ENTERPRISE Transportation Pooled Fund Study TPF-5 (231)

Assess Speed Data for Traffic Management

FINAL REPORT

Prepared by

Athey Creek Consultants

September 2017

Acknowledgements

This document was prepared for the ENTERPRISE Transportation Pooled Fund TPF-5(231) program (enterprise.prog.org/). The main purpose of ENTERPRISE is to use the pooled resources of its North America agency members and the United States federal government to develop, evaluate, and deploy Intelligent Transportation Systems (ITS).

Project Champion

Doug Tomlinson, Pennsylvania Department of Transportation, was the ENTERPRISE Project Champion for this effort. The Project Champion serves as the overall lead for the project.

ENTERPRISE Members

The ENTERPRISE Board consists of a representative from each of the following member entities of the program:

- Illinois Department of Transportation
- Iowa Department of Transportation
- Kansas Department of Transportation
- Michigan Department of Transportation
- Minnesota Department of Transportation
- Ontario Ministry of Transportation
- Pennsylvania Department of Transportation
- Texas Department of Transportation
- Transport Canada
- USDOT Federal Highway Administration

Interview Participants

Information presented in this final report was gathered through discussions with the following individuals:

- Wes Mauer, Colorado DOT
- John Williams, Colorado DOT
- Terri Johnson, HERE
- Denise Markow, I-95 Corridor Coalition
- Sinclair Stolle, Iowa DOT
- Ted Trepanier, INRIX
- Nancy Powell, Kansas City Scout
- Kristi Erickson, Kansas DOT
- Sharri Hilliard, Kansas DOT
- Elise Feldpausch, Michigan DOT
- Goran Nikolic, Ministry of Transportation Ontario
- Brian Kary, Minnesota DOT
- Alex Wassman, Missouri DOT
- Kelly Wells, North Carolina DOT
- Sal Cowan, New Jersey DOT
- John MacAdam, Ohio DOT
- Steve Gault, Pennsylvania DOT
- Eric Sponsler, Pennsylvania DOT
- Doug Tomlinson, Pennsylvania DOT
- Adam Clauss, Southwest Research Institute (Texas DOT)
- Rob Bamford, Transcom
- Michael Pack, University of Maryland
- Mike Fontaine, Virginia DOT
- Vinh Dang, Washington State DOT
- Liz Schneider, Wisconsin DOT

Table of Contents

1.0 Introduction

In recent years, probe data generated by fleet vehicles and travelers themselves is being aggregated by third-party data providers and offered to agencies as comprehensive, real-time speed data. Before third-party data became available, the costs associated with deployment and maintenance of detectors and sensors often limited the coverage area for which state and local department of transportations (DOTs could monitor traffic. The benefits of this new probe speed data may include statewide coverage without the need for wide scale deployment of sensors. As many agencies have begun to use real-time speed data from third-party providers, the uses of this data have increased.

This ENTERPRISE project "Assess Speed Data for Traffic Management" involved first conducting an online search and contacting individuals at agencies to gather information on the use of real-time third-party speed data in operations. Then these findings were analyzed to document and share common practices, challenges, and solutions to these challenges. Specifically, the objectives of this project included efforts to understand:

- Which DOTs are using the real-time speed data provided by third-party providers?
- How are DOTs using the data, especially for real-time traffic operations?
- What benefits are DOTs recognizing from the data?
- What challenges are DOTs facing (e.g., conflation of data) and how have these been addressed?
- How is speed data being combined with sensor data, and what is the potential for speed data to supplement a more limited deployment of sensors for volume collection?

This Final Report document presents the findings of this project and contains the following sections:

- 2.0 Research Approach
- 3.0 Context of Each Agency's Use of Speed Data
- 4.0 Current Uses of Real-Time Speed Data
- 5.0 Findings on Identified Issues and Potential Solutions
- 6.0 Findings on Identified Data Shortcomings
- 7.0 References

2.0 Research Approach

The research approach for this project consisted of two primary efforts: an online search and conducting interviews with agency practitioners and third-party data and service providers.

The online search used the Transportation Research Board (TRB) Transport Research International Documentation (TRID) database to identify published reports regarding the use of real-time third-party data in operations. Search terms included "INRIX", "Navteq", "TomTom", "probe data", and "third-party". Given the fast-paced evolution in the use of real-time third-party data in operations, search results were limited to those published since 2012. Remaining results were further reduced by removing research efforts that did not involve practices that were actually deployed or used by agencies in operations. Findings from this online search about agency practices and uses of real-time third-party data in operations are presented in the remaining sections of this report and were used to identify the agencies and entities to contact for additional information as a part of this project. Links to these findings are presented in section 8.0 References.

The second effort of this project involved conducting interviews with agency practitioners and third-party data and service providers. A question guide, provided in Appendix A, was developed to help focus the interviews and ensure that key topics of interest to this project were covered. Inputs and recommendations from ENTERPRISE members, third-party data providers, and findings from the online search effort of this project were used to identify agencies to contact. Points of contact were identified and contacted by e-mail to request a discussion. Overall, responses were received from practitioners at 17 out of the 25 agencies that were contacted, as well as individuals from 4 out of the 6 third-party data and data service providers. Agencies and providers that provided responses are listed in ***Table 1***. Key findings from these interviews are presented throughout the remaining sections of this report.

Table 1. Agencies and Third-Party Entities that Provided Responses for this Project

Agency Respondents	Third-Party Data or Data Service Provider Respondents
• Colorado DOT • Iowa DOT • Kansas DOT • Kansas City SCOUT • Michigan DOT • Ministry of Transportation Ontario • Minnesota DOT • Missouri DOT • New Jersey DOT • North Carolina DOT • Ohio DOT • Pennsylvania DOT • Texas DOT (SWRI) • Virginia DOT • Washington State DOT • Wisconsin DOT • Transcom	• HERE • I-95 Corridor Coalition • INRIX • University of Maryland-Center for Advanced Transportation Technology Laboratory (UMD-CATT Lab)

3.0 Context of Each Agency's Use of Speed Data

3.1 Agency Availability of Speed Data

The individuals contacted for this project represent agencies with access to various sources and combinations of speed data for use in operations, reflecting differences in agency policies and philosophies. Table 2 characterizes the nature of these differences, whether agencies use exclusively sensors, exclusively third-party data, or speed data from a combination of sources. Table 3 provides context for the findings in this report by listing the third-party providers currently used by each agency that was contacted for this project and their current use of sensors. Table 3 also identifies providers that agencies work with to develop and access tools for using the real-time third-party data, in addition to those offered by the third-party data provider. This includes UMD CATT Lab tools available through the Regional Integrated Transportation Information System (RITIS) and Probe Data Analytics Suite, the Iteris Performance Monitoring System (iPeMS), and those developed by other university partners and entities like Transcom. Note that UMD CATT Lab and Transcom were interviewed for this project, and both procure INRIX and Here data, while UMD CATT Lab also procures TomTom data. This information is not intended to be a comparison of third-party data providers used by agencies, but rather to observe general differences between agencies contacted for this process, including the use of sensors and multiple third-party data sources.

Table 2. General Categories of Speed Data Availability

Available Speed Data	Description	Contacted Agencies
Sensors only	Robust sensor network in place or not a need for third-party data at this time	• Washington State DOT • Minnesota DOT • Kansas DOT
Third-Party Data only	Statewide coverage, currently phasing out sensors or already have, new sensors only where third-party data is limited or latency needs to be minimized	• North Carolina DOT • Ohio DOT • Pennsylvania DOT
Third-Party Data supplements for specific projects	Procure third-party data only for limited mileage on specified roadways that lack sensors	• Wisconsin DOT
Third-Party Data and Sensor Network	Speed data expands coverage area and fills in the gaps of sensor network	• Iowa DOT • Michigan DOT • Missouri DOT • Texas DOT • Virginia DOT
Multiple Third-Party Data Sources and Sensor Network	2-3 sources of third-party speed data and still deploying sensors	• Colorado DOT • Ontario (Ministry of Transportation) • New Jersey DOT • Transcom (NYC-area 16-agency coalition)
Other	Third-party data and service providers, and research entities	• INRIX • HERE • University of Maryland CATT Lab • I-95 Corridor Coalition

Table 3. Availability of Speed Data and Tools as documented through Agency Interviews in this Project

Agency	Sensors	Third-Party Data				Additional Tool Providers[2]		
		INRIX	HERE	TomTom	Other	UMD CATT Lab	iPeMS	Other Partners
Colorado DOT	X	X						
Iowa DOT	X	X						X
Kansas DOT	X							
Michigan DOT	X		X			X		
Ministry of Transportation Ontario	X	X		X	X[1]			
Minnesota DOT	X							
Missouri DOT	X		X			X		
New Jersey DOT	X	X	X			X		X
North Carolina DOT			X			X	X	
Ohio DOT	X	X	X					
Pennsylvania DOT	X	X				X		X
Texas DOT	X	X						
Virginia DOT	X	X				X	X	
Washington State DOT	X							
Wisconsin DOT	X			X				

[1] GreenAisle Data

[2] Not comprehensive for all agencies as this was not an interview question; only known providers are listed

3.2 Third-Party Speed Data Maturity at the Agency

Understandably, agencies are at different places along a capability maturity spectrum regarding the uses of real-time third-party speed data. Agencies that have procured third-party data for many years tend to have more advanced applications and use the data in a variety of ways in operations. However, most agencies are in a less mature state with third-party speed data usage. These agencies tend not to use the data as often or in as many ways and are still developing capability maturity for third-party speed data. These agencies may have only recently begun procuring the third-party speed data or perhaps the third-party speed data is primarily used as a secondary resource to support other available data. Also, third-party data quality and real-time availability can vary from state to state.

At a minimum, every agency contacted that procures third-party speed data uses the color-coded speed map as a layer in the advanced transportation management system (ATMS) or other display on transportation management center (TMC) operator stations, a display on the TMC wall, or on traveler information maps. As capability maturity increases, additional applications are applied over broader areas and more roadways, up to a level that third-party speed data is used for fully automated operational actions.

Generally speaking, agencies contacted for this project fall into one of the five levels described below along a real-time speed data capability maturity spectrum beyond traditional sensor networks to larger datasets and coverage areas, as informally defined in this report and depicted in Figure 1. For the most part, the capabilities presented in each level add to the capabilities highlighted in previous levels.

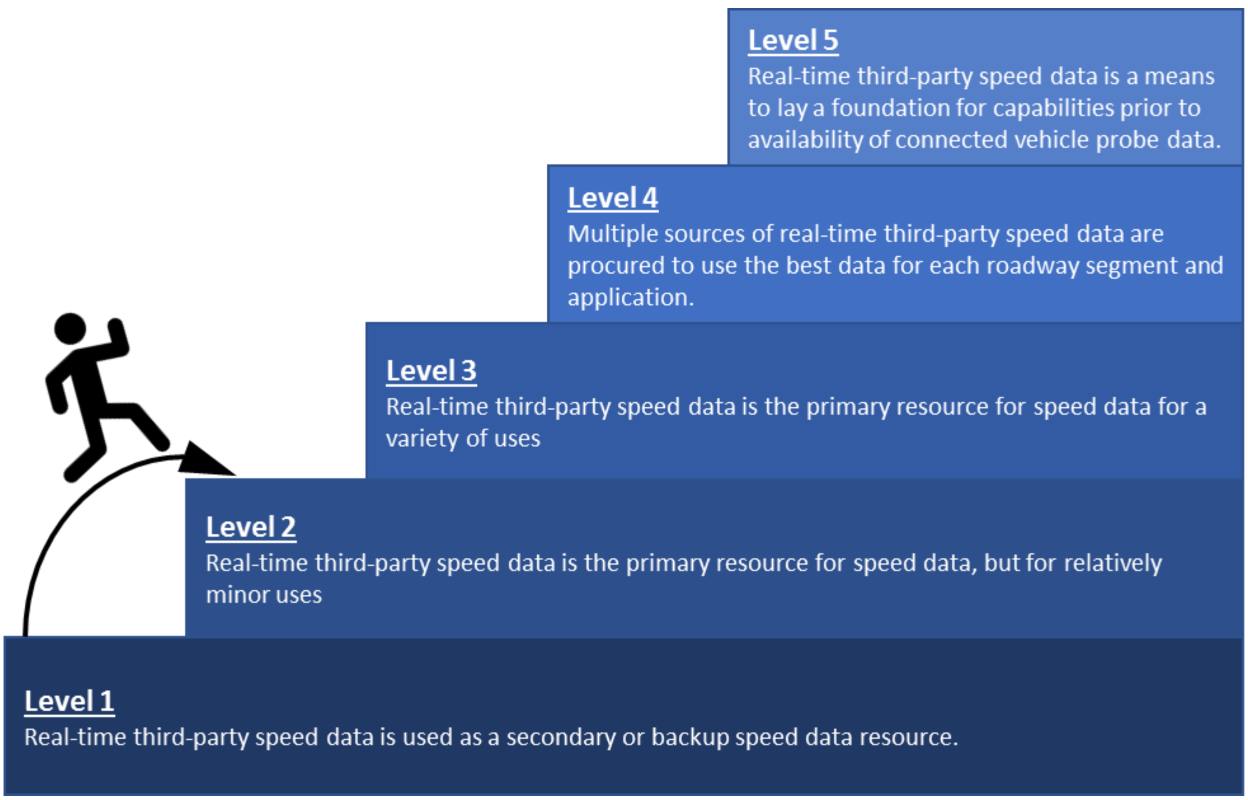

Figure 1. Informal agency capability maturity levels on the use of real-time third-party speed data.

Level 1. Real-time third-party speed data is used as a secondary or backup speed data resource. Agencies in this level generally agree the data is nice to have and may be used to confirm an incident reported by social media in areas that lack camera coverage or sensors. These agencies often use the third-party speed data for very specific items such as travel times on dynamic message signs (DMS) or to display speeds on the 511 map but rely mostly on speed data from sensors at the TMC, except to fill in gaps in the network that lack sensors.

Level 2. Real-time third-party speed data is the primary resource for speed data, but for relatively minor uses. Agencies in this level may monitor a smaller metro area or lower volume roadway network that has relatively fewer incidents with significant congestion impacts. Most agencies at this capability maturity level have integrated the real-time third-party speed data into the ATMS since it is the primary speed data source. Overall, however, operators tend to rely more on cameras for incident detection than monitoring the data or using any automated tools.

Level 3. Real-time third-party speed data is the primary resource for speed data for a variety of uses. Agencies in this level generally have operators conduct active monitoring statewide with the third-party speed data to identify slowdowns and incidents. To assist in these efforts, various tools may be available, such as a dashboard at the TMC to automatically identify slowdowns and provide information such as the end of queue location, duration of the congestion, and percent speed versus normal conditions. The agency may also use third-party speed data for automated applications in the field, such as work zone queue warning systems.

Level 4. Multiple sources of real-time third-party speed data are procured to use the best data for each roadway segment and application. Agencies in this level have a very good understanding of the strengths of each third-party speed data source and a variety of tools that leverage each of them. A data fusion engine may assist with the prioritization and blending of data sources for an informed, cohesive real-time speed data feed that provides the “better” data source for various areas and roadway segments. This process helps the agency to increase reliability and confidence in the data stream, while providing a backup in case of failures.

Level 5. Real-time third-party speed data is a means to lay a foundation for capabilities prior to availability of connected vehicle probe data. Agencies in this level take an even more holistic and forward-thinking view in their use of real-time third-party data combined with traditional and more innovative sources of speed data. While the benefits of third-party data are acknowledged, the advantages are seen as temporary as the connected vehicle environment is deployed to bring much greater potential and much larger datasets. Accordingly, the uses of real-time third-party data and ongoing investments for developing new automated tools from these sources may be secondary to investments in roadside infrastructure and tools to better understand and leverage data from the future connected vehicle environment.

4.0 Current Uses of Real-Time Speed Data

Agencies use real-time third-party speed data in a variety of ways, with some applications of the data being well-established, common uses, while others are more advanced applications that are used by a limited number of agencies. A high-level summary is provided for the widespread practices and more details are included for the less common, but innovative practices. Table 4 presents the agency uses of real-time third-party speed data that were identified through the practitioner interviews.

Table 4. Agency Uses of Third-Party Data as documented through Agency Interviews in this Project

Agency	Travel times on DMS	Map Layer for ATMS/TMC Display	Traveler Information Map	Identify/Confirm Incident	Alternate Route Information	Work Zone Queue Warning	TMC Dashboards and Tools to identify congested conditions	Custom Congestion Notifications
Colorado DOT	X	X		X				
Iowa DOT	X	X		X	X		X	
Michigan DOT	X	X	X	X			X	
Missouri DOT	X	X	X	X		X	X	X
New Jersey DOT	X	X	X	X			X	X
North Carolina DOT	X		X	X	X			
Ohio DOT	X	X	X	X				X
Ontario (Ministry of Transportation)	X	X		X				
Pennsylvania DOT	X	X	X	X	X		X	
Texas DOT	X	X	X					
Virginia DOT	X	X	X	X		X	X	
Wisconsin DOT	X	X			X			

4.1 Common Uses of Real-Time Speed Data

Nearly every agency contacted as a part of this effort used real-time third-party speed data for the following practices:

Automated display of travel times on DMS. Third-party speed data is often the primary source feed for posting travel times on DMS, but it is sometimes used as a source for roadway segments within the desired travel time distance that lacks sensor data. The travel times are automatically posted for each specified DMS for pre-defined segments at pre-defined hours. This involves integration of the third-party data into agency systems.

Map displayed at TMC. TMC operators at agencies with real-time third-party speed data typically have a map displayed that is color-coded with real-time speeds. This map may be displayed on a screen on the TMC wall or a monitor at the operator's work station. Further, this map may be a layer within the agency ATMS or viewed through the data provider's website. Generally, this map provides TMC operators with a much larger coverage area of speed data than a map populated using traditional sensor data. Additionally, the map may allow operators to view the speed data in different ways to separate recurring congestion from non-recurring issues. For example, speed data may be shown as a percentage to compare current conditions versus typical conditions, where even slow-moving traffic would be displayed as green if that level of congestion is normal at that time of day and location. Note that in some cases, operators may have a personal preference to use a free, public source of this information, such as Google Maps.

Traveler information map. Most agencies maintain a traveler information website and, similar to the map displayed at the TMC, most use a source of third-party data to provide traffic speeds as a color-coded layer. Some agencies have a separate license agreement in place with a provider to display publicly-available Google Maps and Traffic on the traveler information website. This agreement may have been made prior to procuring a full real-time third-party dataset, and it may be a lower cost to continue this approach than integrate the third-party data into the traveler information website. As such, the source of speed data on the traveler information map is not always the same as the source for the map displayed at the TMC.

Confirming an incident. TMC operators generally have many sources of data available to identify an incident and typically require verification before taking action and posting DMS messages, for example. While the preferred source of data to identify or verify incidents may be cameras, some areas lack coverage. In this case, third-party speed data may be examined to identify reduced speeds in the area of an incident that is reported on social media or detected via other means. This use requires the map display at the TMC for use by operators, as detailed above.

4.2 Less Common Uses of Real-Time Speed Data

The practices listed below are less commonly used, with few practitioners interviewed for this project identifying the use of these applications within their agency. Note that the applications themselves are not uncommon, rather it is the use of real-time third-party speed data for these applications that is less common.

TMC dashboards for identifying congestion. A number of dashboards are available and used by operators at many agencies. These TMC dashboards vary in the exact information that is presented to operators, but typically automatically generate a list of roadway segments that are experiencing reduced speeds for operators to examine and take action. The list may be populated based on a percentage of the real-time reduced speed versus typical speed (instead of free flow speed) to highlight areas of potential incidents while ignoring recurring congestion. The dashboard may identify the end of the queue, duration of the slowdown, the real-time travel time delay caused by the slowdown, and/or a percentage to indicate the real-time speed versus either free-flow or typical speeds on that segment.

Purdue University developed a dashboard for the Indiana DOT which uses INRIX data to identify congestion. This dashboard has since been adapted and is now offered by INRIX as a tool called Dangerous Slowdowns. Purdue University has also demonstrated the use of a dashboard for incidents and winter weather conditions on interstate highways. A screenshot of a TMC dashboard that was developed by Iowa State University for Iowa DOT is shown in **Error! Reference source not found.** and includes a list of “Active Bottlenecks” developed from real-time third-party speed data.

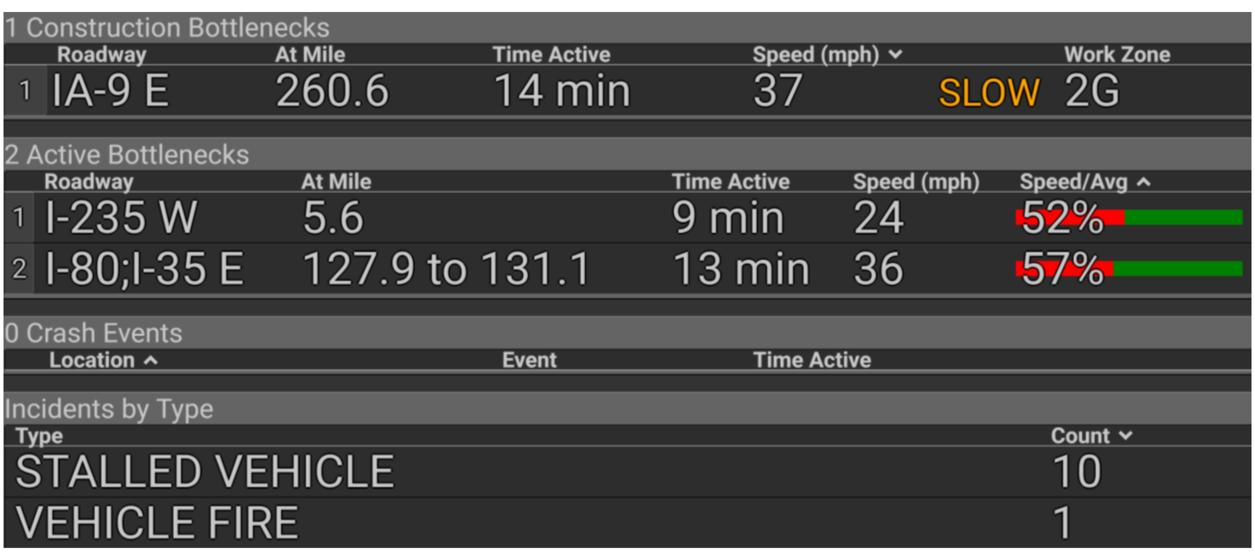

Figure 2. Iowa DOT TMC dashboard that uses third-party speed data to generate a list of “Active Bottlenecks”.

The University of Maryland Center for Advanced Transportation Technology Laboratory (UMD-CATT Lab) Probe Data Analytics Suite provides a number of tools for agencies including a bottlenecks tool shown in Figure 3. This Suite also includes a Work Zone Performance Monitoring Dashboard that was initially developed for the Maryland State Highway Administration that looks specifically at real-time queues in active work zones, including the length of the queue and whether the queue is increasing or decreasing in length as shown in Figure 4, as well as incurred user delay costs and metrics of top critical work zones. UMD-CATT Lab tools are available to the I-95 Corridor Coalition members, including New Jersey DOT, North Carolina DOT, Pennsylvania DOT, and Virginia DOT, and are also used by Michigan DOT and Missouri DOT. Additionally, the Iteris Performance Monitoring System (iPeMS) includes a tool to identify bottlenecks, using real-time third-party data and other speed data sources, as available; iPeMS is used by North Carolina DOT and Virginia DOT. Virginia DOT is conducting additional research to improve the identification and characterization of bottlenecks.

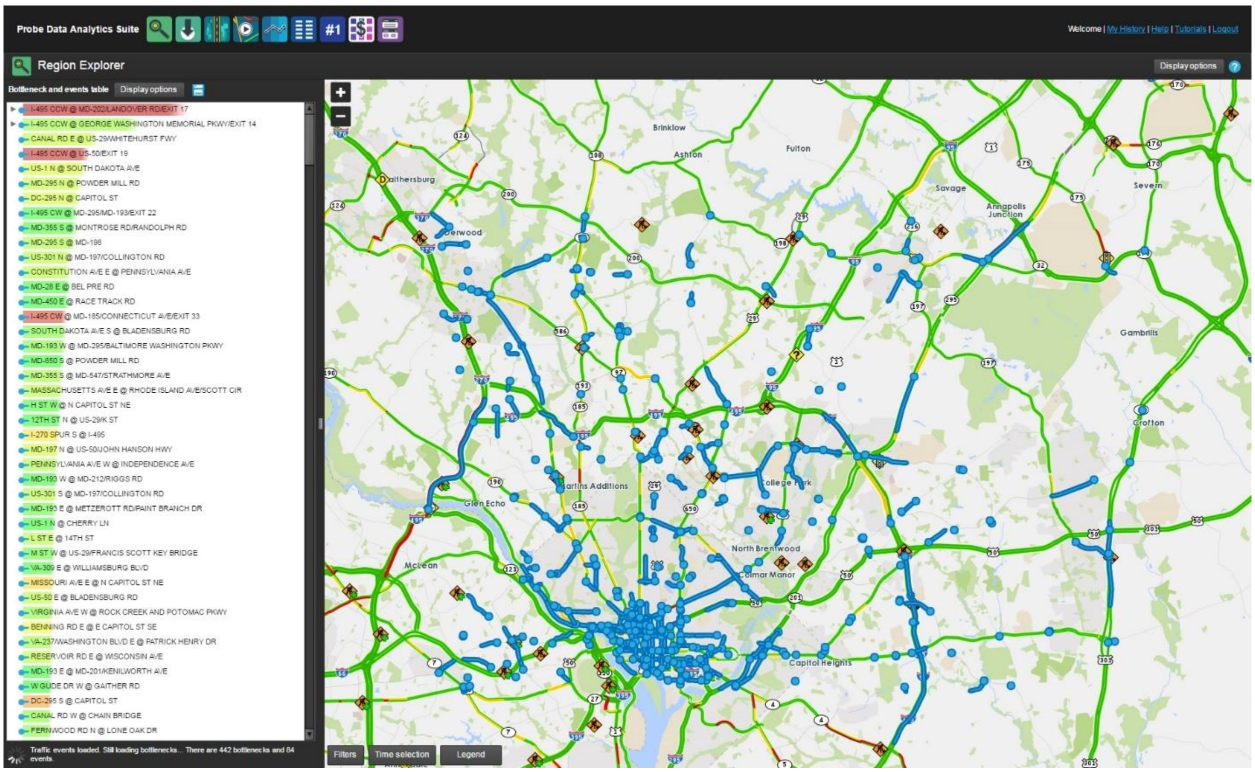

Figure 3. Bottlenecks identification dashboard in the University of Maryland-Center for Advanced Transportation Technology Laboratory (UMD-CATT Lab) Probe Data Analytics Suite.

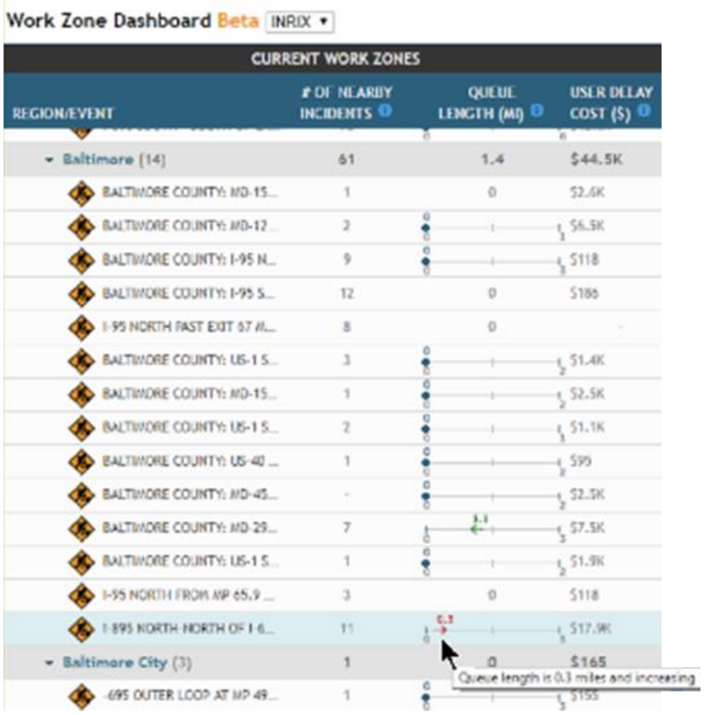

Figure 4. Work Zone Performance Monitoring Dashboard developed by the University of Maryland-Center for Advanced Transportation Technology Laboratory (UMD-CATT Lab) for the Maryland State Highway Administration (SHA).

Automated display of alternate route travel times and recommendations. This practice is similar to the automated display of travel times and is often used to reduce the mobility impacts of a work zone or at locations with two comparable routes. Third-party speed data enables coverage of two routes that travelers may take (e.g., an interstate and secondary roadway or a freeway through a downtown area and bypass freeway) to provide comparative travel times on permanent or portable DMS, or static signs with dynamic numerals. Specifically, the Wisconsin DOT uses third-party data only for travel times on the secondary roadways and uses an existing sensor network on the freeway. An advisory message could also be automatically posted to a DMS for travelers to take the detour route when extreme congestion reaches a pre-defined threshold on the primary highway, and the secondary route is uncongested. North Carolina DOT has also implemented this practice in smart work zones, both to post comparative travel times and also to verify alternate route performance before and while posting recommendations on DMS.

Work zone queue warning system. Real-time third-party speed data can be used in lieu of deploying temporary sensors to monitor speeds and provide information to travelers on a portable DMS about slow or stopped traffic upstream of a work zone. This practice has been tested by Missouri DOT and Virginia DOT with promising results, although Missouri DOT noted that operator intervention was often necessary to change the displayed message after the queue dissipated as the speed data was slow to reflect the improved conditions.

Notifications when slowdowns occur on pre-selected segments. Real-time third-party speed data is also used to generate automatic notifications when speeds drop below a pre-defined threshold for pre-selected segments. These automatic notifications can be sent by the agency systems to DOT staff and contractors. The Missouri DOT, for example, has an in-house text message alert system that is mostly used by construction groups or area engineers responsible for a large area. These individuals can select probe data segments and apply rules for notifications, e.g., traffic speeds below 20 mph for 15 consecutive minutes in segments adjacent to a work zone. Notifications are then transmitted via text or email message through a custom-built web interface.

Traveler information application travel delay features. Traveler information applications feature real-time updates to drivers about slowdowns and incidents. Similar to the previous application, travelers may be able to receive notifications in real-time about slowdowns and delays on pre-defined routes at specified times based on user entries about their typical commute or trips. These visual or audio notifications may be pushed to the user's mobile device before or during the trip, and include recommended changes to the route based on delays and incidents. A similar feature is available on the Ohio DOT traveler information application OHGO, which allows users to see the estimated travel delay on all "red" speed segments of the map and along their route, as shown in Figure 5.

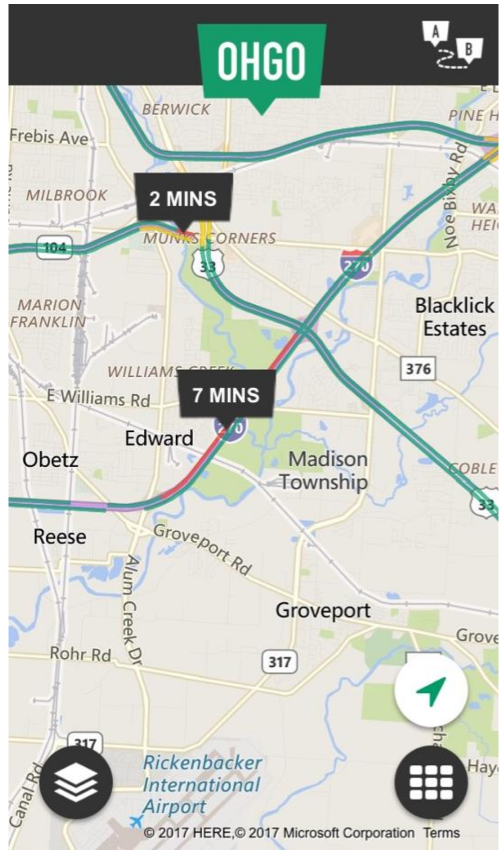

Figure 5. The Ohio DOT traveler information application OHGO can display travel delay on red speed segments.

5.0 Findings on Identified Issues and Potential Solutions

Agency usage of third-party speed data has dramatically increased over the last decade since becoming available, during which time many agencies have encountered challenges when using the data. While some challenges have been overcome by some agencies, other agencies may still be trying to work through these issues. This section presents findings related to issues that were identified and discusses the various approaches and solutions that agencies have developed to overcome these challenges. The issues and corresponding potential solutions are organized in this section as shown in Table 5.

Table 5. Identified Issues and Corresponding Potential Solutions

Issue	Potential Solution
#1: Justifying the Purchase	#1: Identify and compare the costs to deploy, operate, and maintain the existing sensor network.
	#2: Identify and quantify benefits of applications or scenarios using real-time third-party data that are outside of existing sensor coverage areas.
	#3: Identify other potential users of real-time (or historic) third-party data within your agency to strengthen the case for the purchase.
	#4: Tailor the procurement to meet specific agency needs.
	#5: Contact neighboring and peer agencies that use real-time third-party data to get a better sense of the costs and benefits of procuring and integrating the data for operations.
#2: Data Quality	#1: Request a data sample as part of the competitive procurement process.
	#2: Incorporate data quality requirements as part of the procurement contract.
	#3: Conduct a formal data quality check before expanding its use.
	#4: Conduct periodic, formal data quality checks.
	#5: Use as a secondary data source to compare and gain comfort over time.
	#6: Review available studies conducted by others to understand data quality and limitations.
#3: Confidence Scores	#1: Examine the cause of low confidence scores and understand whether this matters for how your agency is using the data.
	#2: Develop Business Rules for disregarding data with low confidence scores.
#4: Data Latency	#1: Understand the variations and causes of data latency in order to respond appropriately.
	#2: Understand the impacts of data latency on automated systems in order to incorporate operator oversight or deploy sensors, as necessary.
#5: ATMS Integration	#1: Use the third-party provided web interface.
	#2: Integrate third-party data as a distinct layer.
	#3: Assign priority data sources for display on the speed layer.
	#4: Upgrade the ATMS software.
#6: Conflation	#1: Incorporate software with automated processes.
	#2: Require third-party data providers to supply data that conforms to agency segmentation.
	#3: Remove historic agency segmentation needs.
	#4: Retain support consultants with expertise in third-party data.
#7: Appropriate Roles of Third-Party Data	#1: Use third-party data to fulfill the initial identified needs, then incrementally test and expand use to other applications over time.
	#2: Conduct analyses to identify tradeoffs of investing in traditional or other innovative systems versus expanding third-party data usage.

5.1 Issue #1: Justifying the Purchase

Identifying and documenting the qualitative or quantitative justification for purchasing third-party data can be a challenge. The purchase and integration costs for real-time third-party data can be expensive and, at times, has resulted in agency staff being unable to secure agency decision-maker support. All agency staff who were contacted as part of this project identified benefits of procuring real-time third-party data, regardless of how extensive the data was used in operations. Nonetheless, feedback from multiple agencies revealed challenges faced internally to receive support for purchasing third-party data.

Although access to real-time data may appear "obviously" beneficial, it can be much more difficult to quantify the actual benefits. The issue goes beyond a need to understand the benefits and compare them to costs. Many agencies do not have clearly defined parameters describing what benefits compared to costs need to be achieved to justify the purchase. Their responses reveal several approaches that practitioners might take to identify the potential costs and benefits of real-time third-party data for operations, as well as other agency functions, compared to existing practices.

Potential Solution #1: Identify and compare the costs to deploy, operate, and maintain the existing sensor network. Many agencies have an existing sensor network to support operations in high-traffic urban areas. In short, the Ministry of Transportation Ontario notes that collecting your own data is expensive. More specifically, Missouri DOT noted the benefit of not having maintenance costs, including the costs of equipment errors that require staff attention.

While agencies with real-time third-party data may still require sensors in specific areas for functions such as dynamic shoulder use, managed lanes, or ramp metering, others have scaled back or eliminated sensors in the field as a result of having real-time third-party data. North Carolina DOT, Ohio DOT, and Pennsylvania DOT have either removed sensors from the field or have stopped installing new ones. Ohio DOT noted that third-party data is half the cost of having to operate and maintain roadside devices, and provides much better coverage. Pennsylvania DOT noted that third-party data is equivalent to statewide sensor data, without being limited by equipment.

Some agencies have found that the expanded coverage area has pre-empted the need to deploy new sensors in areas of growing congestion. Virginia DOT noted a net effect of reducing the need for physical infrastructure, having stopped maintaining some sensors and abandoning some that were not used to control specific devices, while also choosing not deploy sensors in some rural locations because of the third-party data. However, some areas with low third-party data accuracy created a need for new sensors.

Potential Solution #2: Identify the need and quantify benefits of applications or scenarios using real-time third-party data that are outside of existing sensor coverage areas. A major benefit of real-time third-party data is having much more granularity across an expanded coverage area than sensors can provide. This, in turn, can provide much better information for identifying and managing congestion in a variety of scenarios.

Regardless of any shortcomings of third-party data, Iowa DOT notes that having data that is lower quality or higher latency is better than having no data at all. Iowa DOT uses the third-party data to more quickly verify reports of incidents from social media if CCTV cameras are not in the area and law enforcement are not yet on scene. Additionally, although Iowa DOT still deploys sensors for queue warning systems in rural work zones, the third-party data is used to verify the accuracy of those devices. Other agencies, including Missouri DOT and Virginia DOT have deployed queue warning systems in rural work zones that rely exclusively on third-party data, eliminating the expense of sensors.

The Wisconsin DOT found third-party data to be a cost-effective way to monitor traffic on alternate routes during a major, 40-mile interstate construction project in a rural area. While sensors were present on the freeway, the alternate routes that were managed by a variety of local agencies that did not have sensor coverage. A memoranda of understanding (MOU) would have been required with each of these local agencies to install sensors that would only be used for a couple years. Instead, the Wisconsin DOT third-party data contract with TomTom provides flexibility to specify and change which roadways they want data for, without having infrastructure in the field after the project concludes.

Finally, the Ministry of Transportation Ontario notes that because demands on highway capacity are growing, the agency is willing to procure data to better facilitate additional automated processes that improve transportation network performance.

Potential Solution #3: Identify other potential users of real-time (or historic) third-party data within your agency to strengthen the case for the purchase. Both real-time and historic third-party data have a variety of uses in a number of agency programs, including planning, construction, signal timing, performance management, before/after studies, research, and evaluations. Districts or TMCs within the agency may already procure third-party data and these contracts could potentially be modified to expand the coverage area. Existing contracts by other agency departments for historic data might also be leveraged and modified to procure real-time data at a lower cost than a separate contract. The Indiana DOT used third-party data for signal timing, construction planning, and performance management, and the construction department was able to leverage this data for work zone management applications.

Missouri DOT found benefits from real-time third-party data to improve consistency across the state for performance measurement reporting. Before using third-party data as a statewide data source, different districts had different coverage levels and data quality which affected the consistency performance measurement reporting.

Potential Solution #4: Tailor the procurement to meet specific agency needs. Real-time third-party data can be used in many applications, but agencies do not need to make changes to all systems if existing processes meet their needs. For instance, Iowa DOT chose to continue using Google Maps on their traveler information website to display real-time traffic speeds because that was better information and there would be additional costs to integrate the third-party data into the website. As another example, TMC operators at North Carolina DOT prefer using the Bing Maps interface despite having access to third-party data.

Wisconsin DOT wanted to use real-time third-party data only for a specific construction project and secured a contract to receive speed data for limited roadway mileage, instead of a larger statewide contract. Similarly, Indiana DOT had a contract to receive real-time third-party data for all interstate miles plus 200 miles of any non-interstate routes, allowing flexibility for covering work zones or special events, while having lower costs than a contract for statewide coverage.

It is also not necessary to have the same provider for all third-party data applications. For example, the Ohio DOT maintains three separate contracts: a contract with HERE provides real-time speed data for the interstate system, while there are two contracts with INRIX to provide coverage on non-interstate routes and historic data.

Potential Solution #5: Contact neighboring and peer agencies that use real-time third-party data to get a better sense of the costs and benefits of procuring and integrating the data for operations. As an agency explores the needs and potential uses of procuring third-party data, neighboring agencies can be a valuable resource. A site visit or peer exchange webinar could help an agency understand the costs of integrating third-party data, as well as benefits and uses that had not been considered.

Additionally, agencies sometimes work together as part of a cooperative agreement or coalition to procure real-time third-party data. The I-95 Corridor Coalition and Transcom are examples of entities that procure third-party data and perform various functions like integration, data quality and validation analyses, and tool development for the benefit of all member agencies. These arrangements can help reduce the financial and staffing challenges that individual agencies may encounter when procuring and managing third-party data. Agencies might consider inquiring with neighboring agencies, including local agencies and adjacent state agencies, about joining or developing a cooperative agreement to facilitate third-party data.

Outstanding Questions and Considerations that Relate to this Issue

- Real-time third-party data provides agencies with a much larger dataset that may be more challenging to manage than data generated from agency sensor networks, as noted in section 6.2 Data Quantity. Understanding how to effectively leverage and use real-time third-party data may help an agency prepare for the connected vehicle environment, which is expected to be a source of even more data.
- Agencies with an extensive sensor network may be reluctant to shift away from this investment if it works well for them, particularly if their network would require sensors in certain areas for strategies like dynamic shoulder lanes, managed lanes, and ramp metering. These agencies may find limited value in expanding to new areas or roadways if it is felt the existing problem spots are sufficiently covered.
- An agency with limited staff and resources to integrate third-party data into agency systems and then actively monitor may not find this to be a valuable purchase, particularly if the state is mostly low-volume, rural roadways. An extensive dataset should not be purchased if no uses for it have been identified or it will be rarely used.
- What are the various applications and uses of real-time third-party data that agencies can effectively leverage in a TSMO-centric environment? Understanding how to effectively leverage and use real-time third-party data may help justify the purchase.

5.2 Issue #2: Data Quality

Having a good understanding of the data quality is important and agencies report using a variety of methods to ensure they receive consistent and accurate data, while being aware of data limitations. The quality of real-time third-party speed data continues to improve, but agencies should understand the strengths and weaknesses, availability, and relative accuracy of the data. These may vary by roadway type (e.g., freeway, arterial), location (e.g., rural or urban), traffic volume, and time of day. At the same time, agencies might consider how accurate and complete the third-party data needs to be for the ways it will be used; for example, automated tools and applications in operations require higher quality data than is needed for a speed map that TMC operators can use as a secondary resource to confirm an incident.

Data quality checks may occur before procurement and throughout the contracting period and may be conducted with varying levels of formality and scope. A formal data quality analysis may compare the relative accuracy of the data samples using a floating car with documented travel times (Ohio DOT, Wisconsin DOT); deployed sensors and cameras for ground truth (Iowa DOT, Missouri DOT, Texas DOT, Virginia DOT, I-95 Corridor Coalition); or even against other third-party data sources like Google Maps (Wisconsin DOT).

Potential Solution #1: Request a data sample as part of the competitive procurement process. In order to select the best provider of real-time third-party speed data, agencies like Iowa DOT and Wisconsin DOT have found it beneficial to require potential bidders to submit a sample of data with their proposal to compare the relative quality of each dataset. This may be a dataset for a specific duration of time, specific dates, roadways, and/or data elements. Alternatively, the agency might be provided temporary access to a data feed. Agencies may conduct the data quality analysis in-house, with university partners, or hire a support consultant. However, Ontario Ministry of Transportation cautions that accuracy is hard to determine and provided samples can give different results that are hard to compare.

Potential Solution #2: Incorporate data quality requirements as part of the procurement contract. Agencies may specify in the contract that real-time third-party speed data providers maintain quality standards, including thresholds for quality, the frequency and nature of data quality checks, and penalties for failing to meet the established quality standards. Missouri DOT, Ohio DOT, and the I-95 Corridor Coalition all include data quality requirements as a part of their contracts with the data providers. Missouri DOT validates the real-time third-party speed data against sensors, while Ohio DOT staff conduct random data checks using a floating car. The Ohio DOT contracts require the speed data to be accurate within 4 miles per hour for 75 percent of the samples, and data availability greater than 99 percent. Ohio DOT notes instances when one of the third-party data providers did not pass these data quality tests. The I-95 Corridor Coalition request for proposals for third-party data includes details about validation processes. For example, data quality requirements are mandatory for freeways with flows exceeding 500 vehicles per hour, and must have a data latency less than or equal to eight minutes and a maximum average absolute error of eight miles per hour for each of four specified speed ranges. Data collection processes using Bluetooth sensors and calculations to compare third-party and ground truth speed and latency are also included, as well as the resulting impacts to payment for data that does not meet the specified requirements. Additionally, as discussed in more detail as part of Section 5.3, real-time third-party data providers generally provide an indication of the data quality, i.e., confidence score, as it relates to the mix of historic data versus real-time data. Agencies may use this confidence score as a measure of data quality for determining whether or not to use the data in operations for various roadways or applications.

Potential Solution #3: Conduct a formal data quality check before expanding its use. It is common for agencies to roll out different uses and applications of real-time third-party data over time given resource limitations and a desire to gain comfort and confidence in the data. For example, agencies may initially use real-time third-party speed data for heavily-traveled urban and suburban interstates before expanding usage to rural interstates, other freeways, and arterials in both urban and rural areas. Likewise, agencies may initially use real-time third-party speed data for monitoring or to verify an incident before more automated tools and applications are tested and deployed. For both the initial and expanded uses of real-time third-party speed data, agencies may conduct a formal data quality check to confirm accuracy and understand data shortcomings.

There are many examples of this approach. For instance, Transcom conducted an independent analysis of both HERE and INRIX data for the New Jersey DOT and is continuing to conduct analyses for arterials. This analysis has revealed that it can take several days for the third-party data to re-calibrate after agencies adjust traffic signal timings. Both the Iowa DOT and Nebraska DOT relied on a university partner to initially validate the data and review data coverage (i.e., percent real-time data statewide), latency (i.e., identification of events with data versus cameras), and the observed differences in speed between the data and sensors. Iowa DOT is building on these findings as the data is used in different ways. The Colorado DOT hired a support consultant to analyze INRIX data and other speed data sources. Additionally, the I-95 Corridor Coalition has continued to conduct data quality and verification analyses for a variety of roadways, including tunnels and arterials.

Virginia DOT conducted a significant number of studies to validate third-party data beginning in 2010, verifying data accuracy for all interstates in Virginia over the years, including at night and on rural interstates and specific arterial routes. As part of these studies, Virginia DOT has conducted studies where there were concerns about problem locations. They also deployed Bluetooth sensors to improve the data comparisons and better understand the data shortcomings. Virginia DOT has a good understanding of arterial characteristics that negatively impact the data, including super-dense access points and over-saturated conditions. Before expanding the arterial data displayed on the 511 website, Virginia DOT conducts a high-level analysis on those routes based on several weeks or months of data.

Potential Solution #4: Conduct periodic, formal data quality checks. Some agencies conduct regular data quality checks at specified periods of time. This may be done to verify data quality per agency policy, per the contract with the third-party data provider, or per a multi-agency collective agreement. For example, Transcom, the coalition of 16 agencies in the New York City metro area that includes the New Jersey DOT, New York State DOT, and Connecticut DOT, conducts a formal data quality check of the third-party data every two years for the member agencies. Similarly, the I-95 Corridor Coalition conducts periodic data quality and validation studies for the member agencies on various roadways around their region. Agencies like North Carolina DOT and Pennsylvania DOT rely on these studies for data validation.

Potential Solution #5: Use as a secondary data source to compare and gain comfort over time. Depending on how an agency intends to use the real-time third-party data, a formal analysis may not be necessary. For example, Texas DOT operators in each district initially viewed the real-time third-party speed data in a "shadow mode" in order to casually compare and validate the data before it was fully launched in each district, as each felt comfortable with the data quality. Note that the Texas DOT has since initiated a formal analysis. Similarly, MTO procures and uses three real-time third-party data sources side-by-side for operators to compare and verify the quality.

Potential Solution #6: Review available studies conducted by others to understand data quality and limitations. One way for agencies to understand the strengths and limitations of real-time third-party speed data is to review studies that other organizations have conducted. For example, the I-95 Corridor Coalition has been conducting a Vehicle Probe Project since 2008 and posts data quality and validation studies of INRIX, HERE, and TomTom data on the project website that are conducted for the member agencies. Agencies outside of the I-95 Corridor Coalition can view these reports to develop an approach for conducting a local analysis, or as a basis for understanding data shortcomings that may require attention and precautions when using the data. However, the Iowa DOT has noted that the data quality on Iowa roadways is not always consistent with what is reported in I-95 Corridor Coalition reports, e.g., there is a higher latency in Iowa of up to 5-10 minutes.

Outstanding Questions and Considerations that Relate to this Issue

- What are the appropriate levels of data quality and validation analyses, which may depend on how data will be used? Similarly, because data quality varies by geographic area and roadway type, what is an acceptable level of data quality for various areas and applications to meet agency needs?
- Can the data quality of real-time third-party speed data be quantified and analyzed in a standard way based on agency-specific needs? What standard measures might be used to define a minimum acceptable quality level, standard quality level, and high-quality level? An example quality level might define speed accuracy to be within +/- 5 mph on all freeways, U.S. highways, and state routes between 4:00 a.m. and 11:00 p.m. with a latency of no more than 3 minutes.
- Agencies may find it challenging to compare third-party datasets to one another.

5.3 Issue #3: Confidence Scores

Agencies report having different strategies for using data with less than maximum confidence scores, which indicate a blend of historical data in the speed data and thus reduced value in real-time operations. Real-time third-party data providers generally have a measure associated with each speed measurement to indicate the quality of that data. For example, INRIX provides confidence scores, where a score of 30 being entirely real-time data and lower scores indicating that the speed measurement is a blend of historical and real-time data or the historical average.

While some agencies have found low confidence scores to indicate low traffic volumes other agencies call into question the validity of data with lower confidence scores. As an example for an incident during off-peak hours, the absence of real-time probe data due to a road closure could result in the use of historic data incorrectly showing that segment operating at free flow conditions.

Potential Solution #1: Examine the cause of low confidence scores and understand whether this matters for how your agency is using the data. Some agencies continue to use real-time third-party data, despite low confidence scores. Both Missouri DOT and Ohio DOT note that in their experiences a low confidence score is generally due to low traffic volumes, such that historical travel time would be accurate. Wisconsin DOT echoed this, having never observed historic data being used under congested conditions, but only mixed in when there is not enough traffic to cause congestion.

Potential Solution #2: Develop Business Rules for disregarding data with low confidence scores. Other agencies have taken precautions to remove data that does not meet specified thresholds. Pennsylvania DOT does not use any data with an INRIX confidence score less than 30 in their ATMS. The Ministry of Transportation Ontario developed an approach to flag and disregard all data with less than a 95 percent confidence level. Virginia DOT developed business rules based on an empirical policy decision for displaying travel times on DMS that works well for the agency. Specifically, the rule requires 85 percent of the roadway segments in the interval to have an INRIX confidence score of 30.

Outstanding Questions and Considerations that Relate to this Issue

- How often does the incorporation of historic data signify a phenomenon other than a low volume of traffic?
- Aside from low volume traffic, what factors or events may cause a low confidence score to occur and how can an agency identify those instances (e.g. an unplanned road closure that diverts all thru traffic to other roadways)?

5.4 Issue #4: Data Latency

High data latency may reduce responsiveness and effectiveness of certain applications. Timeliness is important in real-time operations to identify problems on the roadway faster, reduce response times, prevent secondary incidents, and improve mobility. Higher third-party data latency could thus result in delays in providing accurate information to the traveling public.

Some applications, such as dynamic shoulder lanes, managed lanes, and variable speed limit systems require low data latency to provide accurate information to drivers, and agencies continue to deploy sensors for these applications. Other systems like dynamic queue warning systems have been successful in areas with sufficiently low data latency.

Potential Solution #1: Understand the variations and causes of data latency in order to respond appropriately. Agencies report varying levels of data latency, with lower volume roads in rural areas generally having higher data latency. Specifically, INRIX notes that in addition to the density of probe data, which is based on the volume of traffic and types of vehicles, that areas of cellular dead zones with only one or two carriers providing coverage can also increase latency. Understanding the locations and reasons for data latency may inform agency decisions to deploy certain types of applications in specific areas, allowing for better use of the data where the data performs well.

Overall, Iowa DOT noted that data latency is higher than values reported by the I-95 Corridor Coalition for states on the east coast. Iowa DOT worked with a university partner to conduct a study to understand the data latency and identify appropriate uses of the third-party data. The study examined sensor speed data and operator identification of events using cameras to measure latency of the third-party data. On the other hand, Ohio DOT believes that the manner in which their ATMS ingests the third-party data could be introducing some delays.

Potential Solution #2: Understand the impacts of data latency on automated systems in order to incorporate operator oversight or deploy sensors, as necessary. In some locations, data latency may be high enough that additional oversight or sensors are necessary for monitoring or specific applications. Iowa DOT concluded that the third-party data with a 5-10 minute latency was too high for a work zone queue warning system and chose to use sensors, with the third-party data available as a secondary source to validate the sensor data. On the other hand, the Missouri DOT deployed a queue warning system and found that relatively higher data latency was acceptable even though some traffic would pass the DMS without seeing a warning as a queue began to form. In essence, the majority of traffic impacted by a queue would see the message. Missouri DOT found that a bigger issue was the latency in the data after incident was cleared, which required operators to manually intervene to change DMS displays.

Outstanding Questions and Considerations that Relate to this Issue

- What is the acceptable data latency for work zone queue warning systems, variable speed limit systems, and other applications?
- What is the relative impact of timeliness to convey a message to all travelers versus the ability to convey information that is slightly delayed but reaches most travelers?

5.5 Issue #5: ATMS Integration

Integration of third-party data feeds into an agency's ATMS was cited as a major expense and a major effort by some agencies. In order to use procured third-party data in an effective manner, agencies typically work to integrate the data feeds with agency systems, including the ATMS. This process may be more challenging for agencies that use older systems, have multiple sources of speed data, or segmentation that differs significantly from the third-party data.

Iowa DOT noted that mapping the third-party segments into the ATMS was the biggest challenge, and Wisconsin DOT stated that integration took months because their ATMS software is older. Upgrading the ATMS software may not be practical in the short-term, and agencies have taken other approaches to make the third-party data work with their existing systems. The integration, use, and display of multiple sources of speed data is generally not an issue, but agencies have taken different solutions to display the real-time third-party speed data alongside other data sources. For all of the potential solutions listed below, regular and detailed communications with the third-party provider will help facilitate this process. Pennsylvania DOT offers a lesson learned to make sure the ATMS vendor communicates with the third-party provider, including during upgrades to better understand what will change. Overall, this will help minimize software compatibility issues.

Potential Solution #1: Use the third-party provided web interface. Third-party providers typically have a web interface available for agencies to use. However, when TMC operators monitor traffic while managing devices through an ATMS, this may not be the ideal approach. The North Carolina DOT does not have an ATMS and instead views third-party data using the web interface supplied by the provider. Due to anticipated challenges with integrating the HERE data feed with agency systems, North Carolina DOT opted to integrate the free Bing service for traveler information. However, challenges due to Bing business rules have resulted in North Carolina DOT seriously considering integrating the HERE data feed. Further, while operators must use the HERE interface to verify an incident or slowdown before taking action, they prefer Google Maps for monitoring traffic speeds.

Potential Solution #2: Integrate third-party data as a distinct layer. For agencies that only use a single source of third-party speed data and do not have sensors, integration of the third-party data into the ATMS will be done as a distinct layer on the user interface. However, if multiple speed data sources are available, the third-party data may be integrated as a distinct layer that is separate from speed data from sensors or other providers. In this case, TMC operators generally have a preferred data source for monitoring traffic speeds, and manually switch between layers of speed data as necessary for confirming an incident. To facilitate operators trying to view multiple layers of speed data, the University of Maryland is developing a tool that stacks the layers of speed data with varying levels of opacity for ease of comparison.

The Ministry of Transportation Ontario is the only agency contacted for this project that discussed averaging multiple real-time speed data sources into a single layer on the user interface. It was noted that they tried to develop an approach for averaging multiple data sources, but that challenges arose when the segments changed.

Potential Solution #3: Assign priority data sources for display on the speed layer. Some agencies have multiple data sources for speed data and have established a prioritization structure for various segments to display a single speed data layer using the highest quality data source for each segment. For example, the Missouri DOT ATMS prioritizes the display of sensor data in Kansas City and fills in third-party data for all remaining segments, whereas third-party data is shown for all segments in St. Louis. On this speed data layer, the operator does not know the data source for a given segment speed, but can have confidence in the data displayed. Texas DOT similarly assigns a priority to sensor data for each segment, but can use third-party data to fill in those segments if there are issues with a sensor.

New Jersey DOT relies on the Transcom data fusion engine for the speed data layer in the ATMS, which includes business rules for prioritization of the data source, type, and location. The confidence levels specified by each third-party data provider are incorporated into this prioritization so that the data source for a given segment can change in real-time based on specified data quality, and there are different thresholds for each data source. Understanding the strengths and weaknesses of each speed data source allows the data fusion engine to select the best qualities of each in order to derive an optimal, single speed data feed.

Potential Solution #4: Upgrade the ATMS software. Some agencies acknowledge that upgrading their ATMS software would facilitate the integration and use of third-party data. Ohio DOT believes the manner in which the third-party data is ingested into their ATMS causes additional latency. Given an overall ATMS change would be a major undertaking, upgrades could be made to modules within the existing ATMS to improve agency processes for using third-party data sources. In the past, Texas DOT staff manually selected segments for the display of travel times, but modified the software due to increased granularity of INRIX XD segments (e.g. 30 INRIX XD segments versus 6 TMC segments on a stretch of roadway). Operators now request travel times from point to point, rather than manually selecting the segments.

Outstanding Questions and Considerations that Relate to this Issue

- The availability of resources may be the predominant factor in the solution an agency chooses for this issue, as well as how frequently and for what applications TMC operators use the data.

5.6 Issue #6: Conflation

Agencies may periodically have to conflate third-party data to match the roadway segmentation in their systems. Third-party data providers periodically update map segmentation, which requires agencies to make adjustments in their systems. Historically, agencies programmed their systems to define traffic message channel segments (referred to as TMC segments) that were as much as 10 times longer than the more granular third-party datasets available today. As an example, INRIX XD segments may be about 0.5 miles in length with more granular sub-XD segmentation being roughly 250 meters.

INRIX notes that segmentation is getting better over time, but changes are still occasionally necessary to make enhancements. Third-party data providers may give a sample test feed of data to agencies in advance of the change so that they can make system adjustments before the change goes live. Agencies have found that to be helpful. While some providers make an abrupt change from one version to another version, Google Maps keeps both the old and new version active for a period of time to allow agencies more time to work out issues with the new version as they are implemented. This approach was found to be very helpful for Colorado DOT. Regardless, Virginia DOT notes that making segmentation adjustments can be a significant effort in a short timeframe.

Potential Solution #1: Incorporate software with automated processes. Some agencies have developed automated processes within their system to assist with conversions when segmentation is changed. INRIX has noted that more tools are now available to help with these efforts. Ohio DOT said they have automated some of this process in their system, although this was not as a big of a challenge when compared to the previous demands of constantly switching sensor IDs before phasing out that sensor network. Texas DOT also uses software to help redo the segmentation. However, this process is easier in some districts than others and can take several days to make adjustments. Additionally, since the use of third-party data is still growing at Texas DOT, there are a limited number of segments that need to be reconfigured.

Potential Solution #2: Require third-party data providers to supply data that conforms to agency segmentation. Instead of devoting agency resources to adjust the segmentation, an agency may require data to be provided in a specified segmentation. As an example, Ministry of Transportation Ontario worked with a university partner to develop roadway segments based on nodes and road geometries in their transportation network and require providers to recalculate their segments based on the agency-developed segments. This approach is based on a philosophy that any adjustments made by the agency would compromise the data, since only the third-party data provider understands the data well. Similarly, North Carolina DOT selected a third-party data provider because the segmentation offered adhered more closely to the standard segmentation they preferred. The University of Maryland noted that it is also possible to directly license the maps from the third-party data providers, which is more expensive up front but cheaper than making adjustments given the large network that is included in the Regional Integrated Transportation Information System (RITIS).

Potential Solution #3: Remove historic agency segmentation needs. Given the improved granularity of third-party data versus the historic traffic message channel segments, agencies may choose to remove the segmentation from their systems. This approach allows agencies to use the third-party data in operations as it is provided, making conflation unnecessary. As an example, Missouri DOT does not have links in their ATMS so the map can be populated with the new segmentation rather than having to translate it to match a defined system. Despite this solution, Missouri DOT notes that data conflation remains an issue when meshing third-party data with volume data from other sources for performance reporting.

Conversely, North Carolina DOT expressed hesitation to this approach, considering the traditional traffic message channel segmentation is more established versus a more proprietary approach. The proprietary approach could restrict their options of switching providers in the future, depending on how the agency may have adapted or developed their systems.

Potential Solution #4: Retain support consultants with expertise in third-party data. One reason some agencies may struggle with conflation is that it occurs relatively infrequently so agency staff do not have a great deal of experience with it. As such, one solution is for agencies to contract a support consultant that has more experience to handle conflation quicker and easier. Virginia DOT supports this solution, noting that support consultants can respond faster given the short notice because they are more familiar with the process.

Outstanding Questions and Considerations that Relate to this Issue

- Agencies might consider working with third-party data providers to request flexibility with the timeline between a notification and actually changing the feed. This timeline could potentially be specified in the data contract.
- What are the specific cost considerations and tradeoffs between hiring support staff, contractual requirements for segmentation, and developing automated systems to translate segmentation changes?

5.7 Issue #7: Appropriate Roles of Third-Party Data

Given the many applications and potential for third-party data, agencies may struggle to determine how to balance use alongside other speed data sources. Real-time third-party data has potential for use in many applications and agency staff may feel overwhelmed and pressured to implement so many new uses simultaneously. The abruptly expanded coverage area provided by the data may be a lot for operators to manage, let alone understanding the data and accompanying tools. As agencies transition to a Transportation Systems Management and Operations (TSMO) environment, related planning efforts might identify and define cross-cutting roles of data, including third-party data.

Real-time third-party data can be used in many applications, but agencies do not need to use third-party data for all systems if existing processes meet their needs. The following potential solutions recommend agencies take a measured approach and use third-party data for additional applications only after gaining comfort with the data and conducting an analysis to confirm the data will meet their needs in other areas.

Potential Solution #1: Use third-party data to fulfill the initial identified needs, then incrementally test and expand use to other applications over time. The initial procurement of third-party data may be for a number of reasons including filling gaps in the existing sensor network and coverage expansion to suburbs and routes beyond a metro area. The data provided may be more than is needed, i.e. statewide data, and other tools and uses for the data may be available. However, a measured approach to understanding the data before widespread use in operations is preferred by many agencies. As one example, Iowa DOT spent time focusing on historical performance measures, then figuring out the limitations of the data to better identify real-time uses. Iowa DOT has been working with university partners to incrementally develop, test, and implement various new tools using the data. Similarly, the Nebraska DOT is currently developing a framework for evaluating the third-party data accuracy and reliability in various traffic operational activities.

Potential Solution #2: Conduct analyses to identify tradeoffs of investing in traditional or other innovative systems versus expanding third-party data usage. Although third-party data has a great deal of potential, agencies may have to consider where to make investments such as expanding third-party data usage, deploying traditional sensors, or looking ahead to a connected vehicle environment. Some agencies that procure third-party data still use other systems for some applications, given convenience and low cost. For instance, Iowa DOT choses to use the free Google Maps on their traveler information website to display real-time traffic speeds because this service would incur a higher cost through their third-party data provider. As another example, TMC operators at North Carolina DOT prefer using the Bing Maps interface despite having access to the third-party data interface.

Procuring third-party data can also cause operators to appreciate the improved capabilities seen on some routes, and realize sensors are needed on other routes to achieve similar capabilities. Virginia DOT realized that the third-party data was not reliable on some routes and installed Bluetooth and Wi-Fi devices to achieve results similar to those seen on other arterials with third-party data alone. At the same time, Virginia DOT has been able to stop maintaining some point sensors in urban areas.

Finally, some agencies have taken the approach to synthesize multiple datasets as a way of increasing capabilities and confidence in data outputs. This approach allows an agency to enhance analytics by pooling data sources together, prioritize data based on the relative strengths of each on various roadway types or locations, and may help practitioners have increased comfort and gain confidence in the data and tool outputs. Ministry of Transportation Ontario and New Jersey DOT both continue to deploy sensors, while also procuring multiple sources of third-party data to improve reliability and confidence in the data. Alternatively, Colorado DOT is deploying roadside units (RSUs) to have a robust mix of traditional and new data from connected vehicles, instead of expanding usage of third-party data. Colorado DOT uses INRIX data for travel time information because a contract with Google Maps was limited to 100,000 queries per year. They plan to use sensor data, third-party data, and connected vehicle data to develop more robust capabilities.

Outstanding Questions and Considerations that Relate to this Issue

- The availability of agency resources, as well as university partners and staff expertise may be a determining factor in an agency's selected approach and the speed at which the agency is able to adapt to new systems and data sources in a rapidly evolving field.

6.0 Findings on Identified Data Shortcomings

Despite consensus that third-party data has dramatically improved over the last decade, many agree that improvements are still necessary to increase confidence before incorporating the data into additional applications and more automated processes. Specific data shortcomings that were identified will be presented in this section as follows:

- 6.1 Data Needs
- 6.2 Data Quantity
- 6.3 Data Reliability

6.1 Data Needs

Despite all the advantages of third-party speed data, it does not yet meet all agency needs. In general, Ministry of Transportation Ontario notes the need for more reliable third-party data before it can be used for more automated processes. However, sensors are still needed to support a variety of applications that use volume and occupancy data, require low latency data, or are in various rural areas or on certain arterial roadways.

No agency contacted as a part of this study uses third-party speed data in ramp metering applications. In addition, even agencies with a great deal of experience using third-party data still deploy sensors to facilitate operations of managed lanes like high-occupancy toll (HOT), high-occupancy vehicle (HOV), or reversible lanes. Virginia DOT noted that INRIX data has issues distinguishing between speeds in adjacent lanes. Virginia DOT also intentionally uses sensors for variable speed limit systems, and Ohio DOT is planning to deploy sensors for a planned variable speed limit system. Likewise, Virginia DOT uses sensors for a dynamic shoulder lane, and Pennsylvania DOT plans to do the same. While some agencies have used third-party data for a work zone queue warning system, the Iowa DOT does not feel comfortable enough with third-party data latency to deploy a system without sensors. Agencies also require volume data for before-after studies. Although third-party speed data can be used, Virginia DOT noted that given ongoing improvements to third-party data, it can be difficult to confirm whether an identified trend is the result of a roadway improvement or change in the data itself. Missouri DOT also identified the need for volume data to help assist with deploying courtesy patrol vehicles.

Additionally, feedback from users suggest that speed data is still insufficiently accurate on many non-freeway routes. Virginia DOT has noted that the data works well on high-volume arterials that do not have super-dense access points, and that arterial travel times are overly optimistic during over-saturated conditions. Third-party data is also inconsistent on low-volume rural facilities. In some cases, cellular coverage by no or only one carrier in remote areas leaves gaps in the speed data, as noted by Colorado DOT.

Overall, third-party data has been continuously improving. Ten years ago, third-party data quality was questionable even for urban interstates, but has improved to a point that it can be used with confidence on most freeways in urban areas and higher-volume freeways in rural areas, and even some arterial routes. It can be assumed that this trend toward higher quality data will continue in remote areas and on arterial roadways. The nature of third-party data is a challenge for deriving a proxy for volume and occupancy data, however given the need for these types of data, third-party data providers are working to find a solution. Specifically, the I-95 Corridor Coalition is leading a study to examine the potential for deriving volumes and turning counts from third-party data, and offering inputs to a third-party provider to collaboratively assist them in their efforts.

6.2 Data Quantity

Real-time third-party data is a much larger dataset given larger coverage areas and higher granularity than traditional sensor networks, as noted by Virginia DOT and the University of Maryland. While sensors are typically deployed at a manageable rate to allow operators to manage and use new data, the procurement of third-party data may generate a sudden and overwhelming amount of data. Texas DOT and Virginia DOT noted that the larger quantities of data can make it difficult for the computers to process. Operators will likely require some time and training to understand the best ways to use and manage the data. Iowa DOT recognizes the value in third-party data and its many potential applications for operations, and realizes the data is not being used to its full potential given the time and resources that are needed to develop new tools. North Carolina DOT similarly acknowledged that they have many ideas for using third-party data that have not been implemented yet. Virginia DOT noted that good data management and analysis tools are needed to create good usable studies and measurements from third-party data. Colorado DOT has a goal to synthesize multiple datasets to make them stronger for enhanced analytics, and have more meaningful outputs which can be leveraged for the government to help drive innovation in the market.

Understanding how to effectively leverage and use the large real-time third-party datasets in operations may help an agency prepare for the connected vehicle environment, which will be a source of even more data. The hardware and software being developed and implemented to manage third-party data and automate processes may assist agencies with the transition to a connected vehicle environment which will require major processing power to effectively leverage the large datasets generated from vehicles and infrastructure.

Many agencies and entities like the University of Maryland with RITIS and Purdue University continue to develop new tools that use the third-party data to improve traffic management and operations. Likewise, third-party data providers continue to release new tools. For instance, the INRIX Dangerous Slowdowns tool is based on a tool developed by Purdue University for the Indiana DOT that identifies slow traffic and detects the end of queues to send notifications to operators.

6.3 Data Reliability

Several agencies shared concerns about data reliability and approaches to help overcome those concerns. Both the New Jersey DOT and Ministry of Transportation Ontario use multiple third-party data sources in addition to sensors, in order to increase redundancy and increase confidence against errors or a system outage. The Ministry of Transportation Ontario would like to implement more automated functions once the third-party data is more reliable. To this end, the Virginia DOT has installed Wi-Fi and Bluetooth on some arterials for better consistency in areas where the third-party data does not perform well. For agencies that have dramatically scaled back the number of deployed sensors, Ohio DOT notes that individual segment data is no longer lost due to a sensor malfunction. However, a provider service system outage caused by a translation service error would cause the agency to lose all coverage. This is a rare occurrence, but a concern nonetheless.

Both Missouri DOT and Virginia DOT said that third-party data provides better consistency for statewide performance reporting. However, Virginia DOT pointed out that the continuous improvement of third-party data creates inconsistency in the data stream over time. Because data quality is a moving target, historic trend analysis is challenging because there is uncertainty about whether the data has changed or roadway conditions have improved.

Similarly, agencies have expressed concerns about data quality (Section 5.2), confidence scores (Section 5.3), and data latency (Section 5.4). These shortcomings are likely to be reduced as third-party data continues to improve.

7.0 References

All links to reports and websites were accessed and valid as of September 2017.

7.1 Probe Data Providers

- HERE. https://here.com.
- INRIX. http://inrix.com.
- TomTom. https://www.tomtom.com.

7.2 Probe Data Quality and Validation Studies

- I-95 Corridor Coalition Vehicle Probe Project: Data Validation. http://i95coalition.org/projects/vehicle-probe-project.
- *Assessment of Speed Information Based on Probe Vehicle Data: A Case Study in New Jersey.* Transportation Research Board Annual Meeting, 2014. https://trid.trb.org/View/1289396.
- *Quality of Private Sector Travel-Time Data on Arterials*. Virginia DOT, 2016. https://trid.trb.org/View/1395557.
- *Validation and Augmentation of INRIX Arterial Travel Time Data Using Independent Sources*. University of Maryland and Maryland State Highway Administration. 2015. http://www.roads.maryland.gov/OPR_Research/MD-15-SHA-UM-3-6_INRIX-Arterial_Report.pdf.
- *A Framework for Evaluating the Reliability of Wide Area Probe Data*. Transportation Research Board Annual Meeting, 2017. https://trid.trb.org/View/1437829.
- *Comprehensive Data Driven Evaluation of Wide Area Probe Data: Opportunities and Challenges.* Transportation Research Board Annual Meeting, 2016. https://trid.trb.org/View/1394420.

7.3 Case Studies and Resources on Tools Developed for the Use of Real-Time Third-Party Speed Data

- *Real-Time Probe Data Dashboards for Interstate Performance Monitoring During Winter Weather and Incidents*. Purdue University, 2016. http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1013&context=civeng.
- Probe Data Analytics Suite. University of Maryland Center for Advanced Transportation Technology Laboratory (UMD-CATT Lab). http://www.cattlab.umd.edu/?portfolio=vehicle-probe-project-suite.
- Iteris Performance Monitoring System (iPeMS). https://www.iteris.com/products/performance-analytics/ipems.
- *Algorithms for Identifying and Ranking Bottlenecks Using Probe Data.* University of Maryland Center for Advanced Transportation Technology Laboratory (UMD-CATT Lab). Transportation Research Board, 2016. http://docs.trb.org/prp/17-05332.pdf.
- *Improving Estimates of Real-Time Traffic Speeds During Weather Events for Winter Maintenance Performance Measurements.* Iowa State University, 2017. http://www.intrans.iastate.edu/research/documents/research-reports/winter_mtc_performance_msrmts_w_cvr.pdf.
- *Indiana Department of Transportation Use of Third-Party Data for Work Zone Management.* Federal Highway Administration, 2016. https://www.workzonesafety.org/files/documents/SWZ/IN_TA_case_study.pdf.
- *The Show-Me State Shows Missourians Reliable Transportation.* HERE, 2015. https://here.com/en/file/7811/download?token=0zIG-H0l.
- *Making the Journey Better in the State of Michigan.* HERE, 2015. https://here.com/en/file/7806/download?token=VUCYI5sv.
- *On the Move: Measuring Mobility to Ensure Effective Travel in the Sunshine State.* HERE, 2016. https://here.com/en/file/7816/download?token=ELJ2ywyK.
- *Using Real-Time Probe Vehicle Data to Manage Unplanned Detour Routes.* ITE Journal Vol. 85, Issue 12. 2015. https://trid.trb.org/View/1403278.

7.4 Ongoing Research on the Use of Real-Time Third-Party Speed Data

- Volume and Turning Movement Study. I-95 Corridor Coalition Vehicle Probe Project. http://i95coalition.org/projects/vehicle-probe-project.
- *Evaluation of Opportunities and Challenges for using INRIX Data for Real-Time Performance Monitoring and Historical Trend Assessment.* Iowa State University for Nebraska DOT. https://trid.trb.org/View/1410527.
- *Identification and Characterization of Traffic Bottlenecks.* Virginia DOT. https://trid.trb.org/View/1472473.
- *Synthesis of Information Related to Highway Practices. Topic 49-14. Best Practices on Acquiring Non-Traditional Data for Transportation Applications.* National Cooperative Highway Research Program (NCHRP) Synthesis, Transportation Research Board. https://trid.trb.org/View/1467151.

Appendix A: Question Guide for Agency Staff who Use Real-Time Third-Party Speed Data in Operations

The question guide below was developed by the project team to facilitate discussions with agencies that use real-time third-party speed data in operations. As a guide, not all questions were asked for all interviews; for instance, depending on how agencies use the data, not all questions would always apply.

1. Assuming agencies we are talking with are procuring real-time speed data from a 3rd party provider, prompt with questions to verify and gather more information:
    - What provider(s)? For how many years? Has your agency procured real-time speed data from other 3rd party vendors in the past? Are you doing anything with archived historical data besides building a performance measurement program?

2. How have you validated the quality of real-time 3rd party speed data for various settings (rural, arterials, etc.)? Have you experienced any issues or concerns with the quality or reliability of the data? How would you compare the quality and value of real-time 3rd party speed data versus sensor data? Do you use real-time data for operations [when the INRIX confidence score is less than 30]? What is the data latency, and is it different in rural versus urban areas?

3. How does your agency use the real-time 3rd party speed data?

    Based on how they answer, Prompt with questions such as:
    - Do your TMC operators rely on the real-time data to manually make operations decisions?
    - Is the 3rd party data ingested into any of your TMC systems (e.g. used for incident detection, travel time generation, DMS message recommendations, etc.)
    - Does your agency publish the 3rd party data for others to ingest? If so, do you repackage or adjust the data at all?

4. What information is generated/calculated from this data (either manually or automatically),
    - Prompt with options such as: delays, speeds, etc.?
        a) Urban areas only, or also in rural corridors?
        b) Freeways only, or also arterials?
        c) Recurring congestion only, or also non-recurring congestion? Incident detection and management? Work zone management? Road weather management?

5. Do you use the real-time 3rd party data in conjunction with sensor data or other data sources?
    - Based on replies, Prompt with questions such as:
        - Do you use the data sets separately or together?
        - How do you merge the two datasets to use them together?
        - Do you use the real-time 3rd party speed data to supplement areas without sensor coverage? If so, is there a defined approach/formula for this?

6. Have you experienced any issues or concerns with conflation of the real-time 3rd party speed data? If so, how have you addressed them?

7. Are there uses of the real-time 3rd party data that work particularly well? Not as well as expected? Do you plan to begin using the real-time 3rd party speed data for other purposes?

8. Please provide details on how operators view the real-time 3rd party speed data alongside / separate from sensor data, and what tools are being used to do so. Were there any challenges integrating the real-time 3rd party speed data with your TMC software?

9. Please provide details on any automated processes that use the real-time 3rd party speed data, e.g., ITS operations, incident detection, etc.

10. Please provide details on benefits you have seen from using the real-time 3rd party speed data.

11. Has your agency used real-time 3rd party speed data instead of deploying a temporary or permanent sensor or detector? Has your agency made any changes regarding the deployment and maintenance of permanent sensors as a result of using the real-time 3rd party speed data, or are there any expectations this may occur?

12. Has your agency used real-time 3rd party data to support traffic operations in the fringe areas of the metro area? For example, has it allowed you to delay deploying additional sensors as the volumes increase in the fringe areas?

13. How do you use volume data for operations (or in general), and are you connecting it with the real-time third-party speed data?

14. Please provide any other comments regarding related challenges or benefits with obtaining, integrating, and using the real-time 3rd party speed data.