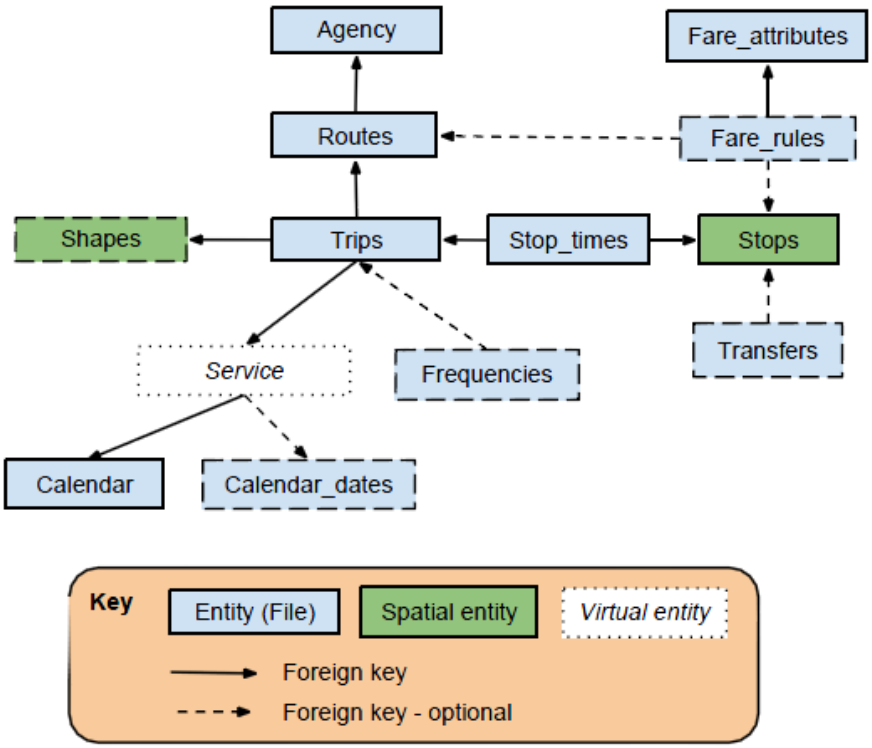


Figure 3: GTFS Specification Overview

According to Bibiana McHugh of TriMet who wrote an article about GTFS in the book *Beyond Transparency*:

> We chose to keep the files in the text-based Comma Separated Values (CSV) format. We wanted it to be as simple as possible so that agencies could easily edit the data, using any editor. This approach received substantial criticism and was even called technically old fashioned and brittle but it was important to us to keep the barrier of participation low so that even smaller less-resourced agencies could join in.

According to Google team members, CSV is also the simplest and more appropriate format to use.

According to Bibiana McHugh, when the pilot route planning service for Portland first opened in 2005 it received an overwhelmingly positive response. By 2011, the vast majority of the transit agencies in the U.S. had adopted GTFS. A comment by Chris Harrelson, the Google Engineer behind the initial project, sheds some light on the factors behind GTFS success:

> It is perhaps easy to jump to the conclusion that Google is the hero in this story, in the typical role of the innovator who overcomes the inefficiencies of the past, but this is not true in this case. This is a success story about a new model of cooperation in order to solve a problem that cannot be addressed directly with either market forces or a classic government solution. Everyone had an equally important role to play, and without TriMet and other government advocates, the story would not be possible.

McHugh cites the following ingredients that made the GTFS initiative and eventual standard successful:

- A collaborative team that started small and designed for a very specific use;
- Releasing the transit data specification in an open standard;

- The simplicity of the specification and format;
- A tangible business incentive for the transit agencies to interoperate with web and mobile routing apps and for private partners to participate; and
- The contributions and involvement from a worldwide community of users.

GTFS has been expanded to include real-time information on transit vehicle location. The standard is called GTFS-realtime and is being used widely to provide live transit updates. A similar standard that is used extensively in Europe is called Service Interface for Real Time Information (SIRI). The Metropolitan Transit Administration in New York and other agencies are using it to disseminate real-time transit information. Other standards include the Transit Communications Interface Protocol (TCIP) sponsored by U.S. DOT and the privately developed NextBus.

Open Matrix FORMAT (OMX)

This is a relatively new specification initiated by a group of transportation practitioners without any government support. According to the OMX website:

> Open matrix (or OMX) is an open file specification for storing a series of matrices with associated attributes. The OMX project is a collaborative effort to define a standard and includes a specification, APIs in different programming languages for reading/writing OMX files, and commercial travel modeling software plug-ins for reading/writing OMX matrices.

Furthermore, the group believes that an open, common format for matrix data across models and software packages will make model development and application easier. The group also believes that the ability to readily share data for a key data structure in a standard format will help spur industry research and innovation, similar to what the ESRI Shapefile did for GIS.

OMX team members made a presentation about the format at 2014 TRB Applications Conference and have held informal meetings among interested parties at recent TRB annual conferences. Based on their website, TransCad has already built OMX support, and other planning software providers, such as Cube and VISUM are planning to add similar functionality soon. The OMX team has developed interfaces to many of the most popular programming languages, including C#, Java, python, Ruby, and R.

Although it is not easy to predict the industry's adoption of OMX, it clearly solves the problem of efficiently storing trip tables and other rectangular-shaped data. Trip tables are ubiquitous in transportation planning since many performance measures, such as VMT (vehicle miles traveled), are based on them. Even though such tables can be processed completely inside transportation software, more and more often analysts turn to R, an open-source statistical package, or other data analytics packages to take advantage of state-of-the-art statistical and visualization functionality. OMX is not entirely a new format; it is based on the successful Hierarchical Data Format (HDF) that was developed to store scientific data. OMX customizes HDF for transportation purposes.

Transportation Network Data

Network data are ubiquitous in transportation because they describe the roadway system. Furthermore, some of the most challenging problems of data integration, geo-referencing, and translation between networks inhibit the adoption of new data such NPMRDS. Even though multiple standards for network representation exist, there has not been one that has gained widespread adoption. Each software vendor has their own definition and each network type has its own format, making network association a very time-consuming and error-prone process. Table 2 presents some network formats used in transportation.

Table 2. Transportation Network Formats and Standards

Network Format	Description
OpenStreetMap	Crowd-sourced detailed network information for the entire world
ARNOLD	Federally supported effort to combine all public roads into a single, publicly available, authoritative spatial dataset
UNETRANS	Unified NETwork for Transportation is a geodatabase network model that stores assets, incidents, activities, routes and location referencing, mobile objects, and relationships
ISO 19148	Linear Referencing Standard
Traffic Message Chanel (TMC)	Network partition used to define NPMRDS travel time information. Travel time information providers produce TMC networks that may not be fully compatible
INSPIRE D2.8.I.7	Highly detailed European data standard on Transportation Networks
Geographic Data Format (GDF)	ISO International standard that is used to model, describe, and transfer road networks and other geographic data
OpenLR	Open dynamic location referencing on transportation networks supported by TomTom
TransCad, EMME, VISUM, AIMSUN, VISSIM, Dynameq, TransModeler, Cube	All transportation software packages have their own definition for transportation networks. Import/Export features exist but they cover only a small subset of available formats

In general, network data in transportation abstract the physical roadway data into nodes and links. Nodes represent intersections, junctions, or any roadway element when a turn can happen. Links represent physical roadway links with one or more lanes on which cars or trains can travel. An overview of the network data for transportation planning can be found in the Travel Demand Forecasting Manual.

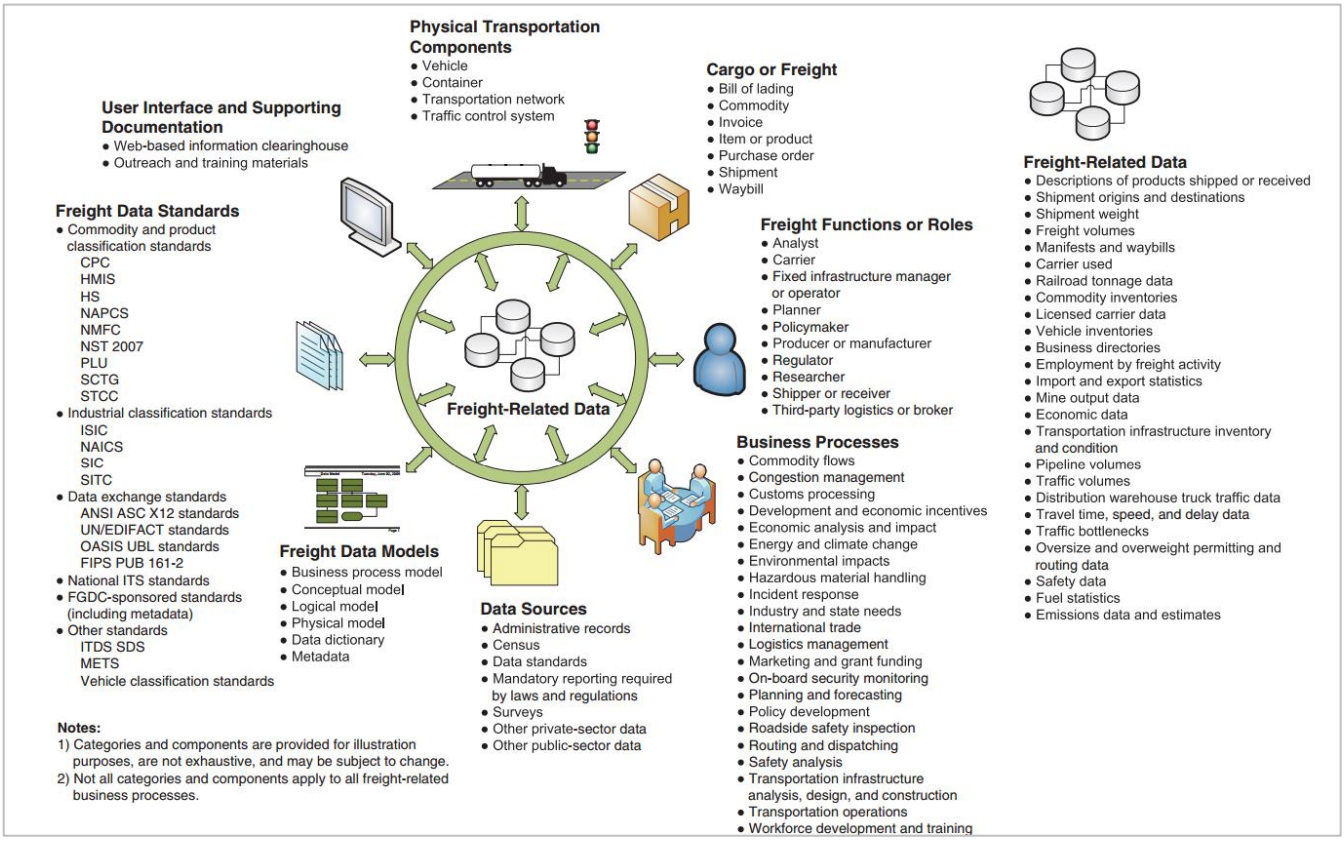
A significant amount of effort is often required to build, maintain, and update a network. The initial starting point can be network data from the FHWA National Highway Planning Network, Highway Performance Monitoring System, Freight Analysis Framework Version 3 (FAF3) Highway Network, National Transportation Atlas Database, U.S. Census TIGER data, OpenStreetMap, and various state transportation networks. Significant manual effort is often embedded in the working network - effort that often impedes the adoption of a newer network format unless an automated way is found to merge the two networks. Manual edits and updates to planning or operational transportation networks are very resource intensive.

Crowd-sourced network data from OpenStreetMap contain all roadway infrastructure, in addition to points of interest or even 3D models of metropolitan areas. OpenStreetMap is often used in navigation apps targeted to millions of users; Waze, a Google company, is an example. While not flawless, it is generally accepted that collaborative, or crowd-sourced map data in OpenStreetMap have a level of detail and accuracy that surpasses the network accuracy that can be achieved by a few number of network modelers, even if they work full time. On the other hand, ARNOLD is a state-of-the-art effort to combine all public roads into a single, publicly available, authoritative spatial dataset. FHWA has been exploring links between OpenStreetMap and ARNOLD to the benefit of both projects.

Freight Standards

A large number of standards exist to describe freight movements. However, in many areas, a standardized terminology and definitions are needed. NCFRP Report 9 provides guidance for developing a nationwide freight transportation data architecture. Specifically, it provides an extensive list of freight

data standards as well as differences in terminology, data item definitions and implementations that impede information flow. The following schematic from the same report proposes a national freight data architecture framework and components.



Source: NCFRP Report 9, Guidance for Developing a Freight Transportation Data Architecture

Figure 4: National Freight Data Architecture Framework and Components

Summary

Table 3 summarizes the standardization efforts discussed in this section. The standardization efforts are compared across different dimensions, based on insights about principles of data standardization and on any lessons learned from each data standard review.

Table 3. Elements of Reviewed Data Standards

	HPMS	TransXML	GTFS	OMX	Competing Network Formats
Business Case	Common framework for performance measurement	One format all transportation data	Provide transit routing info	Storing, exchanging, querying matrix data is hard	One format for all network data
Alternative case or competing formats	State specific procedures and formats	Different formats for different domains that are not linked	Fragmentation by agency and application domain	A number of csv or propriatory tables	Network formats tailored to each application domain
Number of intended users	States and FHWA	Transportation Industry	All Transit Agencies	Modelers	Transportation Industry
Incentive to comply or adopt	Mandatory	Medium to low, not clear to users	Very high to allow transit service to be integrated in apps	Medium to high for data analysts to transfer data	Low unless another application in the same domain uses the format
Simplicity to comply	Low	Medium	High	Medium	Low
Compatibility with existing software	Medium	Low	High	Medium	Low
Variation in data content	Medium	Very high	Low	Low	High
Self sustained community support	No	No	Yes	Yes	For some network formats only

Additional standards not reviewed in this document include the Traffic Management Data Dictionary (TMDD) and the Service Interface or Real-time public transit Information (SIRI). These two standards are combined in the Data Exchange Format Specification (DXFS) to address real-time system management and dissemination of transportation information.

Big Data

Virtually every industry has a new data source at its disposal from sources traditionally outside the industry's domain. In the transportation field, examples are the National Performance Management Research Dataset (NPMRDS), connected vehicle, or cellphone data. More mobility datasets from additional sources or the Internet of Things may be available in the future. In their raw form, the velocity and volume of GPS breadcrumb or cellphone records can be accepted as Big Data across many industries. However, even aggregated mobility datasets can be challenging to transportation engineers. In this section, a brief overview of the Big Data paradigm is provided from the perspective of the transportation practitioner.

There is little consensus about what Big Data actually refers to. In different industries, the term may have different meanings. The most often cited definition is the one included in two reports by ISO and NIST on Big Data:

The heart of the Big Data paradigm is that it is too big (volume), arrives too fast (velocity), changes too fast (variability), contains too much noise (veracity), or is too diverse (variety) to be processed within a local computing structure using traditional approaches and techniques.

Please note that all the terms in the definition above are qualitative in nature. If a specific data volume (e.g. Terabytes) is mentioned in the definition, what Big Data is today may not be in the future. In the transportation field, a real-time traffic prediction application that uses probe vehicle data, roadway sensors, and accident and weather data may legitimately use the terms volume, velocity, variability from the definition. However, most off-line data analytics tasks in transportation practice that build reports or uncover trends and patterns from a variety of sources may not legitimately use the terms in the ISO or NIST definition.

A more relevant definition can be found in the TeraData Magazine as follows: "Big data exceeds the reach of commonly used hardware environments and software tools to capture, manage, and process data within a tolerable elapsed time for its user population." In the transportation field, processing one year of NPMRDS data for a big city may not be done in-memory on a single workstation with less than 32GB of RAM. Another definition was published by the McKinsey Global Institute in May 2011, as follows: "Big data refers to data sets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze." Typical database software tools are traditional relational databases such as Oracle, PostgreSQL, Microsft SQL Server, and MySQL. In 2007, Facebook's former chief data engineer, Jeff Hammerbacher, found that traditional database software such as Oracle could not easily handle more than one Terabyte of data. Frequently, transportation engineers deal with data that is less than the one Terabyte limit. For example, the entire NPMRDS dataset for the State of Florida for a single year is 200 Gigabytes. However, the spatiotemporal nature of many transportation datasets imposes an additional burden on database operations that slows execution compared to numeric or text-based tabular data.

Big data is often different from traditional data sources because a machine automatically generates it. For example, GPS or other sensors generate vast amounts of data that is cheap to store. Unlike traditional data sources, which were carefully designed to be 100 percent relevant and had well-defined variables, a lot of the information from a big data source may need to be filtered out to be useful. Furthermore, parsing and filtering the data may not be easy because the data may be semi-structured or unstructured and not user friendly. Therefore, the data preparation and cleaning step can take a lot longer.

In some respects, working with transportation Big Data will not radically alter the transportation practice, as practitioners will likely work with aggregated and filtered data and not with the raw data

streams. Private companies with extensive expertise in parsing raw GPS breadcrumb data or cellphone records produce aggregated measures that can be used directly by transportation engineers. However, the data generated may not be representative of the entire population or may not have socioeconomic information or other necessary info due to privacy reasons. It will be up to the transportation practitioner to identify the quality of the data and limitations in their use. For example, drivers with GPS devices monitoring their speed may not go over the speed limit or may alter their behavior in some way that is limiting to transportation applications. Regardless, from a data volume perspective, even the aggregated version of raw Big Data sources poses significant challenges to transportation practitioners because the majority of them do not have the necessary database and programming skills required to analyze hundreds of gigabytes or terabytes of data.

Using Big Data alone is limited, but mixing Big Data with traditional data sources holds the highest potential. For this to happen, building a common database that has many of the datasets in one place can help the practitioner intermix and enhance the data. Typically, gathering and preparing a clean dataset takes about 70 or 80 percent of an analyst's time. Understanding the properties of the new data sources and their limitation, however, may require 90% or 95% of the analysis time. Big Data standards for such as the ones being developed by NIST can lower the barrier of entry and ease integration.

Conclusion

Transportation professionals often work with software that adopted a "walled garden approach." This approach involved, and to some extent still involves, using proprietary formats to store transportation data that were developed using public funds. This approach ensured compatibility and interoperability between all components of a single vendor but left the transportation professional clueless when data had to be exchanged between different formats or software tools, unless the vendor had already built in such capability or provided documentation about proprietary file formats.

Recently, it is more and more widely recognized that one software tool, however well-designed and supported, cannot perform all modeling and data analytics tasks. For example, conducting statistical analysis in R using demand data is often preferable to attempting the same tasks with a transportation planning package. Even though software vendors continue to build import/export features and transportation professionals keep abreast of the newest data definitions, the lack of common formats of many data elements that transportation practitioners use inhibits efficiency. As a result, significant manual labor and in-depth knowledge is required to transition data from one format to the next.

In this report, a number of transportation standardization efforts that have met with various degrees of success are presented. Given the fact that the number and size of available data streams are continually increasing, it is increasingly important to identify common formats and data standards to rely on. In data analytics, it is a well-accepted rule that data manipulation or cleaning takes up to 80% of the professional's time. Therefore, to unlock the potential of data and data analytics, it is crucial that data standardization efforts become more successful and more widely adopted.

CHAPTER 2

Feasibility of Data Standards

Overview

The approach to the Feasibility Task involved the following steps:

- Develop a framework for evaluating the data that could be subject to standardization;
- Define the performance measures of interest, users and the data that support them (Task 3A of work plan) and identify categories (Task 3B);
- Document the challenges associated with developing standards for the relevant data (Task 3C); and
- Assess the ability to overcome challenges and analyze the likelihood of adoption (Task 3D).

Framework for Assessing the Need for Data Standards

The analysis framework is based on three major dimensions, as shown in Figure 5 and discussed in the following sections.

1. Performance measures and data types: In the mobility and network characteristics categories, a description of the specific type of data is provided.

2. Life cycle stage: This dimension relates to the stage in the life cycle of a data type, recognizing that standards may be necessary at different stages.

3. Evaluation criteria: The third dimension is the criteria on which we assess the need for and likelihood of adoption of a standard.

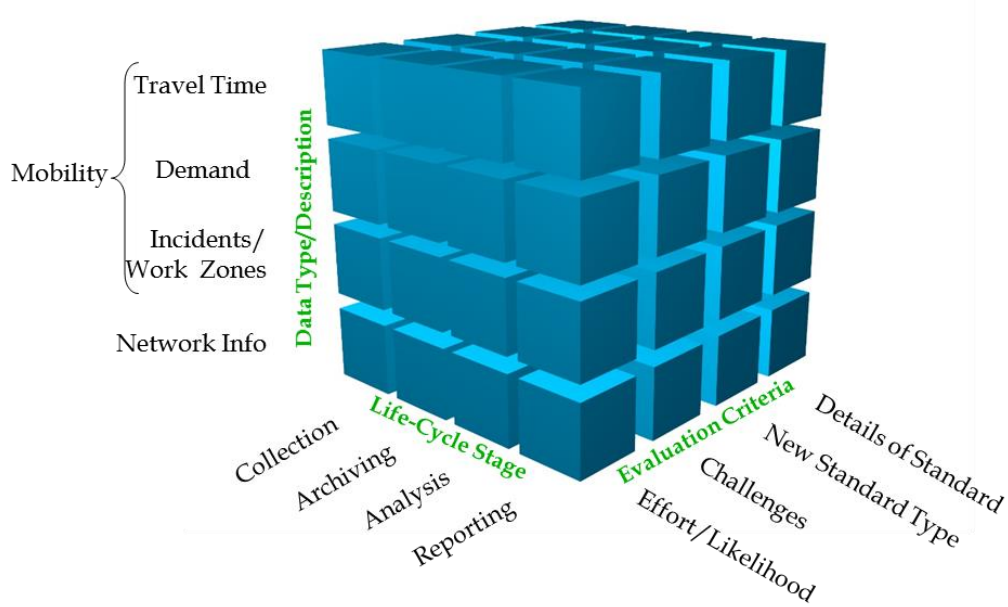


Figure 5: Framework for Evaluating Potential Data Standards

Define Performance Measures and Data Types

Narrowing the Focus

The study team first targeted data types for standardization based on the MAP-21 performance measurement requirements. These were targeted as a way to focus the effort on the measures and data types most relevant for planning and operations. MAP-21 has seven national goals areas:

- Safety;
- Infrastructure condition;
- Congestion reduction;
- System reliability;
- Freight movement and economic vitality;
- Environmental sustainability; and
- Reduced project delivery details.

To support these goals areas, MAP-21 requires the establishment of the following performance measures:

- Pavement condition on the Interstate System and on remainder of the National Highway System (NHS);
- Performance of the Interstate System and the remainder of the NHS;
- Bridge condition on the NHS;
- Fatalities and serious injuries—both number and rate per vehicle mile traveled--on all public roads;
- Traffic congestion;
- On-road mobile source emissions; and
- Freight movement on the Interstate System.

Several of these required performance measures and supporting data have a long history of use in the profession and are subject to several standards. Safety performance measures are well established and the crash data that are used to develop them are subject to several standards, including:

- ANSI D16.1-2007, Manual on Classification of Motor Vehicle Traffic Accidents;
- Model Minimum Uniform Crash Criteria Guideline (MMUCC);
- NEMSIS for EMS data; and
- Model Inventory of Roadway Elements (MIRE) for roadway inventory data.

Likewise, pavement and bridge measures and data have been collected and used for many years, and guidelines and standards already exist, such as FHWA's Practical Guide for Quality Management of Pavement Condition Data Collection. In addition, the measurement of the International Roughness Index (IRI) is required for data provided to FHWA, and is covered in several standards from ASTM International (ASTM E1926 – 08 and ASTM E1364 - 95(2005)).

That leaves performance measures related to traffic congestion, on-road mobile source emissions, and freight movement as the focal point for this study. We have collapsed these measures into two broad categories: mobility and network information. It should be noted that we have not limited our consideration to only the measures being required by FHWA. Also, as will be shown in this document, a wide variety of supporting data – many of which have applications beyond these performance measures – are also considered.

Performance Measures and Data Types

The first column of Table 4 refers to broad *performance measurement types* as defined in NCHRP 551 (Cambridge Systematics et al. 2006). Performance measurement types include roadway congestion, accessibility, and travel time reliability. The second column includes *examples of performance measures* belonging to the same performance measurement type. Examples include level-of-service, mode split, accessibility, vehicle miles traveled, incident report time, and travel cost. The last column, which is of particular interest in this study, lists the *data types* performance measures depend on. Examples include, travel times, network information, and travel demand.

Note that data type does not refer to the actual method of data collection. For example, travel times can be collected using a variety of methods including GPS and connected vehicle data. This is because performance measurement is primarily concerned with the properties and quality of the data and not with the method of data collection that can vary or change based on technology.

Table 4. Example Performance Measures and Data Types

Type of Performance Measure	Performance measures	Data type
Roadway Congestion	Level of Service (LOS based on volume to capacity ratio, travel time index, travel rate index, misery index, speed reduction index, percent of roadway miles operating under capacity, approaching capacity or over capacity, lane-mile duration index, maximum queue length	Link counts, travel times for multiple days, networks to calculate path travel times and lane information, queues
Accessibility to destinations, facilities and services	Percent of target population that can conveniently reach a specific destination (in X minutes), percent of working population within X miles of employment,	Link travel times, roadway networks to calculate path travel

Type of Performance Measure	Performance measures	Data type
	person miles traveled per capita, percent of population within X miles or minutes of the state highway system, percent of facilities that are ADA compliant, number of driveway cuts per mile, percent of urban population with convenient access to public transit, access time to passenger or intermodal facilities	times, transit stops to calculate time to the nearest transit stop, employment location information, census data
Accessibility to different modes	Modal split by trip purpose, average automobile ownership, transit service availability by county, ability of shippers to access desired suppliers, percent of bicycle-miles with bicycle accommodations, percent of roadway-miles with pedestrian accommodations, number of autos per capita	Travel surveys on modal usage or model estimated modal usage, transit network data, freight supplier information, bicycle and pedestrian networks, socioeconomics
Speed	Average speed on a roadway segment or origin-destination pair, ratio between bus and auto speed, percent of high priority highways with average speed of 60 mph, mobility index (VMT at different speed bins)	Link and path travel times for different modes, networks, link counts
Travel time, reliability, delay	Average travel time by mode or across modes for an OD pair, 95-percent reliable travel time, travel time for freight intermodal facilities to highway intermodal facilities, average shipment time (by commodity, mode, local versus long-distance), changes in average, median, and 90th percentile travel time over time, percent difference between second fastest emergency route and the fastest route, buffer time index, deviation from average trip time, number of days when peak period travel time exceeds twice the free flow travel time, total hours of delay, hours of incident related delay, relative delay rate, congestion severity index, hours stopped time per traveler, percent of peak-period travelers delayed	Link and path travel times for multiple days, travel time for freight intermodal facilities, freight shipment times by mode, networks to calculate path travel times, incident location, counts to calculate delayed travelers
System operations efficiency	VMT per lane-mile, passengers per vehicle-mile or hour, percent of road-miles with low speed limits, average circuitry of truck trips between a selected OD pair	Counts by time of day, travel times by time of day for multiple days, networks to calculate fastest and next to fastest paths
Incident Characteristics	Average incident response time, average time to clear incident, percent of incidents cleared in less than X minutes	Timestamped incident activities (timeline)
Crashes	Number of crashes per type, mode, system, location type, crash rates per 100 million VMT, percent reduction of rashes, number of crashes in which speed or traffic violation is a factor, number of chases in highway construction zones, number of fatalities, number of workers killed, number of injuries, crash costs, hours of delay	Crash data elements derived from police accident reports
Travel cost	Trip cost for different modes, shipment cost, vehicle operating costs , dollar loses due to freight delays, economic efficiency/net discounted benefits	Reported or calculated trip costs for different modes, gas and toll information, value of

Type of Performance Measure	Performance measures	Data type
		time
Occupancy	Percent or number of multiple-occupant vehicles, percent work trips by single-occupant vehicles (SOVs) or non-SOVs, average vehicle occupancy	Traveler surveys
Cost efficiency, Fuel Efficiency, economic costs and benefits, direct user costs	Average cost per lane-mile constructed, average maintenance cost per mile, construction and maintenance expenditures per VMT, cost per passenger trip, average fuel consumption per trip by type, annual fuel consumption per VMT	Transportation infrastructure costs

Although a significant number of data types is presented in Table 4, it is also evident that the majority of the performance measures depend on the following subset of data types specific to each mode:

- Travel times;
- Demand (counts, OD flows);
- Incident or work zone data; and
- Network information.

Travel time and transportation demand are fundamental data types that describe the usage of the transportation system and the level of service experienced by travelers. The following paragraphs elaborate on each of the core data types presented in Table 4.

Travel Time Data

Travel times frequently refer to link travel times specific to single mode and a particular transportation network. However, they can also refer to origin to destination or zone-to-zone travel times when accessibility or user experience is measured. Traditional methods transportation practitioners have used include collecting spot speeds using detector data or conducting floating car or GPS surveys. In the past, due to limited of resources, travel time data were confined to a relatively small portion of the transportation network and were not sufficient to describe congestion and network performance in detail.

In the last five years, transportation practitioners have obtained system-wide performance datasets that could not possibly have been obtained with traditional means. Ubiquitous computing and the presence of GPS devices in smartphones has made each traveler a potential data source. The data obtained from a smartphone, bus AVL, or any other device with positional information is called a trajectory. Bicycle and pedestrian trajectory datasets are increasingly made available by equipping volunteers with apps that record and transmit traveler positions. It is expected that this trend will continue as more sensors and the connected vehicle data become available.

Trajectory data even for 1% of a city's population for a single day can be very voluminous, easily reaching many terabytes in size. Transportation practitioners do not usually deal with raw trajectory data. Rather, they are asked to work with probe vehicle data that have already been processed and aggregated to time intervals and predefined network links. In terms of data size, even the aggregated travel time data can be many hundreds of gigabytes for an entire state such as Florida and a single year. Such data can be purchased by multiple vendors and they are frequently reported over Traffic Message Chanel links or over the data vendor's proprietary network.

The data collection methodology, characteristics of the vehicle fleet, and sample size is not made transparent to the user. As a result, confidence intervals, a universal practice in statistics, are not provided. Furthermore, the data preparation process is often a black box; some vendors may substitute missing values with the historical average while some others may interpolate from neighboring observations. Frequently, TMC links have gaps or overlaps between them making the calculation of path travel times harder.

Travel Demand Data

Travel demand data can take two forms, link count data and origin-destination trip data. Count data on links is the most frequently used data type; vehicle miles traveled and other very important measures depend on count data. System performance measures related to cost/benefit analysis that need to include the entire population use OD flows. In the past, OD flow data were estimated based on region-wide travel demand models that used household travel surveys. However, in the last five-years, cellphone companies have started selling phone records from which traveler trajectories and traveler trips can be analytically extracted. Data aggregators have emerged whom, after buying and anonymizing the phone records, sell the aggregated OD tables to public agencies and any other interested parties.

Transportation practitioners have been responsible for the collection and analysis of count data collected by infrastructure owned by public agencies. In many cases, public agencies decide to provide public access to count data in the form of roadway link counts, bus boardings, or bicycle availability by location using in-house or publicly available formats. The primary technology for collecting roadway counts has been inductive loop detectors. A wide range of new technologies including magnetometers, infrared, video systems, Bluetooth, and toll tags is expected to increase data availability and coverage.

In almost all cases, raw values of count or travel time data are not directly used in performance measures. Data need to be filtered, cleaned, and aggregated to produce values that can form the basis for performance measurement. This extraction process is very laborious and requires contextual knowledge of the data collection method and its limitations. Although a small number of transportation practitioners are engaged in such tasks it is important to define quality control measures for the data being reported.

Cellphone information due is another source of travel demand data. Smartphones triangulate their position based on the nearby cellphone towers many times a day, often more than 200. The time-stamped locational information has the potential to identify trip patterns for the entire city. However, because the data are still a sample, sample-biases, data gaps, and missing trip purpose information pose significant problems.

Due to privacy considerations, many Big Data sources including cellphone and GPS data may not be readily available to the public or the transportation engineer. Even if data are anonymized, there may still be ways to identify specific individuals. For example, New York City time-stamped origin-destination taxi data that had been previously anonymized were combined with available photos and used to identify trips of public persons and publish how much they tipped the driver. Research shows that in some cases when disaggregate data are included, it may be impossible to anonymize a dataset. In a recent and well-cited article in Nature's scientific reports that examined the privacy bounds of human mobility researchers (Montjoye, Yves-Alexandre, et al. 2013) conclude that:

> We study fifteen months of human mobility data for one and a half million individuals and find that human mobility traces are highly unique. In fact, in a dataset where the location of an individual is specified hourly, and with a spatial resolution equal to that given by the carrier's antennas, four spatio-temporal points are enough to uniquely identify 95% of the individuals.

Based on the article, privacy issues are likely to emerge when large disaggregate datasets become public. Reporting only aggregate statistics by small number of data vendors may circumvent the problem.

Incident Data

Incident characteristics are an extremely important data source for conducting planning analyses. These data describe incident location, characteristics, and the nature of blockages (duration of lane or shoulder blockages). Although incident data are being collected today by transportation and emergency response agencies, complete uniformity has not yet been achieved.

The duration of the blockages caused by incidents is the most important piece of information needed for planning applications. These data are usually as readily available from databases, especially those that already have incident blockage/duration information; therefore, there is no additional effort involved in assembling these data.

Possible sources of data include the following (these are not mutually exclusive, and multiples may be used in combination to feed data into the system):

- Service Patrol/Incident Response Team data entry – Personal digital assistant/laptop in the vehicles;
- Transportation management center (TMC) operator data entry – Automate the conversion of web information to database entries; and
- Police computer-aided dispatch (CAD) files.

Work Zone Data

For planning applications, work zone data are useful in understanding congestion patterns and explaining why certain congestion trends emerged. Work zone data are sometimes collected as part of incident data, but also independently by construction and maintenance personnel. Major types of data on work zones characteristics are lane information, shoulder information, and signing information. For each of those categories, there are data such as lane modification, lane shift, and shoulder restrictions.

Recent work on work zone performance measures has focused on mobility, throughput, and safety specifically for work zones (Ulman, Gerry et al. 2013) have also been recommended as performance measures. In general, the types of information that should be collected are work zone type, longitudinal characteristics and extent (including details on transition zones and tapers), duration of work zone characteristics, and major cross-section characteristics of the work zone. Separate tracking for both actual and planned changes should be done.

Network Data

Transportation networks describe the infrastructure on which movement and mobility options are exercised. Transportation network information is an integral, often implicit, part of the performance measurement process. Transportation network information for any mode is becoming increasingly important because it can facilitate or impede the flow of data for off-line or real-time applications.

In a connected world, accurate transportation network information is fundamental. Recognizing this, in the last 10 years several private companies have been deploying fleets of instrumented vehicles collecting GPS, Lidar, video and other information to record the continuously changing roadway infrastructure. The information collected includes roadway alignment, number of lanes, lane positioning, signage, traffic light positions, and work zone locations. This information is used to build routable roadway networks and is

becoming increasingly valuable to automakers for connected vehicle purposes (Economist, April 16 2016).

A large number of transportation data that describe human mobility such as travel times, incidents, and counts are defined on a specific network. When mobility data have not been mapped to a network or when only the longitude and latitude information exists, a laborious human supervised procedure is required to assign the count or the incident on the proper link or intersection.

From the data archiving and reporting perspective, networks are defined as a set of nodes and links. Nodes correspond to intersections or junctions, locations where drivers can select between one or more options. Links correspond to roadway segments between nodes. Both nodes and links usually have multiple attributes that define intersection type, the presence of a signal, the number of lanes, etc.

For the same region, it is typical to find several transportation networks in use by different public agencies or by different departments of the same agency. This is often the case because different agencies are interested in different attributes of the same network. Another source of disparity has to do with update frequency; some agencies may update their networks more frequent than others or use different sources of input for the update.

Network information can be represented in data tables. However, unlike database data tables that can be merged easily using primary and foreign keys, it is hard and laborious to associate two networks, transfer information, or update the one using data from the other. In developing associations between the elements of different networks, a complicated many-to-many table needs to be constructed. There are several reasons for this:

- Node locations do not match and link geometry is not the same;
- Nodes and links have different identifiers so there is not an easy one-to-one association;
- Number of nodes and links are different as many networks contain different amount of detail; and
- Links have been grouped together for reporting purposes or intersections have been simplified or omitted.

Transferring information between networks is far from trivial and is one of the main reasons, from a technical perspective, to have data reside in silos. For example, it often takes significant effort to transfer incident, TMC speeds, or count data from latitude/longitude or from one network used for reporting to another one used for modeling or simulation. Furthermore, it is hard to identify an update mechanism between the networks so that when new information is introduced in one network it is also transferred and updated everywhere else.

Another example that demonstrates the difficulties in transferring information between networks is the usage of General Transit Feed Specification (GTFS) data in transportation modeling. GTFS contains information on transit network alignment and on scheduling. GTFS information is generally considered more accurate than the legacy information assembled by transportation modelers manually over the years. However, even if GTFS is generally recognized as a better source for transit information, adoption is very slow because it is not easy to integrate the GTFS and planning networks.

Challenges to Developing Data Standards

Each data type is unique and developing standards for them may involve different obstacles or challenges. This section provides background for each challenge and for the challenges with dealing with

Big Data to exemplify data standard barriers in general. The challenges listed below will be used later to evaluate the potential of different standardization efforts.

- There are disincentives, or even conflicts of interest, for the data producers to collaborate on a common specification. For example, probe vehicle data aggregators may oppose publishing the number of probe vehicles behind each measurement based on revealing trade secrets.
- Multiple competing standards. Agencies have already been using different data formats, different procedures, and may use different definitions for the same data type. A typical example in this category is the transit field before the adoption of GTFS.
- Wide variation in data content. A common denominator may still be found but exceptions may be so frequent that they endanger the universal applicability of the standard. Transportation network data is a typical example. One agency may require a very detailed transportation network while an MPO can replace local streets with zone connectors.
- High barrier of entry. Using the proposed standard requires resources and knowledge that is not widely available by the majority of the stakeholders. Such an example may be the adoption of network algorithms for conflation and data transfer.
- The structure of data content is highly dynamic. By the time the standard is developed, it may already be outdated. New datasets from cellphones or other sensors may fall into this category.

Challenges Associated with Big Transportation Data

Much of the data available for the travel time component of the mobility category can be considered under the rubric of Big Data. While a strict definition of Big Data is elusive, defining characteristics include: (1) voluminous data continuously collected data at very fine spatial and temporal levels and (2) data that are unstructured or without context, that is, additional data are needed to make it usable. Data from sensors and probes fall into this category. In this section, we provide some of the challenges associated with these data.

Data collection in the transportation field was tailored to project-specific transportation needs while the volume of data was small due to resource limitations. Data sufficiency was the major problem transportation engineers faced. Currently, due to crowdsourcing and ubiquitous instrumentation (Internet of Things) the volume of data is often higher than transportation practitioners are used to handling. The new datasets from smartphones, GPS devices, and connected vehicles help transportation practitioners monitor mobility in ways that would have been impossible with traditional means. However, integrating datasets from other domains that use crowdsourcing or were collected for different purposes results in integration and data quality issues that should not be overlooked. Based on an authoritative report by the National Academies Press (Council, N. 2013), the following challenges are associated with massive data analysis and Big Data:

- Coping with sampling biases and heterogeneity. From a statistical standpoint, this is probably the biggest challenge transportation practitioners face when trying to integrate Big Data that have been aggregated using non-transparent procedures. Examples are numerous; free flow speeds may not be reliably obtained from a probe vehicle dataset that contains too many trucks or does not have records over the speed limit. When data collection and data processing is not transparent, the transportation practitioner does not have a way to correct for sample bias. As a result, if data are used without any adjustment they can distort the calculation of performance measures.
- Working with different data formats and structures. Although there are multiple formats that deviate from the well-known tabular csv files, this is less of an issue for transportation practitioners. Software tools exist to convert between data formats.

- Tracking data provenance from generation through data preparation. When data manipulation is not transparent the users of the data do not know when and who modified the content of the data. This can also happen when multiple agencies share data and manipulate the same data.
- Data validation. Frequently, third parties conduct an evaluation of probe vehicle or other Big Data. However, conclusions are often nuanced and accessible only to those with substantial background in the field.
- Developing methods for visualizing massive data. User-friendly spreadsheet software is not capable of storing and analyzing the available data streams.
- Coping with the need for real-time analysis and decision-making. Real-time analysis such as the prediction of travel time information to be displayed in variable message signs necessitates automated procedures that in real-time assess the quality of the data, discard records, and make corrections.
- Enabling discovery and integration. This is a challenge related to data analytics and not standardization.
- Enabling data sharing. Frequently transportation data exist in separate silos. The reasons can be technical but are also institutional. Resolving the technical issues by using common formats and standards can help break down barriers.
- Developing scalable algorithms that exploit parallel architectures. Tabular data that can be stored in spreadsheets can be easily and efficiently analyzed. However, transportation agencies are starting to amass many gigabytes of data that can stress traditional SQL databases to their limit.

Regarding sampling bias, the fact that we are dealing with larger datasets of the same population does not mean that we need to overlook how representative our sample is. For example, information about activity participation from popular apps such as Foursquare may come from a younger crowd that is actively using the app. Likewise, commuters may be underrepresented (OECD 2015) in Strava, a bicyclist app that has collected 80 million bicycle rides and has made the data available to planners. In addition to sampling bias, the lack of socioeconomic data for privacy or other reasons in many datasets can limit their use in transportation applications.

It should be noted that there are still rural areas in the country where the use of probe data is limited due to the lack of sensors and/or emitting devices. There are still areas of the country where there is coverage by only one or two cell towers of only one vendor, making location triangulation problematic and reducing the collected sample size relative to the population to where it may no longer be statistically representative.

Overcoming Challenges and Assessing the Likelihood of Adoption

There are several factors that affect the ability to adopt a new standard. Each of the potential standards are evaluated based on the factors presented below.

Business case: A clear problem needs to be presented to stakeholders and solved by the standardization process. The benefits should be widespread and universal. The cost of compliance for agencies that have adopted a competing format or have yet to establish a policy should be proportional to the perceived benefit from adopting the standard.

Simplicity to comply**:** A clear path should be presented to agencies or different stakeholders. The benefits of compliance may not always be evident and unanimity may not be achieved at the beginning of the effort. However, a critical mass of stakeholders should recognize the benefits of compliance and

should be able to demonstrate them to the rest of the community. As the standard gains momentum it should be clear to the non-adopters that the costs of inaction are higher than the costs of compliance.

Data/software provider support: Many of the newer datasets from probe vehicles or other sensors are purchased from data aggregators. Transportation agencies may be only a small subset of the market and as a result, the transportation sector may have a limited influence changing the collection or reporting methods. In other fields, such as Internet standardization, it is common practice to observe competing companies collaborate and contribute to a common format. Even if competition forces may prevent data providers to reveal private information there should be clear incentives or mandates for them to cooperate on a common standard that is not designed to influence competition.

Alternative of competing formats: Alternative formats are often associated with competing interests. Government's actions should not directly influence or alter the competitive landscape. Nevertheless, transportation data that are collected with public funds should be made available in a format that other parties or agencies can access. In the past, transportation practitioners have worked with software that adopted a "walled garden approach." This approach ensures compatibility and interoperability between all components of a single vendor but leaves the transportation professional clueless when data have to be exchanged between software tools or agencies.

Level of effort to build a standard: The level of effort to build a standard depends on the prior research that has been conducted. In some cases, such as the analysis of travel time data, significant research already exists. In other cases, such as network conflation, it may require significant research to develop procedures that can be applied to map or conflate two networks.

Results

This section presents the results of applying the evaluation framework to the mobility and network information data categories. Table 5 shows the data types selected for possible standardization along with standards that already exist for these data types. In all cases, the existing standards are not sufficient to meet the needs of planners and operators. Some of them are communication standards for transmitting messages. Others do not tackle the issues of interest.

Table 6 provides the matrix evaluation of the data types. The life cycle stages were used to identify specific topics that potential standards would address. In the last column, "Challenges and Likelihood of Adoption", we have only provided entries where a standard has an aspect that is unique from the general barriers to standards development discussed in earlier sections.

Also in Table 6, we have included an attribute for "Type of Standard". This is meant to define the nature of the standard, using the classification scheme developed by ASTM. It does not imply that the topic should be developed as a formal standard, as published by a Standards Development Organization (SDO). The decision on whether to develop the topic as a formal standard or as guidance document is left to the next phase of the research. The ASTM classification scheme is as follows (text taken directly from http://www.astm.org/SNEWS/OCTOBER_2000/oct_howto.html).

Specification

An explicit set of requirements to be satisfied by a material, product, system or service.

A wide variety of subjects are covered in ASTM specifications. Examples of specifications include, but are not limited to, requirements for: physical, mechanical, or chemical properties, and safety, quality, or performance criteria. A specification identifies the test methods for determining whether each of the requirements is satisfied.

Classification

A systematic arrangement or division of materials, products, systems, or services into groups based on similar characteristics such as origin, composition, properties, or use.

Because each ASTM committee is unique, the classifications written by one committee might be quite different from those written by another. The one common aspect to all ASTM classifications is that each must have a mandatory Basis of Classification section. This is the most important part of any classification as it sets up categories in which groupings are made.

Practice

A definitive set of instructions for performing one or more specific operation that does not include a test result.

As mentioned in the Test Methods section, occasionally there is confusion about the differences between practices and test methods. In addition, questions sometimes surface over the differences between practices and guides, which are defined below. The difference between a practice and a guide is that a practice underscores a general usage principle while a guide suggests an approach. A guide connotes accepted procedures for the performance of a given task.

Examples of practices include, but are not limited to application, assessment, cleaning, collection, decontamination, inspection, installation, preparation, sampling, screening, and training.

Guide

A compendium of information or series of options that does not recommend a specific course of action.

A guide may propose a series of options or instructions that offer direction without recommending a definite course of action. The purpose of this type of standard is to offer guidance based on a consensus of viewpoints but not to establish a standard practice to follow in all cases. A guide is intended to increase the awareness of the user concerning available techniques in a given subject area, while providing information from which subsequent testing programs can be derived.

Bundling the individual standards (rows in Table 6) is a good strategy that will aid standard development. For example, it is possible to bundle all the items listed for "travel time" into a single standard rather than treat them as separate standards as the same stakeholders would be involved in all of them.

Table 5. Existing Standards for the Selected Data Types

Data Type	Description	Life Cycle Stage	Existing Standards
TRAVEL TIME			
Vehicle Probe Data: travel times on links	Derived from instantaneous GPS speeds	Collection Archiving Analysis	Emerging only, related AV/CV. IEEE 802.11 & IEEE 1609 Standards in place for wireless communications protocol. SAE J2735 DSRC Message Set standard is intended to for transferring information between vehicles and roadside devices, between vehicles themselves and between vehicles and centers using other wireless mediums for non-time critical applications
Vehicle Probe Data: travel times on links	Derived from vehicle trajectories	Collection Archiving Analysis	
O/D Travel Times	Derived from vehicle trajectories (time and location) collected via GPS; similar to data in Basic Safety Message for AV/CV	Collection Archiving	
Bicycle Travel Times	GPS-based but not currently widespread	Collection	None
Core Travel Time Measures for Extended Facilities and Networks	Standard set of measures to allow cross-study comparisons; both measurements and models	Analysis Reporting	Latest version of Highway Capacity Manual will have some guidance on how to calculate measures. FHWA requirements for MAP-21 will have guidance but only for the unique MAP-21 measures
Performance Measure Development	Metadata specifications	Archiving	ASTM E2468-05 could be adapted but not specific to performance measures right now
DEMAND			
Demand in Congested Flow Conditions	Not possible to observe actual demand from empirical volume data due to queuing	Analysis	Various research methods exist; no standard

Data Type	Description	Life Cycle Stage	Existing Standards
O/D Travel Flows	Number of trips between O/D pairs as measured from personal devices and other sources	Collection	None
Survey Data for Planning and Operations	Guidelines for defining a trip, type of activity, etc. for respondents	Collection	None
INCIDENT AND WORK ZONES			
Traffic Incident and Work Zone Data for Planning and Operations	Detailed data on duration of blockages by incident and work zone type	Collection	FHWA has several rules and guidance documents for Work Zone safety and mobility (e.g., TMPs, work zone operations) for individual construction projects but little associated with the collection, archiving, and analysis of that data for planning purposes and how to improve operations. Incidents: IEEE P512 and National Incident Management System offer standard data elements but these are adequate for some planning and operations analysis
		Archiving	
	Basic work zone characteristics	Collection	
Traffic Incident and Work Zone Performance Measures for Planning and Operations	"Output" level performance measures	Reporting	None
NETWORK INFORMATION			
Network Representation	Highway Network Characteristics	Analysis	FHWA's ARNOLD
	Transit Network Characteristics	Analysis	GTFS
	Conflation of multiple networks	Analysis	None

Table 6. Potential Standards for the Selected Data Types

Data Type	Description	Life Cycle Stage	Type of New Standard	Details of Standard	Challenges and Likelihood of Adoption
TRAVEL TIME					
Vehicle Probe Data: travel times on links	Derived from instantaneous GPS speeds	Collection	Practice	Number of vehicles used in a reporting measurement	Requires vendor participation
		Archiving	Test Method	Quality control procedures	
			Practice or Guide	Network conflation method	Might limit agency flexibility in developing some measures; metadata may be more appropriate (see below)
		Analysis	Practice	Spatial & temporal aggregation procedures	
			Practice or Guide	Integration procedures with "congestion source" data (volume, incident, work zones, weather); important for assigning causality	
Vehicle Probe Data: travel times on links	Derived from vehicle trajectories	Collection	Practice	Number of vehicles used in a reporting measurement	Requires vendor participation
		Archiving	Test Method	Quality control procedures	
			Practice	Method for computing link travel time	
		Analysis		Same as for "Vehicle Probe Data: travel times on links"	
O/D Travel Times	Derived from vehicle trajectories (time and location) collected via GPS; similar to data in Basic Safety Message for AV/CV	Collection	Practice	Number of vehicles used in a reporting measurement	Requires vendor participation
		Archiving	Test Method	Quality control procedures related to mid-trip stop-offs; specification of paths taken between O/D pairs; storage in OMX format	
Bicycle Travel Times	GPS-based but not currently widespread	Collection		Collection and imputation methods	Technology and use of it may not be mature or widespread enough to identify the details of the standard
Core Travel Time	Standard set of	Analysis	Practice or	Computation procedures based	

Data Type	Description	Life Cycle Stage	Type of New Standard	Details of Standard	Challenges and Likelihood of Adoption
Measures for Extended Facilities and Networks	measures to allow cross-study comparisons; both measurements and models		Guide	on the different data sources discussed above	
		Reporting	Specification	Standard reporting levels for key performance measures (e.g., Travel Time Index of 1.0-1.2, 1.2-1.4…)	
Performance Measure Development	Metadata specifications	Archiving	Practice or Guide	Documentation requirements for agencies wishing to create their own travel time-based performance measures	
DEMAND					
Demand in Congested Flow Conditions	Not possible to observe actual demand from empirical volume data due to queuing	Analysis	Practice or Guide	Development of a standard methodology for computing demand from field measurements under congested condition	
O/D Travel Flows	Number of trips between O/D pairs as measured from personal devices and other sources	Collection	Practice or Guide	Development of a standard methodology for computing O/D flows from raw location data	Requires vendor participation; may directly affect proprietary vendor methods, resulting in unwillingness to change
Survey Data for Planning and Operations	Guidelines for defining a trip, type of activity, etc. for respondents	Collection	Specification	Standard would cover both definitions for common data items as well survey methods and experimental design	
INCIDENT AND WORK ZONES					
Traffic Incident and Work Zone Data for Planning and Operations	Detailed data on duration of blockages by incident and work zone type	Collection	Specification	Every time a blockage changes, a new record should be created	Requires reconfiguration of many data collection systems already in place at traffic management centers and construction units
		Archiving	Specification	Detail on lane configuration vs. blockage	
	Basic work zone	Collection	Specification	Standard data elements related	

Data Type	Description	Life Cycle Stage	Type of New Standard	Details of Standard	Challenges and Likelihood of Adoption
	characteristics			to: work zone type; longitudinal characteristics and extent (including details on transition zones and tapers); duration of work zone characteristics; major cross-section characteristics	
Traffic Incident and Work Zone Performance Measures for Planning and Operations	"Output" level performance measures	Reporting	Specification	Lane-hours lost and shoulder-hours lost due to incidents and work zones; also normalized by dividing by number of original lanes; useful in combination with demand and travel time measures for diagnostics	
NETWORK INFORMATION					
Network Representation	Highway Network Characteristics	Analysis	Guide	Not a new standard but guidance on how to adopt existing FHWA ARNOLD standard	Major effort required to revamp legacy systems
	Transit Network Characteristics	Analysis	Guide	Not a new standard but guidance on how to adopt existing GTFS standard to transit applications	
	Conflation of multiple networks	Analysis	Guide	Guidelines on network conflation and transferring of information	Requires vendor involvement

Conclusions

In summary, transportation data standards, guidelines and specifications for planning and operations are feasible and desirable. The areas for standards that should be prioritized are travel time, demand, incidents and work zones, and network information.

Standards, guidelines and specifications are most successful if they:

- Have a clear business purpose;
- Are clear in application, specificity and versioning;
- Are developed with broad outreach and buy-in;
- Are well defined and simple;
- Are open standards;
- Are forward looking; and
- Involve a national or worldwide community.

In terms of data collection, transportation data can be classified into data collected by transportation agencies such as link counts or surveys and those that are purchased from data vendors. When data are collected by an agency, transportation practitioners have full control over the data collection, sampling, preparation, formatting, and aggregation steps. The gradual proliferation and the increased capabilities of sensors such as detectors and video cameras is certain to increase the amount of data practitioners handle and as a result the amount of oversight, measurement, and control over the transportation network. Guidelines already exist for some types of data and can be extended to handle different and increasing numbers of sensor types.

Third party data, purchased from data aggregators, have the potential to change the landscape of performance measurement. The NPMRDS dataset is one such example where the positions of millions of smartphones or other geo-location devices are aggregated to produce, as a first step, link travel times. It is expected that this trend will continue and data from connected vehicles, cellphone tower triangulations, and all kinds of embedded sensors in devices around us (Internet of Things) will become available. As a result, transportation practitioners will have additional opportunities and options to measure human mobility at the micro level such as vehicle acceleration, and pedestrian location or the macro or city level such as trip origins and destinations. In fact, every company and app that has a large user base can track the mobility of its users and after anonymizing and aggregating the data it can build a third party product for all interested parties to use.

Most of the times the volume and variety of such data can safely classify them into the Big Data realm although this term can be ubiquitous. It should be noted though that since the purpose of data collection is not associated with transportation needs and very often, there is little transparency over sample sizes and the procedures used to filter and aggregate the data, some of the third-party datasets might not be best suited for transportation planning and operations.

In general, transportation practitioners, at least those involved in performance measurement, are not likely to be responsible for extracting reliable measurements from often-noisy raw Big Data such as cell-tower triangulated positions. Rather, the great majority of transportation practitioners will use the curated and aggregated data products from third party vendors.

Regardless, it is very important that certain guidelines exist to ensure that the properties and quality of the data is such that can support transportation applications. This is all the more important because unlike data collected by transportation agencies aggregated Big Data may not a) use the definition of specific

transportation terms b) provide enough information to integrate with other transportation specific data sources c) have the properties and quality a transportation professional expects from tailored data collection methods. One such obvious measure that is often missing now is the notion of a confidence interval, a notion that has been a standard reporting methodology in statistics.

Table 7. Potential Standards for the Selected Data Types.

Data Type	Description	Life Cycle Stage	Type of New Standard	Details of Standard	Challenges and Likelihood of Adoption
TRAVEL TIME					
		Archiving	Test Method	Quality control procedures related to mid-trip stop-offs; specification of paths taken between O/D pairs; storage in OMX format	
Bicycle Travel Times	GPS-based but not currently widespread	Collection		Collection and imputation methods	Technology and use of it may not be mature or widespread enough to identify the details of the standard
		Archiving	Specification	Detail on lane configuration vs. blockage	
	Basic work zone characteristics	Collection	Specification	Standard data elements related to: work zone type; longitudinal characteristics and extent (including details on transition zones and tapers); duration of work zone characteristics; major cross-section characteristics	
Traffic Incident and Work Zone Performance Measures for Planning and Operations	"Output" level performance measures	Reporting	Specification	Lane-hours lost and shoulder-hours lost due to incidents and work zones; also normalized by dividing by number of original lanes; useful in combination with demand and travel time measures for diagnostics	
NETWORK INFORMATION					
Network Representation	Highway Network Characteristics	Analysis	Guide	Not a new standard but guidance on how to adopt existing FHWA ARNOLD standard	Major effort required to revamp legacy systems
	Transit Network Characteristics	Analysis	Guide	Not a new standard but guidance on how to adopt	

Data Type	Description	Life Cycle Stage	Type of New Standard	Details of Standard	Challenges and Likelihood of Adoption
				existing GTFS standard to transit applications	
	Conflation of multiple networks	Analysis	Guide	Guidelines on network conflation and transferring of information	Requires vendor involvement

CHAPTER 3

Research Problem Statement Development

This chapter summarizes Task 4 activities of the scoping study to establish standards and guidance for data for transportation planning and traffic operations. The objective of Task 4 was to develop a research problem statement to accomplish the research related to developing standards and guidance.

The chapter is organized in the following sections:

- Problem Statements – This section summarizes details regarding areas most ripe for standards;
- Standard Development Steps – This section describes the general steps that should be taken to develop the standards or guidance;
- Recommendations – This section recommends next steps; and
- Research Problem Statement – This section includes a research problem statement that could be used to proceed to the next step of standards and guidance development for transportation planning and traffic operations.

Note that a "standard" is a method or process that:

- Has a legal basis and can be made mandatory;
- Is a normative document from a formal standard body;
- Is passed though a full and open consensus process; and
- Is implemented on a national level and there is a strong mandate to apply them.

Guidance includes methods and documentation of processes that are informative and do not necessarily require consensus.

Problem Statements

This section documents specific problem statements in five areas or "bundles" that were determined to be ripe for standardization/guidance in Task 3. The five bundles are travel time, demand, incident and work zones, network and transit.

For each bundle, the following is described:

- Business Case: Need for standardization or guidance, description of the problem and specifics as to what needs to be standardized.
- Recommended Standards: Specifics based on Table 3 in the Task 3 report.
- Challenges: Anticipated challenges related to developing standards in each area.
- Research Sufficiency: Discussion of past research on the topic and likelihood of a standards body or panel coming to consensus on the standard.
- Stakeholders: Which stakeholder groups (in general) need to be involved in development and approval of standards and who will use them.
- Level of Effort to Develop and Adopt: The level of effort to build a standard depends on the prior research that has been conducted. In some cases, such as the analysis of travel time data, significant research already exists. In other cases, such as network conflation, it may require significant research to develop procedures that can be applied to integrate geospatial data or conflate transportation networks.

Note: Standards are further delineated as practice, test method, guide or specifications for the purposes of the analysis. These are meant to define the nature of the standard, using the classification scheme developed by the American Society of Testing and Materials (ASTM). It does not imply that the topic should be developed as a formal standard, as published by a Standards Development Organization (SDO). The decision on whether to develop the topic as a formal standard or as guidance document is left to the next phase of the research. The ASTM classification scheme is as follows (text taken directly from http://www.astm.org/SNEWS/OCTOBER_2000/oct_howto.html). This scheme can apply to either.

1. Specification - An explicit set of requirements to be satisfied by a material, product, system or service.

A wide variety of subjects is covered in ASTM specifications. Examples of specifications include, but are not limited to, requirements for: physical, mechanical, or chemical properties, and safety, quality, or performance criteria. A specification identifies the test methods for determining whether each of the requirements is satisfied.

2. Practice - A definitive set of instructions for performing one or more specific operation that does not include a test result.

The difference between a practice and a guide is that a practice underscores a general usage principle while a guide suggests an approach. A guide connotes accepted procedures for the performance of a given task.

Examples of practices include, but are not limited to application, assessment, cleaning, collection, decontamination, inspection, installation, preparation, sampling, screening, and training.

3. Guide - A compendium of information or series of options that does not recommend a specific course of action.

A guide may propose a series of options or instructions that offer direction without recommending a definite course of action. The purpose of this type of standard is to offer guidance based on a consensus of

viewpoints but not to establish a standard practice to follow in all cases. A guide is intended to increase the awareness of the user concerning available techniques in a given subject area, while providing information from which subsequent testing programs can be derived.

Table 8. Potential Standards for the Selected Data Types.

Data Type	Description	Life Cycle Stage	Type of New Standard	Details of Standard	Challenges and Likelihood of Adoption
TRAVEL TIME					
Vehicle Probe Data: travel times on links	Derived from instantaneous GPS speeds	Collection	Practice	Number of vehicles used in a reporting measurement	Requires vendor participation
		Archiving	Test Method	Quality control procedures	
			Practice or Guide	Network conflation method	Might limit agency flexibility in developing some measures; metadata may be more appropriate (see below)
		Analysis	Practice	Spatial & temporal aggregation procedures	
			Practice or Guide	Integration procedures with "congestion source" data (volume, incident, work zones, weather); important for assigning causality	
Vehicle Probe Data: travel times on links	Derived from vehicle trajectories	Collection	Practice	Number of vehicles used in a reporting measurement	Requires vendor participation
		Archiving	Test Method	Quality control procedures	
			Practice	Method for computing link travel time	
		Analysis		Same as for "Vehicle Probe Data: travel times on links"	
O/D Travel Times	Derived from vehicle trajectories (time and location) collected via GPS; similar to data in Basic Safety Message for AV/CV	Collection	Practice	Number of vehicles used in a reporting measurement	Requires vendor participation
		Archiving	Test Method	Quality control procedures related to mid-trip stop-offs; specification of paths taken between O/D pairs; storage in OMX format	
Bicycle Travel Times	GPS-based but not currently widespread	Collection		Collection and imputation methods	Technology and use of it may not be mature or widespread enough to identify the details of the standard

Data Type	Description	Life Cycle Stage	Type of New Standard	Details of Standard	Challenges and Likelihood of Adoption
Core Travel Time Measures for Extended Facilities and Networks	Standard set of measures to allow cross-study comparisons; both measurements and models	Analysis	Practice or Guide	Computation procedures based on the different data sources discussed above	
		Reporting	Specification	Standard reporting levels for key performance measures (e.g., Travel Time Index of 1.0-1.2, 1.2-1.4…)	
Performance Measure Development	Metadata specifications	Archiving	Practice or Guide	Documentation requirements for agencies wishing to create their own travel time-based performance measures	Can build on existing FHWA documentation outline
DEMAND					
Demand in Congested Flow Conditions	Not possible to observe actual demand from empirical volume data due to queuing	Analysis	Practice or Guide	Development of a standard methodology for computing demand from field measurements under congested condition	
O/D Travel Flows	Number of trips between O/D pairs as measured from personal devices and other sources	Collection	Practice or Guide	Development of a standard methodology for computing O/D flows from raw location data	Requires vendor participation; may directly affect proprietary vendor methods, resulting in unwillingness to change
Survey Data for Planning and Operations	Guidelines for defining a trip, type of activity, etc. for respondents	Collection	Specification	Standard would cover both definitions for common data items as well survey methods and experimental design (e.g., http://onlinepubs.trb.org/onlinepubs/nchrp/nchrp_rpt_571.pdf and http://www.travelsurveymanual.org/	
INCIDENT AND WORK ZONES					
Traffic Incident and Work Zone Data for Planning and	Detailed data on duration of blockages by incident and work	Collection	Specification	Every time a blockage changes, a new record should be created	Requires reconfiguration of many data collection systems already in place at traffic management centers and
		Archiving	Specification	Detail on lane configuration vs.	

Data Type	Description	Life Cycle Stage	Type of New Standard	Details of Standard	Challenges and Likelihood of Adoption
Operations	zone type			blockage	construction units
	Basic work zone characteristics	Collection	Specification	Standard data elements related to: work zone type; longitudinal characteristics and extent (including details on transition zones and tapers); duration of work zone characteristics; major cross-section characteristics	
Traffic Incident and Work Zone Performance Measures for Planning and Operations	"Output" level performance measures	Reporting	Specification	Lane-hours lost and shoulder-hours lost due to incidents and work zones; also normalized by dividing by number of original lanes; useful in combination with demand and travel time measures for diagnostics	
NETWORK INFORMATION					
Network Representation	Highway Network Characteristics	Analysis	Guide	Not a new standard but guidance on how to adopt existing FHWA ARNOLD standard	Major effort required to revamp legacy systems
	Transit Network Characteristics	Analysis	Guide	Not a new standard but guidance on how to adopt existing GTFS standard to transit applications	
	Conflation of multiple networks	Analysis	Guide	Standards on network conflation and transferring of information	Requires vendor involvement

Travel Time Standard

Business Case

Historically, transportation agencies have been responsible for collecting travel time data and have had control over the equipment and collection methods. In recent years, new forms of travel time data derived from GPS-enabled devices by the private sector have inundated the market. These data are extremely valuable for a wide variety of real-time applications and historical analyses. However, the conditions under which the data are collected are largely unknown. Standards would ensure that the data meet minimum requirements, are pre-processed in a similar way, and include minimal data quality indicators. Standardization would lead to more accurate comparisons of performance across agencies.

The standards considered here deal with post-collection of the data by private sources. While it might be desirable to have standards governing data collection (e.g., equipment calibration and testing), these would most likely impose undue burdens on the private sector and could lead to a restricted market.

Recommended Standard

Although this bundle represents multiple facets of travel time data, the research team believes that the same stakeholders would have an interest in all of the facets. Therefore, a single standard related to travel time data is recommended. The following is a discussion of the various facets of the standard; these are summarized in Table 8.

Travel Times Derived from Vehicle Probe Data

Number of vehicles used in a reporting measurement should be specified. Requiring vendors to report the sample size associated with a travel time measurement would lead to better use of the data by agencies. Currently, it is possible that a travel time measurement on link for a short time period (e.g., 1-5 minutes) can be based on a single vehicle or several hundred. In some cases, agencies may wish to establish minimum sample sizes for specific types of analyses.

The spatial coverage of the measurements on a reporting segment (e.g., link) should be specified. That is, what portion of the reporting segment had measurements taken on it. This could be implemented as the range from the first measurement to the last measurement along a reporting segment.

Quality Control (QC) Procedures

Very little guidance currently exists on QC procedures for vehicle probe data. The public agency methods are simplistic and are based on excluding very high and very low values based on “statistical trimming” or thresholds. One issue is that it is possible to have very high travel times during extreme events. Quality Control procedures that use additional data that have been integrated with travel times would lead to detection of such extreme events. QC procedures for both link-reported travel times as well as vehicle trajectories are needed.

Spatial and Temporal Aggregation Procedures

Once travel time data have been processed through quality control, the data can be aggregated in multiple ways to produce performance measures. Some of the issues are whether to weight the travel times by volume and how to combine data from multiple links into facility or trip travel times. Performance measure calculations are sensitive to the alternate methods.

Network Conflation Method

Integration of travel time data from vendors with public agency data requires reconciling the spatial systems used by each in a process known as conflation. This process is not fully automated and requires considerable judgment on the part of analysts to set boundaries and other parameters. A standard would not only provide needed guidance but would improve accuracy by eliminating guesswork.

Integration of "Congestion Source" Data

Beyond the spatial matching achieved by conflation, data from multiple sources can be used to decompose congestion into individual causes: incidents, weather, physical capacity, work zones, and demand. However, no standard method exists for assigning the shares of congestion caused by its contributing factors. Making this type of assignment is a centerpiece of mobility performance management, as this knowledge will guide program level investments and aid in project level evaluations.

Method of Computing Link Travel Times from Vehicle Trajectories

Private travel time vendors are beginning to provide data on individual vehicle trajectories for a number of reasons. These data can be used to construct origin/destination (O/D) travel times and are thought to provide more accurate travel times on signalized highways. These data have also been available to the vendors and they have used them to develop link-level travel times using proprietary algorithms. However, if an agency wants to develop link-level travel times from trajectory data, a variety of methods can be used. Having a single method would ensure that the base data in subsequent analyses would be consistent.

Bicycle Travel Times

GPS technology is expanding beyond tracking the movement of motorized vehicles to the movement of individuals. This includes tracking individuals' use of multiple modes for a trip. An immediate need is the use of GPS tracking to measure travel times for bicycle trips. Many of the same issues associated with vehicle probe data also apply for bicycle travel times including quality control and aggregation methods.

Core Travel Time Performance Measures for Extended Facilities and Networks

Developing definitions and systematic calculation measures for a core (i.e., limited) set of mobility performance measures would allow cross-agency comparisons and creation of a library of evaluation results. The intent is that agencies are free to use whatever measures they deem to be most appropriate, but when developing performance reports and evaluations they also produce the core measures. This approach is similar to the proposed rulemaking issued by Federal Highway Administration (FHWA) for the MAP-21 performance requirements. The standard would extend beyond the definition and calculation to include standard reporting levels (i.e., ranges).

Metadata Specifications for Travel Time Measure Development

Agencies are beginning to routinely produce mobility performance reports. However, documentation of how the data were processed is usually lacking. Metadata, which is essentially "data about data," describes the content, quality, lineage, organization, availability, and other characteristics of data. Metadata can be used to determine the availability of certain data, to determine the fitness of data for an intended use, and to enhance data analysis and interpretation by creating a better understanding of the data collection and processing procedures. Metadata can also convey how performance measures were defined by the agency.

Challenges

As discussed in Chapter 3, a set of challenges is common to all standardization efforts. In addition, several unique challenges are identified for this proposed standard related to travel time.

The facets of the standard that are related to or close to the data collection process will require vendor participation as changes will affect their business practices. In the special case of developing link-level travel times from vehicle trajectories, a standard process may be unpalatable to vendors, as they have invested much effort into proprietary algorithms and a standard process could stifle further innovation.

Some agencies may be reluctant to change their data processing methods as these are already embedded in their data systems. Changes to software can be costly. This problem can be ameliorated by phasing in the standard as systems undergo routine upgrades (as opposed to upgrades specifically to implement the standard.

Data processing standards such as for conflation and aggregation may be unduly restrictive and not strictly necessary if adequate metadata is supplied. However, if the goal is to enable cross-agency comparisons, metadata alone will be inadequate.

Research Sufficiency

Standard setting is ultimately a consensus-based process, but it is nonetheless informed by research and technical analyses. For the travel time standard bundle, we have identified several areas where additional research is most likely warranted.

Quality Control of Vehicle Probe Data

Research is needed to develop the algorithms for detecting when very low speeds (high travel times) are legitimate. Such an algorithm would consider conditions in previous and succeeding time periods on the link as well as on upstream and downstream links.

Integration of Congestion Source Data to Derive Their Contributions to Congestion

Several studies exist on this topic. A synthesis of these works is required to identify the most appropriate method.

Stakeholders

The stakeholders that need to be involved in the development of this standard are:

- Private vendors of travel time data;
- Performance Measurement system software vendors; and
- State and Metropolitan Planning Organization (MPO) planners and operators who use travel time data for real-time, planning, and performance management applications.

Applications include:

- Providing traveler information;
- Implementing operations control strategies;
- Performance reporting; and
- Travel demand and simulation modeling.

Level of Effort to Develop and Adopt

Once the necessary research is conducted, the tightly focused nature of the standard's topics should allow consensus to be reached with a small to medium effort.

Traffic Incident and Work Zone Standard

Business Case

Several standards have been developed for incidents and work zones, such as the IEEE 1512 family of standards and the so-called Focus State Initiative measures promoted by FHWA. These standards are necessary but not sufficient for conducting operations and planning analyses for performance management. As with travel time, having a consistent set of data definitions and processing procedures will enable agencies to compare results. Incident and work zone standards specifically related to performance management are needed to allow sophisticated and uniform analyses to be conducted. In some cases, the correct form of data to allow analyses is not currently collected by agencies. These data include not only information on the "incident timeline", which tracks major activities during the "life" of an incident, but also information on the severity of the incident in terms of traffic flow (how many lanes and shoulders are blocked and for how long). These data allow much better assessments of incident management practices and provide data for advanced modeling of future conditions.

The research team recognizes that standards can impose undue hardship on agencies especially when data systems must be retrofitted for compliance. The standards recommended here are not intended for that. Rather, as new data systems are developed or procured, it is hoped the new standards will be followed.

Recommended Standard

Although this bundle represents two distinct types of data, the characteristics we wish to standardize are similar. Therefore, it has been bundled as a single standard. The key characteristics missing from existing standards is information about the duration and the extent of the blockage caused by the incident or work zone. The following is a discussion of the various facets of this standard; these are summarized in Table 1.

Detailed data on duration and extent of blockages is needed by type of disruption. When incident or work zone blockage changes during the course of a single event, a new data entry should be made indicating the blocking pattern of the event, i.e., exactly which lanes and shoulders are blocked. Time stamps for each entry are also required. This would allow accurate tracking of the true nature of the blockage. The sequence of blockage can also indicate how well the incident or work zone is being managed. For work zones, the longitudinal extent of the blockage, as well as the lateral extent, should be captured.

Core blockage-related performance measures should be developed. Once the data above are established, a variety of additional performance measures can be derived. As with travel time, a core (limited) set of measures is recommended. In addition to measures specific to agency needs, several measures should be universally created. The concept is that agencies should report these in performance reports and evaluations of completed projects so that other agencies can compare their performance. The measures recommended in this standard are the lane-hours lost index and shoulder-hours lost index. These are the total lane-hours and shoulder-hours lost due to the disruption divided by the original

number of lanes before the disruption occurred. The indices are computed separately for incidents and work zones.

Challenges

Data collection systems would have to be revamped to capture the blockage data at the level of detail indicated by the standard. This requirement may cause some initial reluctance to the standard, but the standard could be phased in whenever systems are upgraded or replaced as part of their normal software cycle.

Research Sufficiency

No additional research is required in order to begin deliberations on this standard.

Stakeholders

Both data users and data owners must be involved as stakeholders:

- Transportation operators responsible for field data collection of incidents and work zones (e.g., traffic management center and construction personnel);
- Researchers who would use the enhanced data in future research; and
- State and MPO planners and operators who use travel time data for real-time, planning, and performance management applications.

Applications include:

- Providing traveler information;
- Implementing operations control strategies;
- Performance reporting; and
- Travel demand and simulation modeling.

Level of Effort to Develop and Adopt

The tightly focused nature of the standard's topics should allow consensus to be reached with a small to medium effort.

Demand Standard

Business Case

New technologies offer the opportunity to increase the amount of demand data available to transportation agencies. Heretofore, demand data has been limited to vehicle counting technology and forecasts from models. Now, the ability to measure personal movement through the transportation system is possible. Currently, private vendors process data from GPS devices and cellphones to develop demand for origin/destination (O/D) pairs using proprietary algorithms. A standard processing method would ensure that data from one vendor is compatible with and can be compared with others. In lieu of a standard for data processing, a standard test method for documenting the accuracy of the vendor-supplied O/D data may be considered. In addition, because raw data may become available from new sources (rather than the third party vendors who process it), a standard would help agencies process the data consistently.

In the near future, we expect the amount of demand/personal movement data to expand significantly. These data will be generated by personal devices and from automotive technologies such as automated and connected vehicles (AV/CV). However, no standards exist related to the use of new technologies to measure demand. Standards are being developed for the communication of raw data from AV/CV but how demand is derived from these measurements has not been explored.

Recommended Standard

This bundle has three distinct components and may be separated as discussed below because they are fundamentally different in concept. The following is a discussion of the three components of the standard.

Demand in Congested Flow Conditions

Once traffic flow has reached capacity, demand in terms of number of vehicles wishing to travel in a given time period cannot be determined by point-based measurement technologies; the demand for the highway section is stored in a queue. The queue represents the demand for the highway section but it will be counted in a successive time period, not the time period of interest. This part of the standard would provide a standard processing method for assigning the portion of the queue to a time period. This component could be included in the travel time bundle because it is likely that travel time or speed data would be used in the method.

Origin/Destination Travel Demand Flows

As discussed above, this standard could take two forms: (1) a standard processing method for converting raw location data for persons and vehicles into O/D flows or (2) a standard test method for assessing the accuracy of pre-processed O/D flows.

Survey Data for Planning and Operations

Although NCHRP Report 571 (Standardized Procedures for Personal Travel Surveys) covers this topic in depth, it does not include the "weight" of a standard. The recommendation of the research team is that the next phase should determine if the procedures in NCHRP Report 571 should be standardized. In addition, the research in NCHRP Report 571 is currently over 10 years old and new survey methods have been in use since then. As pointed out in NCHRP Report 571, the use of GPS and the internet as survey mechanisms is becoming more common, but little to no guidance in these areas is provided. As stated in the report:

"…there is enormous potential for defining standardized procedures and providing guidelines. These may address such issues of how to provide access to websites, the type of graphics and other materials to be provided, building in cross checks on data and cross-referencing travel of other household members, encryption, and a variety of ethical issues that will arise with Internet surveys. As with GPS, however, this area is considered too under-developed for this project."

Therefore, a second avenue for standardization is a standard governing the use of GPS and the internet as a survey mechanism. The NCHRP Report 571 identifies several potential topics for a potential standard.

Challenges

The O/D travel demand flows standard will require vendor participation as changes will affect their business practices. In the special case of developing link-level travel times from vehicle trajectories, a

standard process may be unpalatable to vendors, as they have invested much effort into proprietary algorithms and a standard may stifle further innovation.

Research Sufficiency

Demand in Congested Flow Conditions

The research team believes that adequate past research has been conducted, but these will need to be assimilated before standard deliberation can proceed.

O/D Travel Demand Flows

Original research would need to be conducted on this topic before standard deliberation can proceed.

Survey Data for Planning and Operations

Original research on the use of GPS and the internet needs to be conducted before standard deliberation can proceed.

Stakeholders

Both data users and data owners must be involved as stakeholders.

Demand in Congested Flow Conditions

- Transportation operators responsible for field data collection of incidents and work zones (e.g. traffic management center and construction personnel);
- Researchers who would use the enhanced data in future research; and
- State and MPO planners and operators who use travel time data for real-time, planning, and performance management applications.

O/D Travel Demand Flows

- Vendors currently supplying O/D travel demand data;
- AV/CV researchers;
- State and MPO planners;
- Survey data for planning and operations;
- State and MPO planners;
- Vehicle and GPS technology experts;
- Internet graphic design; and
- Internet survey design (other fields).

Level of Effort to Develop and Adopt

The Demand in Congested Flow Conditions standard can be undertaken with a minimal effort and probably should be added to the Travel Time Bundle. The O/D Travel Demand Flows and Survey Data standards would require a major effort because foundational research needs to be conducted and in the case of O/D Travel Demand Flows, the standard is likely to be contentious because it intrudes on current business practices of vendors.

Transportation Network Standard

Business Case

Currently, a significant amount of effort is required to integrate geospatially-referenced information from different transportation networks. The difficulties that arise in transferring speed, count, safety, or asset information from one network to the other can significantly hamper performance measurement and the development of data analytics and modeling applications. Methods to associate network features and resolve discrepancies exist but they are semi-automated and no guidance exists on how to validate the outcome.

Different geographic networks support applications such as vehicle navigation, performance measurement, travel time reporting, asset management, address geocoding, travel demand modeling, and emergency management. All these networks legitimately define roads differently in terms of geographic detail, roadway length, accuracy, and corresponding features attached to links. The challenge is to avoid duplicity of efforts and to establish a process for data integration that leads to improvements in accuracy, consistency and completeness.

Currently, this problem is partially addressed by conflation, a procedure that transfers geometry and attributes from one network to another. Conflation is a complex Geographic Information System (GIS) process that is not fully automated and results in many-to-many relationships between geographic features that are often a) imprecise b) hard to validate c) require significant amount of manual labor to produce. When network updates occur, as it is often the case, the previously developed many-to-many relationships are partially invalidated and need to be updated. Although several technical reports exist on conflation, the authors are not aware of a methodology that is fully automated and error-free. Research and guidance in this direction can be valuable. Equally important is the development of specific GIS metrics that quantitatively evaluate the quality of network conflation.

It is important to note network conflation will become technically easier but still necessary as more and more transportation networks are equipped with linear referencing systems. Network conflation becomes straightforward when a common base network and LRS is used. Over the last 15 years, Linear Referencing Systems (LRS) have become the method of choice in the transportation GIS community for representing transportation networks. Furthermore, as of 2012, building a LRS system is a formal requirement of HPMS called All Road Network of Linear Referenced Data (ARNOLD). The ARNOLD reference manual provides business rules and nationally endorsed industry-wide standard practices on LRS deployments.

Technically, LRS uses distance from reference points to register network features. Attributes of transportation features are represented as point and linear events along the features. For example, a point event can be an incident or the exact location of a sign. A linear event can demarcate the start and end of a work zone, a roadway with a specific number of lanes, or a roadway with a recently rehabilitated pavement. As a result, a LRS does not have to segment the transportation system into links and nodes based on some logic that can change from application to application. Sharing of transportation data involves transferring the point and linear events located using linear referencing. In contrast, transferring TMC speeds or TIGER data involves conflating nodes, links, and features associated to each link.

Based on the Circular No. A-16 published by Office of Management and Budget (OMB) the development of a digital spatial information resource is a national policy. Even though HPMS only recently requested the reporting and standardization of statewide LRS systems, it is anticipated that adequate resources will be devoted for its successful completion. The OMB directive on which the ARNOLD requirement is based is presented below.

A major objective of this Circular is the eventual development of a national digital spatial information resource, with the involvement of Federal, State, and local governments, and the private sector. This national information resource, linked by criteria and standards, will enable sharing and efficient transfer of spatial data between producers and users. Enhanced coordination will build information partnerships among government institutions and the public and private sectors, avoiding wasteful duplication of effort and ensuring effective and economical management of information resources in meeting essential user requirements.

Recommended Standard

Standards can be provided along two fronts: on the adoption and use of LRS in transportation applications that use only nodes and links and on the conflation process that merges two networks without LRS. For each of the two cases additional information is provided below.

Standards for the Use of LRS in Performance Measurement, Transportation Planning and Modeling

Transportation planners and modelers can update their networks to include a linear referencing system that is compatible with the state-sponsored ARNOLD network. Alternatively, transportation practitioners may choose to derive their network from the state's ARNOLD network. Third party data providers may be requested to report travel times, counts or other statistics using a network that has LRS. This can be proprietary or the state-sponsored ARNOLD network. Roadway lengths corresponding to probe vehicle speeds can be based on TMC link lengths or they can be updated in a way that eliminates the gaps and overlaps between TMC links that exist now. Linear reference will help minimize or eliminate the obstacles in transferring information between the source and destination transportation network.

Standards on Network Conflation

The state of practice for network conflation can be documented and the most effective methodologies can be proposed. However, conflation is a complex GIS procedure that is achieved by dedicated GIS software. Guidance on algorithm selection or algorithm implementation may have limited value to transportation practitioners who are mere users of the association tables that conflation produces. Guidance on validating network conflation may be more appropriate. Currently, many network conflation tasks are only qualitatively validated using spot checks. Benchmarks on how to quantitatively validate the result of conflation will increase transparency and provide quality assurance.

Challenges

State-sponsored LRS networks are in different stages of maturity. The Highway Performance Monitoring System (HPMS), by requiring that all states develop and submit an all-road linear referenced network, is accelerating change in the right direction. Even though an ARNOLD network may not currently exist for all the states, it is expected that data producers and consumers alike will increasingly utilize LRS in the immediate future.

The challenges are significant. From a transportation planning perspective, there are no conflicting interests and the challenges are mainly technical: learning new methods and applying them efficiently, updating existing networks with LRS, or substituting existing networks with more accurate ones derived from the state-sponsored ARNOLD network. The challenges involve building expertise in LRS and using GIS in conjunction with database tools to analyze transportation data. Transportation modeling software need not change internal structures that are based on nodes and links to calculate routes. However,

additional information about the linear referencing and the common reference points need to be stored by existing software.

Research Sufficiency

In the last 25 years, as the ARNOLD reference manual states, there has been a rich body of work on LRS systems in Geographic Information Systems for Transportation (GIS-T). Substantially less effort has been devoted to network conflation. Nevertheless, a critical body of work already exists to identify best practices.

Stakeholders

Network data integration is an ongoing challenge. Transportation practitioners will welcome advice and guidance that will remove obstacles and streamline data exchanges that currently take a disproportional amount of time. A critical mass of advocates is not required. Consensus is not necessary since the issue at hand is not the development or adoption of a new data format.

Level of Effort to Develop and Adopt

Multidisciplinary expertise will be required to build simple and effective guidance.

Using GTFS in Transportation Planning

Business Case

The General Transit Feed Specification (GTFS) has been universally adopted by transit agencies as the de-facto standard to publish transit route and schedule information. These data driven route planning websites and mobile apps are used by millions of transit travelers every day. To serve the public, transit agencies devote significant resources to develop, publish and update the GTFS data so that the latter reflect the actual service that is being offered.

Demand forecasting, in contrast, has not taken advantage of GTFS. Instead, travel demand forecasting models rely on legacy transit information that has been assembled manually and in layers over many years. As a result, the transit information may not be as accurate or up-to-date with what is contained in GTFS. Support for GTFS in demand forecasting software varies. To the author's knowledge, only one software package supports reading GTFS files. Although there is significant interest and GTFS is universally recognized as the way of the future, there is no guidance for its adoption.

Recommended Standard

Using GTFS will simplify and expedite the network development process. Furthermore, providing guidance on how to incorporate GTFS into travel demand forecasting will improve accuracy and will ensure that the published level of service is the same as the modeled one. GTFS now stands as the most widely available sources of transit schedule information. As a result, its adoption in planning and modeling is inevitable. Providing guidance will accelerate adoption and help practitioners apply certain necessary quality checks. At a minimum, the guidance should provide information on the following topics:

- How to quantitatively verify that a particular route alignment is the same between GTFS and the planning network;
- How to verify that stop locations in GTFS have been imported with adequate accuracy;
- How to calculate headways and frequencies from schedules; abd
- How complex transit fare schemes should be entered or simplified into planning software.

Challenges

The challenges in providing guidance on GTFS are mainly technical but they are moderate and manageable. The primary challenge involves the conflation of the GTFS network that accurately follows the physical roadway infrastructure with the links included in the transportation-modeling network that may represent the transportation network less accurately.

GTFS, by design, is not based on any particular node-link network. Instead, it contains route coordinates in latitude-longitude format. Such a design allows GTFS to be used with all types of networks. However, when importing route alignments to the planning software, route coordinates need to be converted to sequences of nodes. Given the fact that transportation networks do not always accurately follow the physical roadway network it is very frequent that GTFS routes cannot be transferred without approximations and compromises. The guidance could provide advice on how to recognize and resolve this issue.

There are no incentives to using an alternative format or a particular methodology. Travel demand software will be updated to include support for GTFS. Currently, to the authors' knowledge, only one travel demand software includes GTFS support. The developed guidance will be targeted to the transportation modeler. Software vendors will be free to adopt their own methodology and algorithms to support GTFS. The guidance could provide quality metrics to transportation practitioners to judge how each route has been imported to the transportation model. This will increase the transparency and provide quality assurances.

Research Sufficiency

There is sufficient research in network conflation/mapping techniques in the last 20 years. The technology and methods are well documented; however, the underlying data structure is the problem. Therefore, a series of steps for an analyst to follow would be useful.

Stakeholders

Transportation practitioners are expected to welcome the adoption of GTFS, which will free them from significant manual network editing work. Software vendors have already started supporting GTFS.

Level of Effort to Develop and Adopt

It is suggested that transportation agencies, transportation modelers, and software providers are actively consulted. Consensus on the quality measures on network conversion would be welcome and should be actively sought. Since this is strictly a technical issue similar to the selection of a particular performance measure, it is not anticipated that significant disagreements between stakeholders will emerge.

Standard Development Steps

The previous section in this report outlined several areas (i.e., travel time, travel demand, incidents and work zones, network attributes and transit) that could benefit from further development of standards or guidance. This section describes the steps forward—essentially, the "how to further develop" and meet the needs for additional guidance or standards.

Figure 6 illustrates the proposed approach for making progress in each of the identified data areas. These steps are described in more detail in the following paragraphs.

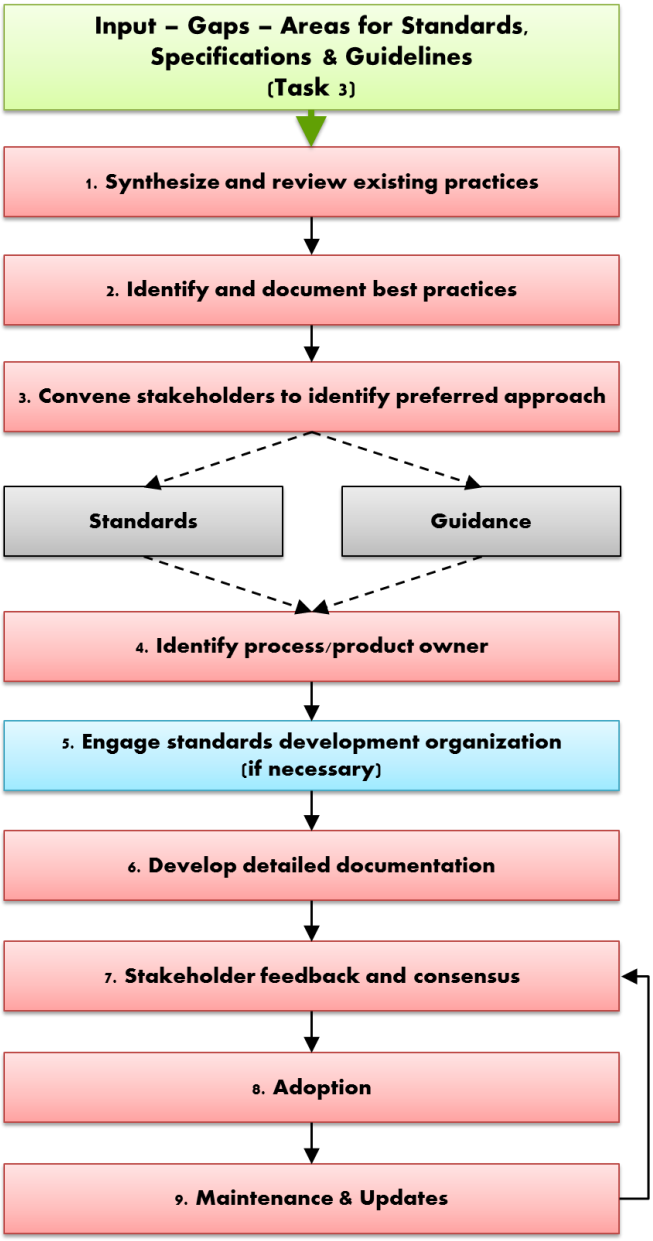


Figure 6: Proposed Approach for Further Development

1. Synthesize and review existing practices

The first step is compiling and reviewing existing practices for each data area. The existing practices could be those used by public agencies or private data providers. In some cases, a variety of approaches are used, although these approaches may have been applied in different contexts.

When compiling existing practices, it will be important to document several elements:

- Context: Who applied the practice, and to solve what problem?
- Extent: How many agencies/companies applied the practice?
- Complexity: How complex is the approach as compared to professional capacity?
- Results: Have agencies/companies achieved benefits from applying the practice?
- Competing practices: Is there more than one satisfactory approach?

The answers to these questions will help guide decisions in upcoming steps when identifying a best practice, determining preferred standardization approach, and in identifying a process/product owner.

2. Identify and document best practice(s)

Once existing practices have been reviewed, the next step is to identify and document best practice(s). Depending upon the number and quality of existing practices, there could be multiple variations in this step, as follows:

1. Multiple existing practices, but clear indication of single best practice – this variation is the simplest to advance, as there is already informal industry consensus (determined by number of implementations) about which of multiple existing practices is the best practice.
2. Multiple existing practices, several equally competitive practices – this variation includes several practices that have seen similar levels of implementation, and no one practice stands out as superior to others. In this case, the several equally competitive practices should be carried forward to the next step, to be considered by stakeholders.
3. One or more existing practices, but none deemed mature or satisfactory – this variation may be uncommon but still possible, in that a practice has been adopted by one but is not considered satisfactory to any others. This could happen in the early stages of problem solving, and generally indicates that a process solution is not in a mature stage.
4. No existing practices, but one or more possible practices – this variation occurs when no practice has been applied, but there are several possible solutions that could be tested. Like the previous variation (#3), this generally indicates that a process solution is in the early stages.

It should be noted that the intent of this step is not to pick a “final winner” to standardize, but to pick those possible practices/solutions with the most potential. If there is already a single best practice (variation #1 above), then yes, that single best practice can be advanced. However, if there are several competing alternatives either in practice (variation #2) or theory (variations #3 or #4), then those competing alternatives should be documented and vetted among stakeholders in the next step.

3. Convene stakeholders to identify preferred approach

Once completing alternatives (either in practice or in theory) have been identified, the next step is to convene relevant stakeholders to identify a preferred approach (e.g., standard or guidance). The stakeholder groups are likely to vary by data area, although there may be a few stakeholders in common across all data areas.

The stakeholders should be guided to consider several factors when identifying a preferred approach:

- Business case – Is the problem clear and pervasive among many agencies, and more likely to be solved by standards development instead of issuing guidance?
- Simplicity to comply – Is there a clear and simple path to implementation?
- Data/software provider support – Is the commercial marketplace healthy and receptive to standardization?
- Alternative of competing formats – Will standardization improve data portability and accessibility?
- Level of effort to develop standard – Is the level of effort required to develop a standard commensurate with the level of benefits once the standard is implemented?

4. Identify process/product owner

Once (or if) stakeholders identify a preferred approach, the next step is to identify a process/product owner who can oversee responsibility for the remaining steps in the process. This owner typically sees the greatest benefit from standard development and adoption. Alternatively, the owner may have domain authority within this particular area, or may have legislatively mandated authority to develop and implement standards within their responsible domain. Depending upon these and other factors, the process/product owner could be either a public or a private entity.

5. Engage standards development organization (if necessary)

If a formal specification is the preferred approach, then the process owner will need to engage a standards development organization (SDO). There are several possible SDOs, depending on the data area:

- American National Standards Institute (ANSI);
- American Society for Testing and Materials (ASTM) International;
- Institute of Electrical and Electronics Engineers (IEEE);
- International Organization for Standardization (ISO); and
- Society of Automotive Engineers (SAE) International.

Each SDO has certain procedures in place to guide every phase of standards development: engaging stakeholders (public and private entities), facilitating discussions, develop technical documents, gaining consensus, and publishing and maintaining standards.

6. Develop detailed documentation

The next step in the process is the development of detailed technical documentation, regardless of whether formal standardization is pursued or not. In some cases, stakeholders within the process may draft this technical documentation, particularly if they are basing it on existing practices that may already be documented. In other cases when practices are not already clearly documented, it may be necessary to use a qualified consultant to develop detailed documentation. Otherwise, development could proceed at a slow pace if there is not a significant time incentive and all contributions are voluntary.

7. Stakeholder feedback and consensus

After detailed documentation has been developed, stakeholder feedback and consensus is necessary to move forward. This step is also necessary regardless of whether formal standardization is pursued or not. If a formal standard is being developed, the SDO often has an established review process for standards that are in development.

8. Adoption

This step of formal adoption is typically necessary only if a formal standard is being pursued. If detailed standards or guidance is being developed, then the adoption is essentially the owner making the technical documentation publicly available. As with other steps in formal standards, the SDO typically has an established process for official adoption of each standard.

9. Maintenance

Most SDOs have a process for standards maintenance and updates. This maintenance process ensures that a published standard is still needed and up-to-date. For example, ASTM International requires that all published standards be re-balloted on a regular basis. The standard owner should be aware of this maintenance process, as they are likely to have responsibility for convening stakeholders in any updates or reballoting.

CHAPTER 4

Proposed Research Statement

I. PROBLEM NUMBER

2017-xxx

II. PROBLEM TITLE

Developing Data Standards and Guidance for Transportation Planning and Traffic Operations – Phase 1

III. RESEARCH PROBLEM STATEMENT

Planning and operating transportation systems involve the exchange of large volumes of data. The lack of common data formats has been a limiting factor for transportation agencies and all practitioners involved in data analysis and reporting. Well-designed data standards can be a viable solution to this problem since they improve the efficiency of data-driven processes and can support innovation.

NCHRP 8-36 Task 129 examined the feasibility of developing standards for transportation planning and traffic operations. The report revealed that it is difficult to predict standard adoption. There are many well-designed and technically superior standards that that have failed and become marketing case studies. Based on the research, a business case and clear incentives for a critical mass of supportive stakeholders is required for market adoption.

Transportation data standards and guidance for planning and operations are feasible and desirable.

Standards are most successful if they:

- Have a clear business purpose;
- Are clear in application, specificity and versioning;
- Are developed with broad outreach and buy-in;
- Are well defined and simple;
- Are open standards;
- Are forward looking; and
- Involve a national or worldwide community.

There are also many challenges associated with standard development. These include:

- Reluctant data vendors;
- Dynamic data content - Standard may be outdated soon after it release;
- Complexity of data to be standardized;
- Standardization process takes too long to complete;

- Standardization process does not take into account a critical mass of would be users or decision makers;
- There are significant disincentives or conflicts of interest;
- Limited outreach - Agencies may be unaware of the benefits of adoption; and
- Inadequate resources to overcome the barrier of entry.

The report concluded that standards are feasible and desired. Five specific areas or "bundles" of standards were identified to be ripe for standardization. They are travel time, demand, incident and work zones, network and transit.

IV. LITERATURE SEARCH SUMMARY

The following are reports and research that have addressed national-level data standards issues.

- NCHRP 8-36 Task 129 – Scoping Study to Establish Standards and Guidance for Transportation Planning and Traffic Operations – Also see Literature review for this report.
- TRB Special Report 304, How We Travel, which defines a sustainable approach to national travel data monitoring.
- The literature review accomplished as part of NCHRP 8-36 Task 129 is comprehensive and should be referenced.

V. RESEARCH OBJECTIVE

The objective of this study is to prioritize transportation planning and traffic operations standard areas and develop standards and/or guidance to be used and adopted by the transportation community.

The five standard data areas or "bundles" to be evaluated for further standard development include travel time, demand, incident and work zones, network and transit. Each bundle includes various data categories. The researcher may evaluate bundles or data categories as appropriate for moving to the next stage of standard development.

Direct reference must be made to NCHRP 8-36, Task 129.

Note that a "standard" is a method or process that:

- Has a legal basis and can be made mandatory;
- Is a normative document from a formal standard body;
- Is passed through a full and open consensus process; and
- Is implemented on a national level and there is a strong mandate to apply them.

Guidance includes methods and documentation of processes that are informative and do not necessarily require consensus.

Proposed tasks include the following:

1. Develop a comprehensive stakeholder outreach plan.

2. Prioritize and select for further standard development – One or more of the bundles described in NCHRP 8-36 Task 129 (Chapter 3) (depending on available funding).

3. Follow steps identified in NCHRP Task 8-36 Task 129 Standards Development Steps (provide specific reference) to develop each standard.

4. Summarize lessons learned.

5. Develop Phase 1 of Transportation Standards and Guidance.

6. Provide recommendations for the next phase of standards and/or guidance development.

VI. ESTIMATE OF PROBLEM FUNDING AND RESEARCH PERIOD

Recommended Funding:

$250,000 per year

Research Period:

24 months

VII. URGENCY, PAYOFF POTENTIAL, AND IMPLEMENTATION

This research project was developed by NCHRP Task 8-36 Task 129. States and MPOs are currently faced with a plethora of issues related to data compilation for MAP-21. Standards will be extremely useful in resolving many of the issues related to system data integration, compilation and reporting.

This project was identified as a High Priority project by SCOP membership because of the current MAP-21 performance management requirements and concerns over broader data collection, management and analysis issues.

The potential benefits of this research include the following:

1. Provide a framework for a phased approach to standard/guidance development for transportation planning and traffic operations.
2. Develop at least one bundle of standards or guidance as deemed high priority by the transportation community.
3. Provide feedback on lessons learned for future standard and guidance development.

References

Butler, J. Allison. Designing geodatabases for transportation. ESRI, Inc., 2008.

Cambridge Systematics (2008), "XML Schemas for Exchange of Transportation Data", Final Report, Prepared for the National Cooperative Highway Research Program Transportation Research Board National Research Council

Cambridge Systematics et al., *Analytic Procedures for Determining the Impacts of Reliability Mitigation Strategies*, SHRP 2 Project L03, Transportation Research Board, 2013, http://onlinepubs.trb.org/onlinepubs/shrp2/SHRP2_S2-L03-RR-1.pdf

Catala, Martin, Samuel Dowling, and Donald Hayward. Expanding the google transit feed specification to support operations and planning. No. FDOT BDK85# 977-15. 2011.

Chandler, Brian E. et al., *Signalized Intersections Informational Guide, Second Edition*, FHWA, 2013, http://safety.fhwa.dot.gov/intersection/conventional/signalized/fhwasa13027/form.pdf

Chopra, Aneesh. Innovative state: How new technologies can transform government. Grove/Atlantic, Inc., 2014.

Dyson, Lauren, ed. Beyond Transparency: Open Data and the Future of Civic Innovation. Code for America Press, 2013.

http://2isfo.eng.hawaii.edu/Presentations/Session%2032/32%20-%20Li.pdf
http://bigdatawg.nist.gov/_uploadfiles/NIST.SP.1500-1.pdf
http://inspire.ec.europa.eu/
http://inspire.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_TN_v3.1.pdf
http://inspire.ec.europa.eu/index.cfm/newsid/11883
http://onlinepubs.trb.org/onlinepubs/circulars/ec185.pdf
http://onlinepubs.trb.org/onlinepubs/conferences/2014/ITM/Presentations/Wednesday/AllAboutMatrices/Stabler.pdf
http://onlinepubs.trb.org/onlinepubs/ncfrp/ncfrp_rpt_009.pdf
http://onlinepubs.trb.org/onlinepubs/nchrp/nchrp_rpt_571.pdf
http://ops.fhwa.dot.gov/publications/fhwahop13047/index.htm
http://stateofthemap.us/osm-and-arnold/
http://tfresource.org/Open_Matrix_Format
http://tisa.org/technologies/tmc/, http://wiki.openstreetmap.org/wiki/TMC
http://user47094.vs.easily.co.uk/siri/
http://user47094.vs.easily.co.uk/siri/
http://welcometobaltimorehon.com/the-1904-fire-and-the-baltimore-standard
http://wiki.openstreetmap.org/wiki/OSM_XML
http://www.aptatcip.com/
http://www.astm.org/SNEWS/OCTOBER_2000/oct_howto.html
http://www.bpmn.org/
http://www.fgdc.gov/standards/process
http://www.fhwa.dot.gov/policyinformation/hpms/documents/arnold_reference_manual_2014.pdf
http://www.fta.dot.gov/grants/15682.html
http://www.gsdi.org/
http://www.gtfs-data-exchange.com/
http://www.iso.org/iso/big_data_report-jtc1.pdf
http://www.iso.org/iso/bottom_line.pdf

http://www.iso.org/iso/home/standards.htm
http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=54610
http://www.isotc211.org/hmmg/HTML/
http://www.ite.org/standards/tmdd/
http://www.nextbus.com/xmlFeedDocs/NextBusXMLFeed.pdf
http://www.omg.org/
http://www.opengeospatial.org
http://www.opengeospatial.org
http://www.openlr.org/
http://www.rasch.org/rmt/rmt151h.htm
http://www.teradatamagazine.com/v11n01/Features/Big-Data/
http://www.travelsurveymanual.org/
http://www.w3.org/2004/copras/docu/D15.html#benefits
https://developers.google.com/transit/gtfs-realtime/
https://www.fgdc.gov/
https://www.fgdc.gov/
https://www.fhwa.dot.gov/policyinformation/hpms/documents/arnold_reference_manual_2014.pdf
https://www.fhwa.dot.gov/policyinformation/hpms/documents/arnold_reference_manual_2014.pdf
https://www.oasis-open.org/
https://www.openstreetmap.org/
https://www.r-project.org/
https://www.standards.its.dot.gov/LearnAboutStandards/ApplicationAreas
https://www.waze.com/
https://www.whitehouse.gov/omb/circulars_a119
https://www.whitehouse.gov/sites/default/files/omb/memoranda/2012/m-12-08.pdf

Li, Honglong Li, Wellander, Chris, and Shenghong, L., *Freeway Traffic Demand Under Oversaturated Conditions*, presented at 2nd International Symposium on Freeway and Tollway Operations, June 2009,

Manyika, James, et al. "Big data: The next frontier for innovation, competition, and productivity." (2011).

National Cooperative Highway Research Program, et al. Travel Demand Forecasting: Parameters and Techniques. Vol. 716. Transportation Research Board, 2012.

Segaran, Toby, and Jeff Hammerbacher. Beautiful data: the stories behind elegant data solutions. O'Reilly Media, Inc., 2009.

Acronyms and Abbreviations

ADA	Americans Disabilities Act
ANSI	American National Standards Institute
ARNOLD	All Roads Network Linear Referenced Data
ASTM	American Society of Testing and Materials
AV/CV	Automated Vehicles/Connected Vehicles
AVL	Automated Vehicle Locator
BMC	Baltimore Metropolitan Council
CAD	Computer Aided Dispatch
CVS	Comma Separated Values
CWM	Common Warehouse Meta-data
DOT	Department of Transportation
DSRC	Dedicated Short-range Communications
DXFS	Data Exchange Format Specification
EU	European Union
FAF3	Freight Analysis Framework Version 3
FGDC	Federal Geographic Data Committee
FHWA	Federal Highway Administration
FTA	Federal Transit Administration
GDF	Geographic Data Format
GIS	Geographic Information System
GIS-T	Geographic Information Systems for Transportation
GPS	Global Positioning System
GSDI	Global Spatial Data Infrastructure
GTFS	General Transit Feed Specification
HDF	Hierarchical Data Format
HPMS	Highway Performance Monitoring System
HTML	Hyper Text Markup Language
IBM	International Business Machines
IEEE	Institute of Electrical and Electronics Engineers
INSPIRE	Infrastructure for Spatial Information in the European Community
ISO	International Standards Organization
ITS	Intelligent Transportation System
LOS	Level of Service
LR	Linear Referencing
LRS	Linear Referencing System
MIRE	Model Inventory of Roadway Elements
MMUCC	Model Minimum Uniform Crash Criteria Guideline
MPO	Metropolitan Planning Organization
NBS	National Bureau of Standards
NCHRP	National Cooperative Highway Research Program
NEMSIS	National Emergency Management System Information System

NHS	National Highway Safety
NIST	National Institute of Standards and Technology
NPMRDS	National Performance Management Research Dataset
NYMTC	New York Metropolitan Transportation Council
O/D	Origin/Destination
OASIS	Organization for the Advancement of Structured Standards
OMB	Office of Management and Budget
OMG	Object Management Group
OMX	Open Matrix Format
QC	Quality Control
SAE	Society of Automotive Engineers International
SDO	Standards Development Organization
SIRI	Service Interface for Real Time Information
SOV	Single Occupant Vehicle
SQL	Structured Query Language
TCIP	Transit Communications Interface Protocol
TMC	Traffic Message Chanel
TMC	Transportation Management Center
TMDD	Traffic Management Data Dictionary
TRB	Transportation Research Board
U.S. DOT	United State Department of Transportation
UML	Unified Modeling Language
UNETRANS	Unified NETwork for Transportation
US	United States
VMT	Vehicle Miles Traveled
W3C	World Wide Web Consortium
XML	eXtensible Markup Language

Appendix A - Project Presentation Slides

NCHRP

NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM

NCHRP 8-36 Task 129
Scoping study to establish standards and guidance for data used for transportation planning and traffic operations purposes

PROJECT OBJECTIVES

Agenda

- Project Objectives
- Task 2 – Literature Review
- Task 3 – Feasibility of Standards
- Task 4 – Draft Research Statement

Project Objectives

1. Examine the feasibility of conducting research that will produce potential data standards and guidance

2. Create a research problem statement for introducing data standards in specific fields mature for standardization

Project Overview

1. Document current literature on data standards and guidance
2. Assess opportunities and challenges for establishing standards and guidance
3. Form a problem statement to support a larger research effort for producing standards and guidance for the collection and maintenance of data

Types of Standards

Formal or *de jure* standards	Legal basis and can be made mandatory
	Normative documents from formal standards bodies
	Passed through a full and open consensus process
	Implemented on a national level and there is strong mandate to apply them. Considerable time (up to 4 years) may be needed for completing the full approval process

TASK 2 - LITERATURE REVIEW

Types of Standards

Technical or industry specifications	Based on consensus among members of standards bodies, consortia or trade organizations
	No formal character or legal basis
	When widely accepted and used in practice = ***de facto* standards**
Guidelines, conformance, or reference implementations	Informative and do not require consensus

Task 2: Literature Review

- Best practices
- Challenges
- Survey of current practice
- Short telephone interviews
- Annotated bibliography, key findings

7

Steps for Standard Development

1. A **market need** for a new standard or standardization activity identified and recognized
2. **Set of requirements** drafted (commercial, user or functional requirements)
3. **Specification drafted** by a group of experts - Based on a group of technical experts
4. **Formal approval** process
5. **Testing** or (self-) certification by the industry, may also encompass developing implementation guidelines
6. **Maintenance or periodic review process** to ensure the standard will remain in sync with market requirements

Representative Standards

- FHWA's HPMS:
  - » Successful.
  - » Mandated by Government
- NCHRP's TransXML:
  - » Well designed.
  - » Not frequently used
- Open Matrix Format:
  - » Independently proposed by interested practitioners
- ISO's and many others Network Formats that have not been adopted in our industry

Transportation Network Standards

Network Format	Description
OpenStreetMap	Crowd-sourced detailed network information for the entire world
ARNOLD	Federally supported effort to combine all public roads into a single, publicly available, authoritative spatial dataset
UNETRANS	Unified NETwork for Transportation is a geodatabase network model that stores assets, incidents, activities, routes and location referencing, mobile objects, and relationships
ISO 19148	Linear Referencing Standard
Traffic Message Chanel (TMC)	Network partition used to define NPMRDS travel time information. Travel time information providers produce TMC networks that may not be fully compatible
INSPIRE D2.8.I.7	Highly detailed European data standard on Transportation Networks
Geographic Data Format (GDF)	ISO International standard that is used to model, describe, and transfer road networks and other geographic data
OpenLR	Open dynamic location referencing on transportation networks supported by Tom Tom
TransCad, EMME, VISUM, AIMSUN, VISSIM, Dynameq, TransModeler, Cube	All transportation software packages have their own definition for transportation networks. Import/Export features exist but they cover only a small subset of available formats

Representative Standards

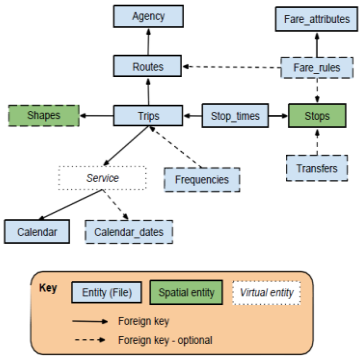
- The Object Management Group (OMG)
  - » International, open -membership, not-for-profit technology standards consortium – The Unified Modeling Language (UML).
- The Federal Geographic Data Committee (FGDC)
- Infrastructure for Spatial Information in the European Community (INSPIRE)
- The Organization for the Advancement of Structured Information Standards (OASIS)
  - » Nonprofit consortium that drives the development, convergence and adoption of open standards for the global information society
- The Organization for the Advancement of Structured Information Standards (OASIS)
  - » Nonprofit consortium that drives the development, convergence and adoption of open standards for the global information society

Examples of Guidance

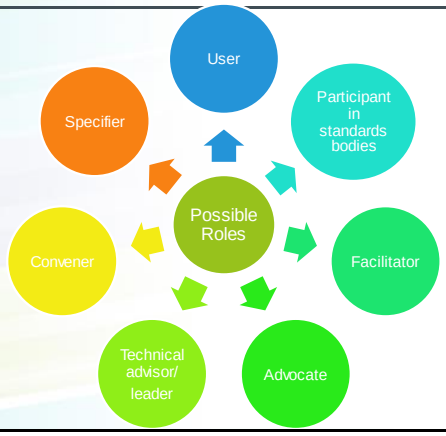
- NCHRP 666 – Target Setting Methods and Data Management to Support Performance Based Resource Allocation
- NCHRP 8-36 (100) – Data Assessment
- FHWA Office of Safety – Data Business Planning Guide
- FHWA Office of Operations – Data Business Planning Guide

Representative Standards

- TriMet General Transit Feed Specification (GTFS):
  - » De-facto standard that prevaile competing efforts from governr



Government Role In Standardization



Government Role In Standardization

Option 1: Help build **Voluntary Consensus** on a pre-existing or new specification that practitioners will design

Option 2: Mandate or design a new standard

Feasibility of Data Standards

- Task 3A
  - » Define performance measures and data users and uses
- Task 3B
  - » Identify categories of possible standards
- Task 3C
  - » Document challenges
- Task 3D
  - » Assess ability to overcome challenges
- Task 3E
  - » Analyze likelihood of adoption

Success Factors

- Clear business purpose
- Clarity in application, specificity and versioning
- Developed with broad outreach and buy-in
- Well defined and simple
- Open standard
- Forward looking
- Involve National or Worldwide community

Task 3A - National Goal Areas

- Safety
- Infrastructure condition
- Congestion reduction
- System reliability
- Freight movement and economic vitality
- Economic sustainability
- Reduced project delivery details

TASK 3 – FEASIBILITY OF STANDARDS

Task 3A – Data Categories

- Mobility
  - » Travel times (links, probe vehicles, OD)
  - » Demand/Volumes (by Auto, Transit, Bike, Peds)
  - » Incidents
  - » Work zones
- Networks

22

Task 3A - Example Performance Measures

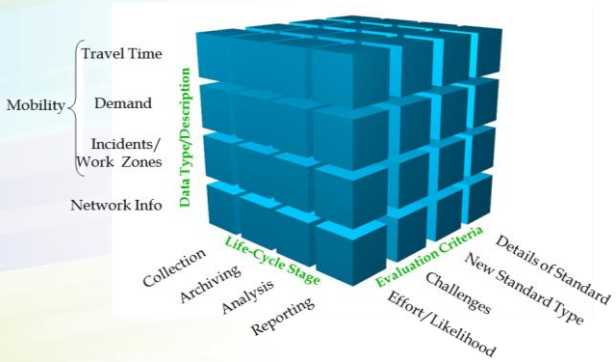
- travel time reliability
- delay
- accessibility to destinations
- travel time index
- incident response time
- time to clear incident
- number of crashes
- highway lane miles
- percent of road miles with low speed limits
- number of hours of road closures, transit headway
- delay
- waiting time
- vehicle counts
- VMT
- VHT

Core Data Areas

- Travel Time
- Travel Demand
- Incident
- Work Zone
- Network

26

Task 3B - Framework



Task 3C - Challenges

- Reluctant data vendors
- Dynamic data content - Standard may be outdated soon after it release
- Complexity of data to be standardized
- Standardization process takes too long to complete
- Standardization process does not take into account a critical mass of would be users or decision makers
- There are significant disincentives or conflicts of interest
- Limited outreach - Agencies may be unaware of the benefits of adoption
- Inadequate resources to overcome the barrier of entry

27

Task 3B - Standardizing Travel Times

Life Cycle	Type	Type of New Standard	Details	Challenges	Effort or Likelihood
Collection Auto		Specification	Quality indicators, confidence intervals	Reluctant data providers	Government mandate
Collection Auto	TMC Links	Revise TMC Specification	New roadway endpoints	Vendor resistance	Moderate
Collection Bike		Guidance	Collection methods	Limited data availability	Limited data applicability
Archiving for All modes		Specification	Quality control	?	?
Archiving for All modes		Guidance	Conflation methods	Reliable conflation methods	Moderate effort
Analysis for All modes		Guidance	Spatial & temporal aggregation	Minor effort	Substantial research exists
Analysis for All modes		Guidance	Integration with congestion sources	Research needed	Intermediate effort

Standardization Analysis

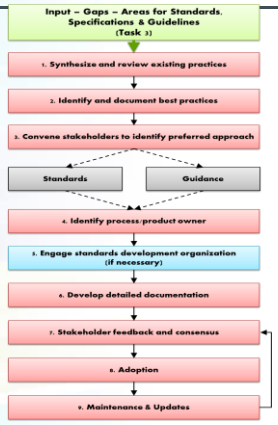
- "Bundles" ripe for standardization:
  - Travel Time
  - Demand
  - Incident
  - Work Zones
  - Network
  - Transit
- For each bundle, the following was assessed:
  - Business Case
  - Recommended Standards
  - Challenges
  - Research Sufficiency
  - Stakeholders
  - Level of Effort to Adopt

28

Outcome of Task 3

- Specific applications areas or categories that need guidelines or standards
- Criteria for evaluating the adoption of new standards
- Challenges in adopting new standards and likelihood of adoption
- The input of Task 3 will inform the research statement of Task 4

Standard Development Steps



32

TASK 4 – RESEARCH STATEMENT

Research Problem Statement

- Intent: Develop an RPS in sufficient detail so that an in-depth study can be conducted on the most critical data requiring standardization
- Current study will define the need, benefit, and barriers for standards for specific data types (no need to cover again in next study)
- Content:
  - Identify specific data types and what aspects of those data that need to be standardized
  - Bundle data types if manageable (e.g., all forms of travel time data)
  - Identify sources of technical information for each data type that would provide the basis for setting a standard
  - Stakeholder outreach to gather information at the formative stage and for reaction to recommended standard

Task 4: Research Problem Statement

- Describe:
  - Types of standards and guidelines
  - Types of users and data collectors
- The statement will include:
  - Desired outcomes
  - Possible tasks
  - Timeframe and resource requirements
- Deliverable: Complete research problem statement

31

Research Problem Statement

- Problem Title
  - Developing Data Standards and Guidance for Transportation Planning and Traffic Operations – Phase 1
- Objective
  - To prioritize transportation planning and traffic operations standard areas and develop standards and/or guidance to be used and adopted by the transportation community
- Recommended Funding
  - $250,000 per year
- Research Period
  - 24 months

34

Tasks in Research Plan

- 1. Develop a comprehensive stakeholder outreach plan
- 2. Prioritize and select for further standard development – One or more of the bundles described in NCHRP 8-36 Task 129 (provide specific reference)(depending on available funding)
- 3. Follow steps identified in NCHRP Task 8-36 Task 129 Standards Development Steps (provide specific reference) to develop each standard.
- 4. Summarize lessons learned
- 5. Develop Phase 1 of Transportation Standards and Guidance
- 6. Provide recommendations for the next phase of standards and/or guidance development.

35

Anita Vandervalk
Avandervalk@camsys.Com

Larry Goldstein
Lgoldstein@nas.Edu

Benefits of Potential Research

- Provide a framework for a phased approach to standards/guidance development for transportation planning and traffic operations
- Develop at least one bundle of standards or guidance as deemed high priority by the transportation community
- Provide feedback on lessons learned for future standard and guidance development

36