NCHRP 08-119 Data Integration, Sharing, and Management for Transportation Planning and Traffic Operations

Phase I Report – Executive Summary

April 16, 2021

Introduction

To effectively plan and operate transportation systems, a wide variety of data must be leveraged. However, existing data silos; a lack of common data formats; hesitation or fear of sharing data; and traditional, rigid data management practices limit transportation agencies and practitioners in their ability to exchange, integrate, and use data for analysis, reporting, and decision-making. Further, as transportation agencies begin to use new data sources and to make their data more accessible, some of the issues faced include gaining confidence in the data and understanding their quality and variability, reconciling data discrepancies among different sources (including probe vehicles and connected vehicles), validating processes for turning data into information, comparing system performance across time taking into account how the data have changed, and understanding the implications of big data approaches/applications. More modern approaches to data management can improve the efficiency of data-driven processes as well as support innovation across transportation agencies.

NCHRP 08-119 Project Objectives

The objective of this research is to develop tools, methods, and guidance for improving data sharing, integration, and management practices to enable transportation agencies, in collaboration with private-sector and public-sector stakeholders, to make better planning and operations decisions. The research will uncover and share:

- Best practices for data sharing, integration, and management
- Emerging practices for data sharing, integration, and management
- Lessons learned associated with data sharing, integration, and management
- Challenges and opportunities associated with data sharing, integration, and management
- Opportunities and information on partnering with private sector data providers
- Needs/approaches to prepare for new issues such as big data

The period of performance for this project runs from September 2019 through September 2022. However, due to the rapidly changing and advancing nature of data, data management, and applications of data within and beyond the transportation industry, the panel directing this research wishes to share information and products throughout the research period of performance to allow transportation agencies to take advantage of the information sooner rather than later. As such, this document provides a summary of the major findings of Phase I of the research, which was conducted between September 2019 and September 2020, as well as plans for developing new tools, methods, and guidance selected by the project panel as a result of the Phase I findings.

Research Approach

Phase I of the research involved gathering information from a wide variety of sources, developing representative transportation planning and operations use cases, assessing challenges and needs, developing a draft data catalog, and identifying potential products for development in subsequent phases of the research. Information gathering involved three primary sources:

- Relevant literature regarding data sharing, integration, and management – the team reviewed over 100 NCHRP reports, Federal Highway Administration (FHWA) reports, state and local reports and white papers, journal articles, web articles, press releases, blogs, etc. across a broad range of transportation operations and planning areas.
- Relevant data standards – the team reviewed 22 standards pertaining to areas data exchange and communications in ITS, mobility, management, work zone, and transit.
- Interviews with public and private stakeholders – the team conducted telephone interviews with individuals representing a wide range of stakeholders, including city, county, and state transportation agencies; metropolitan planning organizations (MPOs); FHWA; transportation consultants; software companies; transportation network companies; data providers; transportation coalitions; and academics.

Initial use case categories, on which to focus subsequent product development, were selected based on information in the NCHRP 08-119 RFP and the experience/expertise of the research team members. As information was gathered and reviewed and telephone interviews were conducted, the research team documented challenges, assessed needs, and identified opportunities associated with data sharing, integration, and management and was able to narrow in on more specific use cases. In the end, 10 use cases for the improvement of data sharing, integration, and management were identified:

1. Integrating work zone event data (WZED) with other data sources
2. TIM performance measurement
3. Managing big data from emerging technologies
4. Next generation ICM
5. Using data for effective operations and enforcement of private shared mobility services
6. Improving transportation planning with emerging data
7. Freight data for planning and operations
8. Integrating ITS transit data for planning and operations
9. Modernization of ATDM Approaches
10. Preparing agencies for smart cities

The findings, including the state of practice, challenges, and existing products to support transportation agencies, are summarized for each of these use case areas in the following section.

Phase I Findings

Integrating Work Zone Event Data (WZED) with Other Data Sources

This use case focuses on the enhancement of work zone data event data (WZED) through the integration of these data and other data sources. As part of a larger effort by the FHWA to standardize and modernize WZED, the Work Zone Data Exchange specification (WZDx) was developed to provide a reference guide or standard method of digitally communicating work zone activities. To date, there have been three versions of the specification, with version 3.0 having been released in September 2020 (1). In April 2021, version 3.1 was developed to enhance and clean up the specification of version 3.0; however, no major changes were introduced (2).

Developing a standard definition for communicating work zone information digitally is essential to the increasing demands of agencies to become more data-driven with their operations and maintenance activities. However, due to a variety of reasons, only a few agencies have implemented the work zone data specification v2.0, including departments of transportation of Iowa, Maricopa County, Arizona; Massachusetts; and Texas (1). Integrating work zone event data with other data sources presents several challenges, most of which focus on data collection and the dynamic nature of work zones:

- Data collection for work zone is done on an as-needed basis, requiring multiple interfaces and significant processing to standardize, which in turn necessitates committing IT services to accomplishing data standardization.
- Data are often not collected in a systematic manner due to difficulties associated with identifying what information to collect and the potential resulting benefits.
- Integrating work zone data elements can be challenging due to these elements often residing in data siloes such as project plans, computer aided dispatch (CAD) databases, or in some type of contract administration software where critical path tasks are input and planned. Additionally, work zones can involve different business units, including but not limited to construction, maintenance, and permitting, and these activities are not considered related and would not have their data pulled together into single work zone database.
- Maintaining work zone data requires planning and commitment of resources. It is not uncommon for an agency to publish seasonal work zone data and not proactively maintain that record unless notified by someone of an inconsistency.
- The WZDx specification includes optional fields due to the inability by some agencies to create and maintain any additional data from what they already produce. Because of this, any given data set conforming to that specification may not have all of the expected data fields available.
- Agencies may not have a good mechanism for verifying if a planned lane closures becomes active.
- Agencies interested in publishing WZED are already doing so in another format, such as the 511 system, which produces WZED using the much older and more established Traffic Management Data Dictionary (TMDD) standard.
- The dynamic nature of work zones (e.g., timing and location of work zones is not always known in advance).
- Work zones may involve various lane closures during their duration, for which information may be difficult to obtain in real time.

Several products exist that could assist agencies with understanding the WZDx specification and adopting a standardized data framework:

- The *WZDx GitHub website* provides detailed high-level overview of the WZDx standard and an entity relationship showing how WZDx data tables interrelate and the required and optional fields, along with detailed descriptions of individual of individual data tables, fields, and enumerations.
- The *Framework For Standardizing and Communicating Work Zone Event Data* provides a conceptual architecture for a WZED warehouse and for work zone data systems (WZDS) for collecting, storing, using, and communicating WZED. This source also includes an introductory implementation guide for agencies that have considered adopting the WZDx standard (3).
- *Deploying and Integrating Smart Devices to Improve Work Zone Data for WZDx* summarizes the efforts undertaken by Iowa DOT to obtain work zone data from the deployment of smart arrow boards with the goal of improving the location and temporal accuracy of work zone information as part of a WZDx feed (4).

Traffic Incident Management (TIM) Performance Measurement

This use case focuses on the sharing, integration, and management of data to support TIM program performance measurement and management. TIM is a planned and coordinated process to detect, respond to, and remove traffic incidents and restore traffic capacity as safely and quickly as possible. This coordinated process involves a number of public and private sector partners, including law enforcement, fire and rescue, emergency medical services, transportation, towing and recovery, and others (5). There are many benefits to measuring and monitoring the performance of TIM programs; however, performance measurement requires data, which historically has been lacking.

FHWA has conducted multi-state efforts and has sponsored numerous initiatives to define and institutionalize the use of three TIM performance measures nationally. One of these initiative was Every Day Counts Round 4: Using Data to Improve TIM and its objective was to increase the quantity, quality, and use of data for TIM performance measurement and analysis. Seventeen states (6) that took part in this initiative improved their TIM performance measure data collection, integration, and use through:

- Including the TIM data elements in crash reporting, extending the analysis of TIM data to agencies beyond the state DOT, and employing a quantitative data systems for TIM data analysis
- Expanding training of data collection
- Making use of other existing data sources (e.g., CAD, Waze, INRIX) to extract TIM data
- Creating a TIM data dictionary to ensure and improve the consistent reporting of the performance data
- Developing dashboards to better visualize and relay information to traffic center management (TMC) operators and subsequent stakeholders
- Increasing the frequency of reporting TIM data performance measures
- Using the data to support decision making in allocating resources and developing countermeasures and to support policies and propose legislation

Despite the continuing efforts in improving TIM performance measures, states continue to face a number of challenges with their own TIM-relevant data for performance measurement, management, and analysis:

- Incomplete coverage – transportation agencies often have few data on traffic incidents beyond the data collected at the TMC and through their safety service patrol (SSP) programs, leaving gaps in times (e.g., overnight), in rural areas, and off highways.
- Timeliness of data –transportation agencies often have to wait until an incident is reported by law enforcement or the public, particularly outside of TMC/camera coverage areas. Additionally, crash data in some states may not be available for months or even a year after they were collected.
- Lack of data quality/reliability – quality issues include lack of sufficient detail in the data, manually entered data, and the use of proprietary "black box" systems, which subsequently lead to difficulties with understanding and reporting TIM performance.

Given the efforts to date to improve data sharing, integration, and management of TIM-relevant data to support TIM performance measurement and analysis, there exist a number of products that could support state DOTs in making progress in this area:

- *Guidance for Implementation of Traffic Incident Management Performance Measurement: Final Guidebook* (7) – this product of NCHRP 07-20 provides a working framework to support an agency or TIM program in understanding TIM performance measurement and in establishing a TIM performance measurement database. The guidance provides consistent definitions, data sources, example reports and data visualizations; discusses common challenges; and presents key points for success. Products include a final report (7) and an associated website (8).

- *TIM Performance Measures Database Schema* (8) – another product of NCYRP 07-20, which organizes 34 TIM data elements to support comprehensive TIM performance measurement and analysis.
- *NCHRP Research Report 904 Leveraging Big Data to Improve Traffic Incident Management*– provides guidelines for states on leveraging a big data approach to improve TIM, including collecting and storing more data, opening and sharing data, breaking down data silos, and integrating various data sources via a common data (cloud) storage and analytics environment, and maintaining data accessibility, to name a few (9).
- *NCHRP 03-138: Application of Big Data Approaches for Traffic Incident Management (TIM)* – this project will address a range of challenges related to Big Data and its use by state DOTs to enhance incident management, performance assessment, and predictive analytics through 1) use cases that identify data sources and possible analyses to enhance TIM, 2) exploring the integration of non-traditional data with traditional TIM data, 3) building a data environment for data storage and analyses, 4) developing case studies, and 5) developing a user guide with identified techniques and tools for use by state DOTs and other agencies (10).

Managing Big Data from Emerging Technologies

This use case focuses on the integration and management of data from emerging technologies. There are many emerging technology data sources today, such as crowdsources data (e.g., Waze, vehicle probes, StreetLight), connected vehicle (CV) data, new mobility service provider data (e.g., scooters), and IoT (e.g., smart device data for work zones, parking, fleet tracking, and traffic signals) that hold great potential to offer new insights. This use case focuses on crowdsourced and CV data.

FHWA's Every Day Counts Round 5 (EDC-5) *Crowdsourcing for Operations* innovation (11), conducted in calendar years 2019 and 2020, assisted agencies in increasing their use of crowdsourced data to better operate the transportation system through new, cost-effective, and proactive operational strategies and applications. FHWA provided support to 30 states in using crowdsourced data to enhance early detection of incidents, traveler information, traffic signal retiming/studies, road weather management, and detour route management. Due to the demand by transportation agencies to further increase their understanding and use of crowdsourced data, the *Crowdsourcing for Operations* innovation was selected again as an EDC-6 innovation (12), and FHWA will be providing assistance to states and local agencies through calendar year 2022.

NCHRP 08-116 *Framework for Managing Data from Emerging Transportation Technologies to Support Decision-Making* reviewed the state of practice at state and local agencies for identifying, collecting, aggregating, analyzing, and disseminating data from emerging public and private transportation technologies, synthesized the kinds of data being collected and used, and assessed data requirements and gaps. There are currently little connected vehicle data being generated save for the pilot projects happening across the country, which often focus more on the application/use of the technology over the use of the data, and the traditional approaches to data collection, sharing, management, and analysis are not sufficient for emerging "big" data (13).

Given the nature of big data, it is not surprising that transportation agencies struggle to manage it. Some of the common big data challenges encountered by agencies are summarized below:

- Most agencies do not recognize or understand the need for a complete paradigm shift in the way data from emerging technologies must be managed.
- There is a lack of trust in external data sources, as well as some internal data sources.

- There is a misperception that locally-hosted data are more secure than online cloud-hosted data, when often the opposite is true.
- Many organizations do not see the eminent need to address the management of emerging technology data and therefore are not active in seeking a solution.
- Vendors have increased control over technology and service agreements.
- Agencies have difficulty accessing and meaningfully using data from third parties.

Probably the biggest barrier to agencies being able to manage and make use of data from emerging technology such as crowdsourcing and connected vehicles, as previously state, relates to their traditional approaches to data management. These traditional systems and practices cannot meet the needs and requirements of data from emerging technologies; new systems and approaches are needed if these data are to be effectively used and managed by transportation agencies.

Two products were identified that could provide support to agencies in managing emerging types of big data:

- *NCHRP Research Report 952 Guidebook for Managing Data from Emerging Technologies for Transportation* – contains over 100 recommendations for managing data from emerging technologies in a modern way and contains a roadmap for implementing the guidance, as well as several tools including a modern data management capability maturity self-assessment. This guidebook is a good resource for agencies grappling with how to manage new, large datasets (14).
- *USDOT's Operational Data Environment (ODE)* – the ODE is a real-time virtual data router that ingests and processes operational data from various connected devices – including vehicles, infrastructure, and traffic management centers – and distributes it to other devices and subscribing transportation management applications. The system allows applications to submit data through a variety of standard interfaces, supporting both the producers and consumers of connected vehicle data. As it is open source, it can be used and/or adapted by localities for their deployment of connected vehicle operations. The code can be found on the USDOT CV Pilot Git Hub repository (15).
- *U.S.DOT: Integrating Emerging Data Sources into Operational Practice—Opportunities for Integration of Emerging Data for Traffic Management and TMCs* – this report provides agencies responsible for traffic management with an introduction to big data tools and technologies that can be used to aggregate, store, and analyze new forms of traveler-related data (i.e., connected travelers, vehicles, and infrastructure) and to identify ways these tools and technologies can be integrated into traffic management systems and TMCs (16).

Next Generation Integrated Corridor Management (ICM)

This use case focuses on the on the integration and management of data to support integrated corridor management (ICM), an approach that involves aggressive, proactive integration of infrastructure along major corridors in order for transportation professionals to fully leverage all existing modal choices and assets (17). ICM gives corridor managers an approach and toolbox to integrate developments in technologies, collaborative practices, and operations innovations from the last decade for improved corridor performance and enhanced traveler experience (17).

The U.S. Department of Transportation (USDOT) selected two corridors to demonstrate the nation's first ICM systems, one in US-75, Dallas, TX and the second in I-15, San Diego, CA (17). However, as the pilot sites have been further enhanced and new ICMs have been implemented, datasets have evolved, having more attributes than were available a decade ago. Some of the challenges that agencies faced in dealing with ICM data included:

- Data architecture as an afterthought – systems are conceptualized without consideration of data architecture and computational needs for fusion. Architectures were only vaguely described at concept and only fully articulated at the end of development.
- Unfamiliarity with cloud systems – agencies are concerned (and fearful) of perceived security challenges with cloud storage and use. Many do not fully understand their options nor the cost implications (or possible savings) of the cloud in lieu of on-premise solutions.
- Open source is difficult – open source development is much harder to realize when some software developers have commercial interests to keep their solutions, data structure, or methodology proprietary. In some instances, sharing of data is done on the condition of omitting certain data elements. Some data providers only provide a "cleaned" version of the data rather than allowing access to the raw data.
- The end goal for ICM is a challenge – traveler behaviors are difficult to change dynamically. Consequently, the jury is still out on how successful ICMs are in achieving transportation modal shift. Travel decisions are made at the beginning of commutes and once that decision is made, the travel back home typically cannot be changed (i.e., travelers that drive in the morning do not take transit home).

Few recent documents or products exist that support the sharing, integration, and management of data for ICM. *Advances in Strategies for Implementing Integrated Corridor Management (ICM), Domestic Scan Team Report* provides a review of several implementations of ICM and proposes lessons learned and best practices (18). Within this document, a Capability Maturity Model (CMM) Approach is defined that provides a disciplined process to address varying levels of maturity of six core building blocks of an ICM.

Using Data for Effective Operations and Enforcement of Private Shared Mobility Services

This use case specifically focuses on the data issues related to effective operations and enforcement of private sector shared mobility services. Shared mobility is "the shared use of a vehicle, motorcycle, scooter, bicycle, or other travel mode; it provides users with short-term access to a travel mode on an as-needed basis" (19). Its scope includes micromobility services, such as bikesharing and electric scooter services, as well as carsharing, micro-transit, Transportation Network Companies (TNCs) and traditional ride-hailing (taxi) services. Shared mobility providers provide varying amounts and types of data to public agencies in essentially every locality where they operate. However, the type, quantity, frequency, and format for the data varies widely. Although dockless bikesharing and scooters came along after TNCs and docked bikeshare programs, there has been more progress made in the area of standardizing data exchanges for these micromobility systems than there have been for motorized modes such as TNCs, taxi services, and microtransit. Various types of micromobility data are currently collected by operators of various cities in the states of Texas, Illinois, Tennessee, Minnesota, Washington, Utah, Oregon, and California (20).

Some of the challenges in improving the sharing and use of private sector shared mobility data are:

- General lack of familiarity with the data that shared mobility companies collect and how these data might be used to inform planning, operations, and the management of the multimodal system.
- Tradeoffs between the legitimate needs of public agencies and the protection of private and proprietary data. This is the largest and most complex challenge. Public agencies need to collect data for planning, operations, and enforcement, yet they often cannot get access to privately-owned data.

- Coordinating data collection and analysis across systems and modes. The combination of the two standards that address micromobility data, Mobility Data Specifications (MDS) and General Bike Feed Specification (GBFS), provides excellent standards for the exchange of micromobility data, and their use continues to grow. However, similar standards do not currently exist for taxi fleets, TNCs microtransit, or car sharing. Currently each jurisdiction sets its own rules, and either the public agency or the private provider determines the format for providing the requested or required data.
- Existing planning models do not account for new shared mobility technology that has just begun to significantly alter travel behavior.
- Data sharing/accessing or private sector (proprietary) data – e.g., negotiating data sharing agreements. Local jurisdictions typically develop their own individual agreements and data sharing requests with shared mobility companies (opportunity lost to share with others).

Several standards, guidance documents, and open source products were identified for exchanging micromobility data, and those include the following:

- *Mobility Data Specification (MDS)* – a widely used standard for exchanging data between micromobility operators and public sector agencies. It supports real-time data exchange and the exchange of detailed trip-specific information, as well as broader categories of information. MDS is currently maintained and governed by the Open Mobility Foundation (21).
- *Data Sharing Glossary and Metrics for Shared Micromobility* – provides a rigorously-defined data dictionary and definitions for several performance metrics (22).
- *General Bikeshare Feed Specification (GBFS)* – complements the MDS and is an open standard for providing public, real-time, read-only data on bikeshare systems (23).
- *Uber Movement* – Uber's first effort to voluntarily provide aggregated open data for use by public sector entities and others. It provides speed and travel time data, as well as mobility heat map visualizations at both the census tract and Transportation Analysis Zone (TAZ) levels (24).
- *SharedStreets Referencing System* – an open approach enabling differences in GIS systems to be reconciled, so that data from different GIS sources can be combined and used together (25).
- *CurbLR and Open Curbs*. CurbLR is an open source effort to develop a standard for digitizing curb information, such as rules and regulations for parking and delivery, as well as information about the physical infrastructure (26).
- Open Curbs is a similar effort by Coord, a Sidewalk Labs spinoff (27). Having curb data in digital form is highly valuable both for public agencies managing mobility and for mobility providers.
- *Managing Mobility Data* and *Guidelines for Mobility Data Governance and Contracting* – two documents that provide a start for developing more specific and concrete model frameworks for data exchange and data governance. They provide a good discussion of the issues in well-structured forms and sets of recommendations. Neither is at the level of specificity to serve as model agreements (20, 28).

Improving Transportation Planning with Emerging Data

This use cases focuses on improving transportation planning through the use of emerging data. Emerging technologies have led to new and emerging datasets, such as passive data from cellular phones, data collected from shared mobility apps, and data from an array of micromobility operation. Such data provide new and innovative ways to analyze transportation demand and traveler behavior and, combined with that collected by more conventional methods, through appropriate preprocessing, are actively being used as a resource to measure and analyze transportation system performance and plan mitigating policies and strategies. However, agencies are encountering a number of challenges both in integration and interpretation of much of these data. Therefore, despite the availability of such new big data sources, transportation demand forecasting models (TDFMs) used primarily for longer range planning are still, almost exclusively, based on conventional data such as travel diary surveys and population census (29).

A number of the specific challenges associated with evolving travel behavior and integrating new data sources into the planning process are itemized below:

- The desire to integrate private and third party datasets into the regional and community planning process is generating new issues in understanding the statistical reliability of these data. Furthermore, passively generated datasets lack ground truth to be validated against. Tools and procedures used to clean and preprocess the information, as well as ambiguities in terms of such measures as sample size and levels of confidence are concerns for agencies trying to utilize this data to calibrate and validate travel models.
- Although very large amounts of data are often available, usable data for travel model development must be stratified in terms by traveler and trip characteristics. In the case of some forms of big data, the location information is generated by non-transport activity (e.g., during phone call, text message etc.); hence, it cannot be converted directly to mobility data for transport studies and often requires significant processing. Lastly, accessing and understanding data derived by interpretation and preprocessed from third parties can be challenging.
- Lack of detail in private sector data (e.g., origins-destinations, geographic resolution, discontinuity in location data, missing information not sufficient to be integrated into regional modeling and forecasting). Some required elements and classifications (e.g., trip purposes, accompanying travelers) are not explicitly recorded. Furthermore, the data may lack personal or socio-demographic information of the user, which are key inputs.
- Data are collected at different periods and spatial units (census data are usually available at the block level, while mobile phone data relate to individual cell towers) creates difficulties in comparing the different datasets, even when they are related.
- Proliferation of geo-referenced data and networks is demanding better methods and conflation to reconcile inconsistent mapping.
- Finding and applying the right models to the given problem and informing those models with appropriate data has become an increasingly challenging process.

A number of efforts have been identified that are underway to standardize the formats and information content associated with various types of transportation planning data. Many of these efforts have broader scope than simply transportation planning data and have significant implications for agency data management in general:

- *All Road Network of Linear Referenced Data (ARNOLD)* – developed and maintained for all publicly-maintained roads in the state. Provides a geographically referenced platform for describing the linear extent of a wide variety of attributes associated with the state road system, including attributes related to fiscal management, safety, freight, and performance management (30).
- *Model Inventory of Roadway Elements (MIRE)* – an extension of ARNOLD, is used for safety analysis that could allow extraction of such characteristics as connectivity of network elements, as well as identification of intermodal crossings (31).
- *TransXML* – a expandable vocabulary/syntax for describing transportation elements using the XML language (32).

- *Open Matrix File Format (OMX)* – a recommended matrix format standard for storing and transferring matrix data. It is used to get zone-to-zone trip flow matrix files out of a proprietary format and into something that can be read by many programs (33).
- General Travel Network Specification (GTNS) – a format and proposed standard for describing travel model networks based on Open Street Map (OSM) (34).

Freight Data for Planning and Operations

This use case focuses on sharing, integration, and management of relevant freight data sources for planning and operational purposes. Timely, comprehensive, and high-quality freight data support agency planning and operational efforts by identifying and aligning critical needs with infrastructure investments. In recent decades, the FHWA, the American Association of State Highway and Transportation Officials (AASHTO), the Transportation Research Board's (TRB) National Cooperative Highway Research Program (NCHRP), and the Strategic Highway Research Program (SHRP2) have invested in a number of freight data related studies. Specific needs for freight data identified through these studies include freight demand modeling, data collaboration, standardization, integration, and sharing, and freight data architecture implementation.

The literature identified several challenges associated with the use of freight data for transportation planning and operations purposes. These challenges are technological, organizational/institutional, and administrative:

- Variations in data standards and data collection programs across state and international boundaries create a challenge for cross-border and interstate goods movement.
- Differences in vehicle classification by data providers make it a challenge to compare and combine datasets. Similarly, there are differences in the fleet mix of data providers. Fleet mix represents the number of vehicles for each vehicle classification or type represented in a data vendors sample.
- Global Positioning System (GPS) and Location Based Services (LBS) data providers depend on permanent traffic counters for the expansion of their sample data to better represent the population of truck traffic. However, the number of permanent traffic counters on the roadway network is limited and varies by state.
- Agencies have limited funds to hire staff and expertise to help identify, develop, process, analyze, and apply freight data to address agency needs.
- Integration of infrastructure planning and operations with available freight data sources remains a challenge for public sector agencies.
- Limited data for validating model outputs makes it a challenge for agencies to test the reliability and sensitivity of their models to real-world conditions and changes.
- Rapidly changing trends in consumer demand are driving e-commerce growth and making it a challenge for state and local agencies to plan accordingly.
- Differing planning horizons for public and private sector agencies (years versus days or weeks) result in a disconnect regarding investments to develop, procure, and share data.
- There is a lack of incentive for private industry to share data. Additionally, it is a challenge to persuade private industry to share data when that data could be mishandled in a way that creates liability for the shipping company or reveals trade secrets to competitors.
- When data can be obtained from private industry partners, they may stipulate certain restrictions on how the data can be used or shared.

There exists a number of products to support freight data sharing, integration, and management including:

- *NCFRP Report 35 Freight Data Dictionary (FDD)* – a product of project NCFRP 47(35), provides recommendations for the effective use of freight data including identifying and resolving differences in data element definitions, as well as providing access to over 6,300 data elements and 13,300 glossary terms from multiple freight data sources.
- *SHRP2 C20 Freight Data Guide* **–** identifies the most relevant data sources that address specific freight data needs. This guide also contains data acquisition (and negotiation) tips, examples of data applications, and a table of common freight data sources in two categories: publicly available and commercial (36).
- *Freight Data Sharing Guidebook* – identifies the most relevant data sources that address specific freight data needs. The guide also contains data acquisition (and negotiation) tips, examples of data applications, and a table of common freight data sources in two categories: publicly available and commercial (37).
- *Freight demand modeling and local freight data improvement case studies* – including those by the I-95 Corridor Coalition, the Washington Metropolitan Council of Governments, the Baltimore Metropolitan Council, Iowa DOT, and Utah DOT (38).
- *Freight and Service Activity Generation (FSAG) Software* – developed by the Center for Infrastructure, Transportation, and the Environment (CITE), estimates the number of daily freight deliveries, freight shipments, and service trips attracted to an establishment or ZIP Code (39).
- *Hybrid freight data ontology* – allows for disparate data sources to be described by their own unique ontologies and then to be mapped to a single upper-level "global" ontology. This approach allows for enough integration at the global ontology level to allow all databases to respond to the same queries, while remaining flexible enough to easily allow for data sources to be added or modified (40).

Integrating ITS Transit Data for Planning and Operations

This use case focuses on transit data, with a particular focus on the data sharing, integration, and management needs of transit agencies for planning and operations. Transit ITS data is a valuable resource for transit agencies. There is great potential to use the data to create enhance effectiveness and efficiency and move towards more data-driven decision-making. But while transit ITS offers promising potential, the use of ITS data is often an afterthought. There are growing trends towards interfaces, coordination, and integration of ITS technologies between transportation and transit agencies. As urban transportation is being transformed through big data, mobility on demand, and connected and automated vehicles, much of the transit industry is poorly positioned to be a full partner in these transformations (41). Some of the biggest challenges are listed below:

- Integrating different ITS sets together and with other sources:
    - Each system (e.g., AVL, APC and AFC from different vendors) has different database structures and data formats.
    - Wide variation in data types exist that are specific to individual applications, limiting the possibility for integration and extending the time needed for employees to gain a working knowledge of applications.
    - Multiple competing standards such that integration between different data systems or applications is difficult or impossible to achieve.
- Few agencies are able to provide broad access to the data within their organization, both due to technical limitations and usability concerns. Even if they could provide organization-wide data access, individual users would be challenged to make sense of the complex heterogenous data.

- Technology and data often take a back seat to ongoing operations and sufficient funding – the transit industry is more of an operations-driven culture than a data-driven decision-making culture.
- Due to the lack of open data standards, guidance, and third-party assistance many agencies develop their own unique approaches and products, often "reinventing the wheel" unnecessarily.

Better integration and management of transit data could improve the use of these data by transit agencies to support transit planning and day-to-day operations. To that extent, several standardized data models and schemas are available to transit agencies for adoption and use:

- *General Transit Feed Specification (GTFS)* – defines a common format for public transportation schedules and associated geographic information. GTFS "feeds" let public transit agencies publish their transit data, and developers can write applications that consume that data in an interoperable way (42).
- *Service Interface for Real Time Information (SIRI)* – allows for the structured exchange of real-time information about schedules, vehicles, and connections together with general informational messages related to the operation of the services (43).
- *GTFS-Ride* – an open, fixed-route transit ridership data standard developed through a partnership between the Oregon Department of Transportation and Oregon State University. GTFS-Ride allows for improved ridership data collection, storing, sharing, reporting, and analysis (44).
- *Transit Communications Interface Profiles (TCIP)* – defines mechanisms for the exchange of information among transit business systems, subsystems, components, and devices (45).
- *Outputs of TCRP G-18 Improving Access and Management of Transit ITS Data* (currently underway) – the objective of this project is to develop a common, practical approach to storing, accessing, and managing fixed-route transit ITS data. Final products of this project (expected in mid-2021) will include a final report documenting the entire research effort and documentation of the data structure / specification along with detailed functional requirements and guidance on use, application, and maintenance of the specification (46).

Modernization of Active Traffic and Demand Management (ATDM) Approaches

This use cases focuses on data acquisition, assessment, validation, and management in supporting proactive and responsive traffic demand management by transportation agencies. Active traffic and demand management (ATDM) strategies aim to follow a cycle of monitoring transportation systems, assessing performance, formulating and recommending dynamic actions, and implementing actions to improve travel reliability, safety, and throughput by balancing traffic demand and network capacity (47). ATDM promotes dynamically influencing, managing, controlling, and linking travel demand and traffic management based on real-time performance feedback. ATDM strategies strive to maximize system efficiency by preventing, delaying, and more quickly mitigating breakdown conditions. Examples of ATDM strategies include:

- Predictive traveler information – relies on the use of historic network trends, current conditions, and expectations of future conditions that change supply or demand (e.g., rain, snow, crash, ending of a sporting event, roadwork, etc.) to predict and share travel times for the next few minutes to few hours. Agencies are still challenged to provide real-time travel times, and significant data integration and analytical techniques are required for predictive traveler information. However, several states are making progress in utilizing and integrating various data sources to improve the provision of travel times either via dynamic message signs (DMS) or their 511 traveler information services (48, 49, 50).

- Queue warning – typically rely on changes in traffic volumes or travel speeds, incident detection and location, and CCTV cameras to identify the formation of queues. In addition to infrastructure-based monitoring equipment, facility operators have begun to leverage IoT, crowdsourced data, and third-party tools that make use of vehicle probe speed data to identify traffic queues. Traditionally, agencies communicate warnings to drivers using DMS or portable changeable message signs; however, more recently, the communication of queue warning is done through integration with private navigation applications.
- Adaptive and integrated ramp metering balance freeway and arterial delays based on real-time traffic flow and incident data. Data needed to support adaptive ramp metering includes traffic speeds and volumes by freeway lane, ramp volumes and geometry (storage, acceleration distance, etc.), and crash history. ATDM metering strategies engage simulation and benefit from more robust field data, including vehicle probe speed, connected vehicle, incident, and Automated Traffic Signal Performance Measures (ATSPM) data. Challenges of achieving adaptive metering include access to verified real-time data, higher costs for complex systems, and difficulty in timely performance measurement. A recent example of an adaptive ramp metering system was developed in-house by the Arizona Department of Transportation. The system monitors upstream and downstream freeway segments and adjusts the ramp metering rate to utilize the freeway on-ramp storage to increase segment and corridor flow rates, and an associated evaluation tool provides before and after comparisons of speed, flow, capacity, delay, and delay cost (49).

Florida DOT launched its Data Integration and Video Aggregation System (DIVAS) is 2017 to integrate and manage actionable real-time traffic data. It is a centralized hub for aggregating, fusing, and disseminating near real-time information and live stream video. It combines information from its SunGuide ATMS, Next Generation FL511 system, and third-party data to integrate and manage actionable real-time traffic data. It is a centralized hub for aggregating, fusing, and disseminating near real-time information and live stream video. It combines information from its SunGuide ATMS, Next Generation FL511 system, and third-party data (51).

A fundamental challenge to ATDM is the traditional approach to system development in DOTs that limits flexibility and adaptability. Agencies often need to build another system to implement and monitor real-time conditions. These considerations impact the cost of implementation and cohesive, comprehensive system implementation. Other challenges related to ATDM include:

- Access to verified real-time data
- Higher costs for complex systems
- Understanding processes to support data verification, fusion, and analytics along with data management – real-time data collection and processing from multiple sources is needed to support ATDM.
- Many agencies are not at the data maturity level to support ICM/TSMO strategies, let alone ATDM. Applications, algorithms, and simulation models often evolve with limited internal oversight and understanding.

Preparing Agencies for Smart Cities

This use case focuses on the transportation side of smart cities and how transportation data, ranging from traditional traffic volumes to pedestrian data derived from the Internet of Things (IoT) devices, can be shared, integrated, and managed within a smart city. Smart cities bring together infrastructure and technology to improve the quality of life of citizens and enhance their interactions with the urban environment (52). A positive quality of life involves enhancing every aspect of the daily existence of citizens. From safe streets to green spaces, from a reasonable commute to access to art and culture, a smart city creates an environment that promotes the best of urban living and minimizes the hassles of city life (53).

A few national and international smart city "challenges" have been issued to explore ideas and develop best practices. The winner of the USDOT Smart City Challenge, Columbus, Ohio was designated as America's Smart City. The Smart Columbus project funded the development of a range of smart city projects – from self-driving shuttles to prenatal trip assistance to a multimodal trip planning application – that demonstrate the value of data in a smart city environment. The Smart Columbus project data backbone, a web-based, dynamic, governed data delivery platform, stores over 3,000 datasets and features an open data platform for the region, syndicating public, anonymized open data accessible to other Columbus city agencies, academic researchers, startups and app developers, and the community at large (54).

Smart City PDX, the smart city project for Portland, OR, is partnering with the community to make Portland a place, where data and technology are used to improve people's lives, particularly in underserved communities. The city of Portland started the Portland Urban Data Lake (PUDL) pilot to address the need for a centralized system to manage, integrate, and analyze data generated by smart city technologies (55). Three of the smart city project pilots included managing eScooter data generated at high volumes, combining Waze crowdsourced data with existing transit data to improve the accuracy of TriMet's "Transit Tracker" app, and measuring how the technology behind PUDL could help the city migrate away from traditional, siloed, server-based data management and the potential impact of this migration (e.g., cost savings).

A smart city initiative comes with a wide range of challenges some of which include:

- Smart cities are built on pervasive and open networking and cheap and disposable hardware (i.e., system of systems), which are the opposite of how cities and states currently develop or procure their information systems (i.e., traditional system design).
- The lack of an "umbrella agreement" available creates difficulties in sharing the data. This can result in cities having to create new data sharing agreements for any new data request, an approach that is increasingly frustrating and inefficient, both for data requestors and cities staff.
- There are few unified standards to address the many possible aspects of a smart city program.
- Differing, and sometimes competing, ideas between city agencies and departments on data architecture solutions and methodologies.
- The difficulty of comparing the costs of on-premise options with cloud-based architecture. Because cloud-based technology fees are based on use, this cost-benefit comparison is nearly impossible to make without first knowing how much data will be collected and how frequently it will be queried. Furthermore, no data exists to baseline the cost of the new system, as there is no equivalent on-premise system using that much data.
- Necessary funding levels for projects can be difficult to define if information about data volumes and access frequency are not yet known.

There are multiple existing products that can help to support smart city data management, ranging from open source data sharing platforms, to common data frameworks, to entire data ecosystems and ontologies:

- *Smart Columbus OS (SCOS)* – serves as an open data platform for the region, syndicating public, anonymized open data accessible to other Columbus city agencies, academic researchers, startups, and app developers and the community at large. SCOS currently stores over 3000 datasets (54).
- *NCHRP Research Report 952 Guidebook on Managing Data from Emerging Technologies* – provides guidance, tools, and a big data management framework, and it lays out a roadmap to guide a move toward the effective use and management of data from new technologies (14).
- *DataSF* – an open data sharing and analysis platform for the city of San Francisco containing over 500 datasets, many of which can be directly relevant to transportation applications (56).
- *Smart City Interoperability Reference Architecture (SCIRA)* – a project of the Open Geospatial Consortium (OGC) Innovation Program sponsored by the US Department of Homeland Security (DHS), Science and Technology (S&T). The SCIRA pilot demonstrations and reports were finalized in January 2020 and the toolkit and guides are currently under development. SCIRA aims to advance standards for Smart Safe Cities and to develop open, interoperable designs for incorporating IoT sensors into city services (57).
- *KM4City Ecosystem* – a complete, open-source, big data solution for smart cities, with an ontology at its core integrating the various datasets common to smart cities projects.
- *SNAP4City platform* – a turnkey, complete, open-source, microservices solution for smart cities, built on KM4City, to more easily manage complex smart city data sources and networks and to deploy, maintain, and evolve smart city applications (58).
- *Smart Cities Data Sharing Framework* – a blueprint for a common framework, a set of critical components, and an evolutionary path from data collection to data monetization (59).
- *Smart Cities Data Exchange Framework* – details the processes for taking data from community development source systems (such as traffic sensor data and affordable housing stock), creating a pipeline for data transformation into a common open schema, merging data from across multiple communities, and serving data through APIs, discovery systems, and visualization tools (60).
- *Open Connectivity Foundation* – one of the main industrial connectivity standards organizations for the IoT; is working to secure interoperability for consumers, businesses, and industries by delivering a standard communications platform, a bridging specification, an open source implementation and a certification program allowing IoT devices to communicate regardless of form factor, operating system, service provider, transport technology or ecosystem (61).
- *FIWARE* – a European effort developed by the FIWARE Foundation, a non-profit organization founded in 2016 that focuses on developing a reference architecture, roadmaps, specifications, and libraries supporting cloud, IoT, networking, data management and analysis, and user interface and security (62).
- *CityKEYS* – a common smart city performance measurement framework from the Technical Research center of Finland that provides a performance measure framework, suggested datasets and sources, suggested indicators and indices, business models and opportunities and policy making recommendations for small and large cities (63).

Summary of Challenges, Needs, and Opportunities

The literature review, stakeholder interviews, and associated needs assessments resulted in the identification of data sharing, integration, and management challenges, needs, and opportunities of which to further explore and address in this NCHRP 08-119 project through the development of various products. The challenges, needs, and opportunities identified for the individual use cases showed significant overlap. The most common challenges seen across the use cases include the following:

- Benefits of new approaches / technologies for data management are not fully understood or appreciated by transportation agencies (particularly at the leadership level).
- Data siloes; lack of internal data sharing; and insufficient tools, technologies, and technical data expertise inhibit the ability for agencies to fully exploit their data.
- Data are not properly maintained, leading to data quality issues.
- Data are insufficient – there is a lack of real-time data, raw data, and sufficiently granular data to support analyses for performance management, operations, planning, and decision-making.
- Data are collected at different periods and spatial units, which creates difficulties in comparing the different datasets, even when they are related.
- Available data sources are not easy to integrate.
- Proprietary systems / data ownership issues prevent the integration, sharing, and use of data as well as the ability for agencies to validate or ensure data quality.
- The increasing availability of data from emerging technologies is creating massive datasets that do not fit in the existing relational database systems of most transportation agencies – existing data architectures are unable to effectively ingest, process, or manage this increased data load.
- There is a reluctance to share data (internally and externally) due to privacy, security, and accuracy concerns.
- Lack of rights to the data – ownership of data produced by technology vendors does not always reside with transportation agencies.
- Misaligned objectives/motivations of the private sector (e.g., private sector processes and formats can differ from those used in the public sector, limiting their ability to integrate data or collaborate with transit agency partners).
- Agencies lack expertise in developing contractual agreements / drafting and enforcing favorable contractual terms.
- Approaches to protect data privacy negatively impact data usability.
- Outdated analyses / processes do not account for new data, new technologies, or new behaviors.
- Administrative barriers, traditional IT approaches, and legal processes hinder the procurement of modern data, tools, and services.
- Misaligned objectives/motivations of the private sector (e.g., private sector processes and formats can differ from those used in the public sector, limiting their ability to integrate data or collaborate with transit agency partners).
- Lack of internal expertise of data aggregation and visualization techniques or supplier-provided tools to match, clean, maintain, and analyze data to create information and standardized reports that provide meaningful information to support planning and operations decision making.
- Keeping pace with the data industry – even small changes to data structures, architecture, or standards requires time for agencies to comply with.

Likewise, the most common needs across the use cases include the following:

- Standards for datasets or data products that are relevant to modern data applications.
- Automated, standardized methods to improve data quality and/or data collection.
- Support for integrating new datasets / enriching existing datasets.
- Guidance on assessing the quality and potential of various datasets, as well as identifying potential use cases for the data.
- Updated internal and external business processes to facilitate data integration, sharing, and management.

- Improved data coverage with detailed, real-time data.
- Improved data sharing and collaboration both internally and with partner agencies.
- Education / training on modern data management practices and other data science skillsets.
- Cost-benefit analysis to support the adoption of modern data architecture and the associated changes in procurement.
- Guidance on how to communicate data effectively with executives and other internal stakeholders.
- Guidance on drafting and enforcing technical contracts with data providers / data users.
- Policies to reconcile the need for usable data with the need to protect sensitive information / personally identifiable information (PII).
- Guidance on providing incentives for private industry partners to provide data.
- System requirements for data product procurement, deployment, and testing.
- Open data platforms for data sharing with access methods that meet the needs of multiple audiences.

Given the findings from the literature review, interviews, challenges, and needs assessment, the team identified products that can help to "move the needle" in the state of practice in data sharing, integration, and management within transportation agencies. Most agencies still operate with siloed data, traditional data systems, and within a culture that does not rely on data to inform decision-making. The goal for NCHRP 08-119, is to develop products that introduce agencies to modern data practices that can ultimately transform the way they do business. Based on the specific challenges/needs associated with each use case category, the team identified a number of potential opportunities to address the challenges and needs identified by developing additional products as part of NCRP 08-119. Compiling the potential products proposed across the use cases categories resulted in 30 potential products. These products generally fell into the following categories:

- Short documents/guides/fact sheets (2-4 pages) on specific topics including lessons learned; results of data quality assessments; best practices for data sharing, integration, and management; and making a business case (e.g., datasets, use of cloud)
- Assessment of real world data for quality, potential use cases, and requirements
- Pilot integration of real world data to demonstrate approach, feasibility, and value, and associated written case studies documenting the process, challenges, lessons learned, and benefits
- Condensed or simplified guidance/guidelines (5-10 pages) on various topics
- Draft data specifications/frameworks to support data integration

Some of the potential products identified were novel; others would build from / expand upon existing products. Some potential products were specific to one use case; others crossed / applied to multiple use cases. Some potential products would leverage / build from other on-going efforts. After a review and discussion of the potential products, the panel selected 10 products. These products are described in the following section.

NCHRP 08-119 Products Under Development

Based on the findings from the Phase I research and the recommendations of the project panel, the research team is developing the following products associated with the use case categories previously described:

- Pilot Integration of Iot data with the WZDX

- Short Documents Synthesizing Outputs/Lessons Learned from Developing TIM Big Data Use Cases/Data Pipelines
- Assessment of Real-World Connected Vehicle Datasets
- Demonstrate Use and Value of Crowdsourced Data Integration
- Pilot Integration of ITS Transit Data in an Integrated Corridor Management (ICM) System
- Data Sharing Resources Guide for Shared Mobility Data
- Pilot Integration of Big Data for Freight Transportation Planning and Operations
- Guidance and Demonstration for Improved Conflation and Geodata Reference Process

In addition to the products associated with the use cases, the project team is developing two additional products:

- Data Catalog of datasets relevant to the use cases
- Website/knowledge base – a centralized, functional, usable, searchable, relationally-linked resource library of relevant knowledge, resources, case studies, and data sets associated with the use cases.

Each of the products is described in more detail herein.

The Pilot Integration of IoT Data with the WZDx

This use case will bring together various approaches, challenges, lessons learned, and successes by states in combining real-time data from smart work zone devices with their WZDx data feeds. This will be achieved by collaborating with agencies engaged in these efforts and willing to share their experiences. The use case will also include an exploratory data analysis of smart work zone data and the opportunities to integrate these data with other sources to improve the identification of active work zones.

The products resulting from this effort will include the following:

- A *compilation of existing and emerging practices* of integrating smart work zone data with state WZED (focusing on WZDx-compliant data feeds), including approaches, experiences, challenges, lessons learned, successes, and benefits.
- A *short guide/recommendations on steps* for other states to repeat these processes.
- An *integrated dataset* containing WZED, smart work zone sensor data, and other data sources.
- An *accompanying data catalog* describing what features exist in each dataset, what relationships exist between features, and what processing steps (if any) were performed for that feature. (This information will also be incorporated into the overall data catalog.)
- A *one-page executive summary* to be used as an outreach document promoting the outputs of the research.

TIM Big Data Use Cases/Data Pipelines

This use case will focus on communicating information to improve sharing, integration, and management of data that support TIM program performance measurement, management, and monitoring. The products, listed below, will be largely based on the efforts and outputs of the ongoing NCHRP 03-138 project, Application of Big Data Approaches for Traffic Incident Management:

- A 4-5 page *TIM Data Sharing Guide*, which will present common barriers, misconceptions, lessons learned, recommendations, and benefits associated with sharing and integrating data across internal DOT groups, external TIM partner agencies, and private data providers in support of TIM analysis and performance management.

- A 4-5 page *TIM Data Quality and Use Guide*, which will present TIM data quality assessments, identified issues, and limitations with certain datasets; recommendations for improving TIM data quality for increased usability; and how various data sources can support a range of TIM use cases.
- A 2-page *outreach document* that transportation agencies can share with their TIM partners/coalitions regarding the uses and benefits of sharing data across agencies (e.g., partner agencies, private towing companies). Having access to this type of data can greatly increase transportation agencies' and TIM coalitions' understanding of when and how responders arrive on/depart from incident scenes, roadway and incident clearance times, and incident type and details, which can lead to improved processes.
- A 2-page *document matching TIM use cases with appropriate datasets*. This document will draw together use cases or "wish list" items identified by transportation agencies with the available and appropriate data (per data quality assessment results) needed to develop the use cases / address the "wish list."
- Two, 5-10-page *detailed case studies* documenting the development of a data pipeline and associated data products (e.g., charts/graphs, report, dashboard) to support some aspect of TIM (e.g., support efficient and effective operational/tactical decision-making, inform TIM planning).
- A *one-page executive summary* to be used as an outreach document promoting the outputs of the research.

Assessment of Real-World Connected Vehicle Datasets

This use case will focus on the availability and quality of CV datasets to support operational and planning analyses conducted by state DOTs. The product will be a *guidance document* as described below:

- The first part of the guidance will identify and clarify the likely path(s) in which CVs will be adopted, what CV data are like, and what will be necessary to access the data. This information will be gleaned via interviews with representatives from USDOT, leading states, private transportation analysis companies, and the automotive industry.
- The second part of the guidance will assess the potential value of CV data to transportation agencies via a detailed assessment of various CV datasets obtained from both public and private sources. This product will focus on using the CV data for managing roadways and planning future operations and also consider the use of data from dedicated CV deployments in conjunction with other types of vehicle data that are often captured by traffic information providers using conventional cellular technologies and provided to public agencies for a fee.

Use and Value of Crowdsourced Data Integration

This use case will focus on the challenges, lessons learned, steps taken, and recommended actions when integrating and using crowdsourced data. The use case will focus on different types of crowdsourced data and the many real-time and archived uses of that data either independently or through integration with other data. Products will include four 8-10 page guides that concisely present key information in a format that contains more depth than a fact sheet while remaining brief enough to be consumed by interested practitioners within 30-60 minutes. The products will include the following:

- *Free Navigation App Data Integration and Use* – will mainly focus on the Waze for Cities partnership event and travel time data. Also touched on in this product are vendors that offer "black box" services connected with the Waze platform.
- *Vehicle Probe Data Integration and Use* – will focus on themes common across vendors such as INRIX, HERE, or TomTom, and their real-time speed and travel time data feed. Also discussed

more abstractly will be existing services by these vendors that some agencies have also developed through in-house tools (e.g., back of queue detection).

- *Social Media Data Integration and Use* – will focus on the use of Twitter feeds for real-time operations in the context of incident detection and event management, and to some extent, virtual public involvement for transportation planning.
- *Origin-Destination Data Integration and Use* – will focus on the individual vehicle or person-level data from vendors such as StreetLight, INRIX Trips, and other sources, mainly for transportation planning applications.

Pilot Integration of ITS Transit Data in an Integrated Corridor Management (ICM) System

This use case will focus on the collection and integration of actionable transit ITS data with an ICM system to better inform traveler modal decisions. Data from ITS transit systems, such as automated passenger counting (APC) systems, can supplement automatic vehicle location (AVL) and schedule adherence data to better inform and potentially influence modal shift from personal vehicle use. Products associated with this use case will include the following:

- An *integrated dataset* containing relevant data, such as highway and arterial speeds, volumes, and travel times; GTFS; and transit ITS data. Additional data features, such as accessibility seating and bike rack information, may be included where available.
- An *accompanying data catalog* describing what features exist in each dataset, what relationships exist between features, and what processing steps were performed for that feature (This information will also be incorporated into the overall data catalog for NCHRP 08-119.)
- An *8-12 page guide* describing the steps taken, challenges faced, and lessons learned during the data gathering and integration process. This information will be concisely presented yet will contain enough detail to enable an agency to reproduce each step in the integration process using modern data management techniques.
- A *one-page executive summary* to be used as an outreach document promoting the outputs of the research.

The Data Sharing Resources Guide for Shared Mobility Data

This use case will focus on compiling for public sector practitioners the key issues facing the management of shared mobility data, detailed guidance on the resources (standards, policies, model documents, organizations, etc.) available, and a detailed assessment of the information or other support that each provides. The product will be a *guidance document and tabbed online resource* that summarizes:

- Relevant literature and online documents
- Sample documents and agreements relating to the provision, management, and sharing of shared mobility data
- Relevant data standards and open source software related to these standards
- Organizations that are active in these topic areas
- Examples of public data sets and dashboards provided by public agencies across the country

The guide will serve as a single, comprehensive document to the material that is available to help public agencies establish and operate an effective system for shared mobility data management. The guide will describe the information that can be found in each of the resources as well as point the reader to other resources that best address their specific issues. The guide will also identify and discuss topics and issues where little information is currently available.

Pilot Integration of Big Data for Freight Transportation Planning and Operations

This use case will focus on the issues facing state and other government agencies that are exploring big data sources to facilitate better decision-making in freight transportation planning and operations. Limited funds, differing planning horizons, proprietary information, insufficient human resources, and misinformation about available data have resulted in a disconnect about the acquisition, integration, and use of freight-related big data.

The use case will showcase how big data from the private sector can be acquired, integrated, and used to address a limited number of more specific freight planning and operations use cases. The products that will be developed through this effort include the following:

- An *interactive tool* to illustrate how big data sources can be integrated and used to address the selected use cases
- An *accompanying reference library of curated studies and reports* that further illustrate how freight planning agencies are using big data
- An *accompanying data catalog* describing the content, strengths, and limitations of the big data sources
- An *accompanying guide* with instructions on acquiring the data from the vendors

Guidance and Demonstration for Improved Conflation and Geodata Reference Process

This use case focuses on addressing the network conflation challenges faced by agencies. Conflation is the process of identifying common points and references to reconcile two or more geo-datasets across overlapping areas. Because of differences in scales, resolutions, and sometimes accuracy or conventions, data referring to the same location often do not have the same geographic reference and can't be combined easily. This leads to defining "near enough" criteria to expect two references to represent the same feature. The user of "near enough criteria," while allowing several corresponding features to be merged effectively, is not perfect. Network conflation is a continuing need and difficult task for agencies using geo-referenced maps and databases.

The products resulting from this effort will provide recommendations regarding best practices for network conflation tools, researching potential improvements to these tools and methods, and identifying transportation agencies where the use of best conflation practices can be demonstrated. Products will include:

- A guidance document describing identified best practices and summarizing their value and appropriate use. This document would be oriented toward practice managers potentially interested in adopting some or all for the product recommendations.
- A software tools providing an implementation pathway for improved georeferencing processing.
- Prototypical reference datasets based on partner agency inputs and demonstration results.

Data Catalog

A data catalog provides a comprehensive reference of available data sets, including their purpose, content, and metadata. Ideally, data catalogs are generic enough to provide users with an overall view of what data are available to them, yet specific enough that users can identify exactly which tables within a dataset are most relevant to their work.

It is the goal for the data catalog developed as a product of NCHRP 08-119 to become a go-to spot for analysts and others trying to find data to build insights, discover trends, and identify new products for their agencies. The datasets used to build the data catalog will include a range of data types, such as data from connected vehicles, crowdsourced, work zone, weather, ITS transit, and linear referencing systems to name a few.

For every table in each data source, existing documentation (e.g., data dictionaries, schema references, API references) will be referenced. Raw data samples will be collected, adapted to a tabular format from the source format (e.g., extensible mark-up language (XML) to comma-separated values (CSV)), and standardized to create descriptive entries for all features in that table. These feature entries will include the feature name, description, data type, data mode, examples, and allowed values where available.

This effort will result in both a master database file and a data catalog document that takes the standardized data references and converts them into a Word document with descriptive introductions and a clear, commonly used format. By automating this process, a new data catalog document can be re-generated with little effort as additional data sources are added to the master database file.

Website/Knowledgebase

The objective of the website is to provide a centralized, functional, usable, searchable, relationally-linked resource library of relevant knowledge, resources, use cases, and data sets to support advancement of data sharing, integration, and management by transportation agencies. The website will compile, across all the use case categories, and provide easy access to best practices, emerging practices, lessons learned, challenges and opportunities, and needs/approaches to prepare for new issues such as big data. A long-term goal of this effort is to create a platform that fosters intra-organizational education, thought leadership, best practice evangelization, discovery, modernization, adoption, implementation, and ongoing advancement of collective knowledge. A key challenge of will be to create a comfortable and delightful user experience that is engaging, simple, natural, and useful.

References

1. FHWA. Work Zone Data Exchange (WZDx). Retrieved April 2021 from https://ops.fhwa.dot.gov/wz/wzdx/index.htm#wzdxspec
2. USDOT. WZDx Specification v3.1. Retrieved April 2021 from https://github.com/usdot-jpo-ode/wzdx/blob/master/RELEASES.md
3. Stephens, D., Schroeder, J., & Ostroff, R. (2020). A Framework for Standardizing and Communicating Work Zone Event Data (WZED). FHWA. Retrieved from https://collaboration.fhwa.dot.gov/wzmp/updateddocuments/FHWA-HOP-20-019_Framework_DRAFT_Report.pdf
4. Knickerbocker, S., Ravichandra-Mouli, V., and A. Venkatachalapathy (2021). Deploying and Integrating Smart Devices to Improve Work Zone Data for WZDx. Presented at Transportation Research Board 100th Annual Meeting.
5. US DOT. (June 2020). Traffic Incident Management. Retrieved June 29, 2020, from Federal Highway Administration Office of Operations, https://ops.fhwa.dot.gov/eto_tim_pse/about/tim.htm
6. FHWA (April 2019). Every Day Counts: An Innovation Partnership with States, Final Report. Retrieved April 2021 from https://www.fhwa.dot.gov/innovation/everydaycounts/reports/edc4_final/#
7. Pecheux, K.K., and R.E. Brydia (October 2014). Guidance for Implementation of Traffic Incident Management Performance Measurement, Final Guidebook. Washington, DC: Transportation

Research Board, National Academy of Sciences. Retrieved from http://onlinepubs.trb.org/Onlinepubs/NCHRP/docs/NCHRP_07-20_Guidance.pdf

8. Pecheux, K.K., and R.E. Brydia (October 2014). Guidance for Implementation of Traffic Incident Management Performance Measurement, TIM PM Schema. Washington, DC: Transportation Research Board, National Academy of Sciences. Retrieved from http://nchrptimpm.timnetwork.org/?page_id=1014
9. Pecheux, K., Pecheux, B., & Carrick, G. (2019). Research Report 904 Leveraging Big Data to Improve Traffic Incident Management. Washington, DC: Transportation Research Board, National Academy of Sciences. Retrieved June 4, 2020, from http://www.trb.org/Main/Blurbs/179756.aspx
10. Pecheux, K.K. (Ongoing). Application of Big Data Approaches for Traffic Incident Management (TIM). Washington, DC: Transportation Research Board, National Academy of Sciences. Retrieved April 2021 from https://apps.trb.org/cmsfeed/TRBNetProjectDisplay.asp?ProjectID=4735
11. FHWA (2020). Every Day Counts Round 5 (EDC-5) Crowdsourcing for Operations. Retrieved April 2021 from https://www.fhwa.dot.gov/innovation/everydaycounts/edc_5/crowdsourcing.cfm
12. FHWA (Ongoing). Every Day Counts Round 6 (EDC-6) Crowdsourcing for Advancing Operations. Retrieved April 2021 from https://www.fhwa.dot.gov/innovation/everydaycounts/edc_6/crowdsourcing.cfm
13. Pecheux, K.K., Pecheux, B.B., Ledbetter, G., Schneeberger, J.D., Hicks, J., Burkhard, B., and M. Campbell (2020). Framework for Managing Data from Emerging Transportation Technologies to Support Decision-Making. Washington, DC: Transportation Research Board, National Academy of Sciences. Retrieved April 2021 from http://www.trb.org/Publications/Blurbs/181365.aspx
14. Pecheux, K.K., Pecheux, B.B., Ledbetter, G., and C. Lambert (2020). Guidebook for Managing Data from Emerging Technologies for Transportation. Washington, DC: Transportation Research Board, National Academy of Sciences. Retrieved April 2021 from http://www.trb.org/main/blurbs/180826.aspx
15. U.S. DOT ITS JPO. (n.d.). Wyoming Connected Vehicle Pilot Integrates Connected Vehicle Data into Traffic Management and Information Dissemination. Retrieved June 4, 2020, from U.S. Department of Transportation Intelligent Transportation Systems Joint Program Office: https://www.its.dot.gov/pilots/wydot_traffic_mang.htm
16. Gettman, D., Toppen, A., Hales, K., Voss, A., Engel, S., & El Azhari, D. (2017). Integrating Emerging Data Sources into Operational Practice: Opportunities for Integration of Emerging Data for Traffic Management and TMCs. Washington, DC: U.S. Department of Transportation Federal Highway Administration. Retrieved June 4, 2020, from https://rosap.ntl.bts.gov/view/dot/34175
17. USDOT. Integrated Corridor Management (ICM) Knowledge and Technology Transfer (KTT). Retrieved April 2021 from Intelligent Transportation Systems Joint Program Office https://www.its.dot.gov/factsheets/icm_ktt.htm
18. Spiller, J. N., Compin, N., Reshadi, A., Umfleet, B., Westhuis, T., & Sadegh, A. (2014). *Scan 12-02 Advances in Strategies for Implementing Integrated Corrdidor Management.* Washington, DC: Transportation Research Board, National Academy of Sciences. Retrieved August 7, 2020, from http://onlinepubs.trb.org/onlinepubs/nchrp/docs/NCHRP20-68A_12-02.pdf
19. SAE International (Septermber 2018). Taxonomy and Definitions for Terms Related to Shared Mobility and Enabling Technologies. Retrieved April 2021 from https://connectedautomateddriving.eu/wp-content/uploads/2019/01/J3163_201809-Taxonomy-Shared-Mobility-002.pdf

20. NACTO. (2019). *Guidelines for Regulating Shared Micromobility, Version 2, Chapter 5.* NACTO. Retrieved from https://nacto.org/wp-content/uploads/2019/09/NACTO_Shared_Micromobility_Guidelines_Web.pdf
21. Open Mobility Foundation. Mobility Data Specification. Retrieved April 2021 from https://github.com/openmobilityfoundation/mobility-data-specification
22. Mobility Data Collaborative. (2020). *Data Sharing Glossary and Metics for Shared Micromobility.* SAE.
23. NABSA. (n.d.). *GBFS & Open Data*. Retrieved June 4, 2020, from North American Bike Share Association: https://nabsa.net/resources/gbfs/
24. Sivaram, S.-p. T. (June 2020). Director of Policy, Cities, and Transportaton at Uber and Global Head of Privacy and Security Policy at Uber. (K. P. Michael McGurrin, Interviewer)
25. Eros, E. (May 2019). *Getting Started with the SharedStreets Referencing System.* Retrieved from SharedStreets: https://sharedstreets.io/getting-started-with-the-sharedstreets-referencing-system/
26. Eros, E. (June 2019). *Towards a Data Standard for Curb Regulations.* Retrieved from Medium: https://medium.com/sharedstreets/towards-a-data-standard-for-curb-regulations-146263bfc60
27. Coord. (July 2020). *Curb Management for Fast-Changing cities*. Retrieved from COORD: https://www.coord.com/
28. Mobility Data Cooperative. (2020). *Guidelines for Mobility Data Sharing Governance and Contracting.* SAE.
29. Cuauhtémoc, A., Fourie, P., & Erath, A. (2016). *Transport Modelling in the Age of Big Data.* Zurich: Future Cities Laboratory. Retrieved June 4, 2020, from https://ethz.ch/content/dam/ethz/special-interest/baug/ivt/ivt-dam/vpl/reports/1101-1200/ab1164.pdf
30. FHWA (2014). Certified Public Road Mileage Data and ARNOLD. Retrieved April 2021 from Office of Highway Policy Information https://www.fhwa.dot.gov/policyinformation/hpms/arnold.cfm
31. FHWA. Roadway Safety Data Program: MIRE. Retrieved from https://safety.fhwa.dot.gov/rsdp/mire.aspx
32. Ziering, E., Harrison, F., Scarponcini, P., Baker, M., and C.E. Campbell (2007). TransXML: XML Schemas for Exchange of Transportation Data. Washington, DC: Transportation Research Board, National Academy of Sciences. Retrieved from http://www.trb.org/Main/Blurbs/158531.aspx
33. Stabler, B. (March 2020). *The Open Matrix File Format (OMX) Documentation*. Retrieved July 29, 2020, from GitHub: https://github.com/osPlanning/omx/wiki
34. Castiglione, J., Mahut, M., Sun, W., Rousseau, G., Joshi, C., Frkonja, J., Gao, S. (n.d.). *General Travel Network Specification.* Retrieved 8 6, 2020, from Zephyr Transport: https://zephyrtransport.org/trb17projects/7-general-travel-network-specification/
35. Walton, C.M. et al. (2015). Implementing the Freight Transportation Data Architecture: Data Element Dictionary. Washington, DC: Transportation Research Board, National Academy of Sciences. Retrieved April 2021 from http://www.trb.org/Main/Blurbs/173083.aspx
36. AASHTO. Freight Demand Modeling and Data Improvement (C20). Retrieved from http://shrp2.transportation.org/Pages/Freight-Demand-Modeling-and-Data-Improvement-(C20).aspx
37. Cambridge Systematics, Inc., North River Consulting Group, and University of Washington (2013). Freight Data Sharing Guidebook. Washington, DC: Transportation Research Board, National Academy of Sciences. Retrieved from https://www.nap.edu/catalog/22569/freight-data-sharing-guidebook

38. FHWA. (2020, April 8). *Freight Demand Modeling and Data Improvement: A Strategic Roadmap for Making Better Freight Investments*. Retrieved from FHWA Freight Management and Operations: https://ops.fhwa.dot.gov/freight/freight_analysis/fdmdi/index.htm
39. Rensselaer: Center for Infrastructure, Transportation, and the Environment. Freight and Service Activity Trip Generation Software (FASTGS). Retrieved from https://cite.rpi.edu/software-and-tools/
40. Seedah, D. (2014). Retrieving Information from Heterogeneous Freight Data Sources to Answer Natural Language Queries.
41. Hemily, B. (2015). *Challenges in Deploying and Achieving the Full Potential of Transit ITS.* Washington, DC: USDOT.
42. GTFS: Making Public Transit Data Universally Accessible. Retrieved from https://gtfs.org/
43. Transmodel. Standard Interface for Real-time Information. Retrieved from http://www.transmodel-cen.eu/standards/siri/
44. Oregon Department of Transportation, Oregon State University. (n.d.). *GTFS Ride*. Retrieved June 4, 2020, from GTFS Ride: https://www.gtfs-ride.org/
45. USDOT (2009). APTA TCIP-S-001 - Standard for Transit Communications Interface Profiles (TCIP). Retrieved from Intelligent Transportation Systems Joint Program Office https://www.standards.its.dot.gov/Factsheets/Factsheet/37
46. TCRP G-18 (Ongoing). Improving Access and Management of Transit ITS Data. Washington, DC: Transportation Research Board, National Academy of Sciences. Retrieved from https://apps.trb.org/cmsfeed/TRBNetProjectDisplay.asp?ProjectID=4687
47. U.S. Department of Transportation Federal Highway Administration. (2017). *ATDM Brief: ATM Implementation and Operations Guide.*
48. Lambert, C. (April 2019). Kentucky Case Study for FHWA EDC-5 Crowdsourcing for Operations. (V. Shah, Interviewer).
49. Roberts, J. (March 2020). Arizona DOT Vehicle Probe Footprint. FHWA EDC-5 Crowdsourcing for Operations Vehicle Probe Peer Exchange. St. Louis, Missouri.
50. VanDyke, B. (March 2020). Oregon DOT Vehicle Probe Data Operations. FHWA EDC-5 Crowdsourcing for Operations Vehicle Probe Peer Exchange. St. Louis, Missouri.
51. Stone, T. (August 2017). *Traffic Technology International*.
52. Ellsmoor, J. (May 2019). *Smart Cities: The Future of Urban Development*. Retrieved June 2020, from Forbes: https://www.forbes.com/sites/jamesellsmoor/2019/05/19/smart-cities-the-future-of-urban-development/#6f45affd2f90
53. Sen, R. (n.d.). Smart Cities of the Future From Vision to Reality. Retrieved June 2020, from Deloitte: https://www2.deloitte.com/us/en/pages/consulting/solutions/smart-cities-of-the-future.html?id=us:2ps:3gl:confidence:eng:cons:31020:nonem:na:6sKjOFEK:1177105429:424633433650:e:Smart_Cities:Vision_To_Reality_Exact:nb
54. Smart Columbus. (n.d.). About Smart Columbus. Retrieved August 2020, from Smart Columbus OS: https://www.smartcolumbusos.com/about/about-smart-columbus
55. Portland Bureau of Transportation. (August 2020). Portland Urban Data Lake (PUDL). Retrieved from Portland Bureau of Transportation: https://www.portlandoregon.gov/transportation/article/681572
56. City of San Francisco. (n.d.). *DataSF*. Retrieved August 7, 2020, from DataSF: https://datasf.org/
57. Open Geospatial Consortium. (2020). Smart City Interoperability Reference Architecture. Retrieved from Smart City Interoperability Reference Architecture: https://scira.ogc.org/
58. Snap4City: Smart aNalytic APp builder for sentient Cities and IOT. Retrieved from https://www.snap4city.org/dashboardSmartCity/management/iframeApp.php?linkUrl=https://

www.snap4city.org/drupal&linkId=snap4cityPortalLink&pageTitle=www.snap4city.org&fromSubmenu=false
59. ATIS Data Sharing Framework for Smart Cities. Retrieved from https://www.atis.org/smart-cities-data-sharing/
60. The Smart Citis Data Exchange: An ATIS/US Ignite Initiative. Retrieved from https://www.atis.org/scde/#:~:text=The%20new%20ATIS%2FUS%20Ignite,sizes%20are%20invited%20to%20participate.
61. Open Connectivity Foundation: Unlocking the Massive Opportunity in the Internet of Things. Retrieved from https://openconnectivity.org/
62. FIWARE Foundation. (2020). FIWARE. Retrieved August 2020, from FIWARE: https://www.fiware.org/
63. Technical Research Centre of Finland. (2020). CityKEYS. Retrieved from CityKEYS. Retrieved from http://www.citykeys-project.eu/