CV Data Prioritization

Prepared by: Rick Schuman, Charles R. Lattimer, Michael L. Pack

CV Data Prioritization

[Publish Date]

The Eastern Transportation Coalition is a partnership of 19 states and the District of Columbia focused on connecting public agencies across modes of travel to increase safety and efficiency. Additional information on the Coalition, including other project reports, can be found on the Coalition's website: www.tetcoalition.org

Table of Contents

Acronyms and Definitions

Aggregator: A data aggregator ingests multiple raw data feeds (typically from different source providers), cleans and anonymizes the data, and combines them into a harmonized dataset. For example, a data aggregator may obtain raw GPS-based data from multiple data suppliers and produce average speed readings on standardized roadway segments. This aggregated dataset is then made available to agencies.

API: Application programming interface. A common type of software interface that enables the automatic retrieval of data from predefined data endpoints.

CATT: The Center for Advanced Transportation Technology at the University of Maryland, which develops and operates the Regional Integrated Transportation Information System (RITIS).

CV: Connected vehicle. For the purpose of this report, a connected vehicle is one that transmits on-board data (such as location or vehicle status) to data aggregators through third-party networks.

C-V2X: Cellular vehicle-to-infrastructure communication. Allows vehicles and roadside infrastructure to share data using a cellular connection.

DaaS: Data as a service. DaaS providers typically supply data to agencies through an API or other standard data exchange method at regular intervals on a subscription basis.

DMS: Dynamic message sign.

DSRC: Dedicated short-range communication. Included a dedicated radio spectrum in the 5.9 GHz band intended exclusively for vehicle-to-infrastructure and other public safety communications. Early connected vehicle deployments relied on agencies to install roadside DSRC units to enable communication with connected vehicle OBUs. DSRC is being supplanted by C-V2X approaches.

MRM: Mobile resource manager. A data company that provides fleet management software and uses CV data from commercial vehicles to improve logistics and vehicle maintenance for fleet owners.

OBU: On board unit. This is the in-vehicle component of a DSRC-based connected vehicle system. It transmits vehicle location and “basic safety information” that can be received by other DSRC-equipped CVs using their OBUs or by roadside units. OBUs can either be installed by the OEM or be aftermarket additions to vehicles.

OEM: Original equipment manufacturer. In the context of this report, OEM refers to both foreign and domestic automakers.

Purchasing models: Different purchasing models can be used to purchase datasets. Two common examples are project-based purchases and subscription-based purchases.

- **Project-based purchases:** Data purchased on a project basis is typically done to support the finite needs of a specific project. Some agencies can use capital funds for project-based data purchases, making them easier to fund initially. However, it can be challenging to keep the project-based datasets current and work them into ongoing performance management programs without ongoing funding for operations and maintenance. Project-based purchases are also less sustainable from an industry perspective, as data vendors incur ongoing costs from continuously ingesting and processing new data from their source suppliers, which is a necessary part of doing business.
- **Subscription-based purchases:** Data purchased on a subscription basis is typically used to support programmatic goals, such as performance management or data-driven

project prioritization. Subscription-based data has the advantage of providing a steady delivery of data sets for an agency's road network, often over a period of many years. This enables agencies to track performance trends on their road network and maintain consistent data sets over time. Data vendors frequently find subscription-based purchases more sustainable, due to the vendor's ongoing cost of acquiring and processing their source data.

RSU: Roadside unit. This is the roadside component of a DSRC-based connected vehicle system. Receives "basic safety information" from vehicles and can transmit agency-specific information to vehicles on known road hazards, travel conditions, etc.

SaaS: Software as a service. Typically, it refers to cloud-based or web-based software that is maintained and updated by the SaaS company and is generally paid for on a subscription basis.

TDM: Transportation Data Marketplace, managed by The Eastern Transportation Coalition.

TETC: The Eastern Transportation Coalition.

TMC: Two definitions of TMC appear in this report, which can be differentiated based on their context:

- **Traffic Messaging Channel**: A standardized system for defining roadway segments used in map-based travel data reporting. Data derived from CV datasets (like speeds and travel times) are often reported in terms of *TMC segments*.
- **Traffic management center**: An operations center where *TMC operators* monitor roadway conditions and dispatch first responders to the location of incidents. TMCs also serve as hubs for transportation data as operators receive data streams from the field and make detailed notes on incident response activities.

VPP: Vehicle Probe Project. A first-of-its-kind TETC initiative that empowered the broad use of CV datasets for Coalition members.

Executive Summary

Nearly 20 years ago, The Eastern Transportation Coalition (TETC) conducted a transformational experiment to procure speed and travel time data across multiple states. Through subsequent rebids of the original procurement, TETC's Traffic Data Marketplace (TDM) has expanded to include additional data sets beyond speed data. The TDM has been a clear success and is by far the country's largest single contracting mechanism for agency use of connected vehicle data-based products.

However, the world of connected vehicle data has rapidly evolved from a technology and policy perspective since the TDM began in mid-2022. With so much changing and the desire to innovate and be transformative at a large scale once again, this SCOOP project focuses on a simple question: "What more (if anything) can TETC do to increase the benefits of new and emerging connected vehicle data for its members?"

Recommendations

Based on member input, industry interviews, and historical research, this report recommends that the Coalition and its members focus activity on three areas:

1. **Stabilize the connected vehicle data supply chain** – Since the TDM started, source data suppliers have changed, and the risks they face in providing data for "non-core" uses, such as powering public sector solutions, have significantly increased. As the largest collective agency purchasers of CV-based datasets, TETC has a voice, and recent market developments suggest that the Coalition should use that voice to lessen the chances of current products being disrupted while also looking to increase future uses for CV data.
2. **Optimize TDM for the next generation of the program** – Several recommendations are provided to maintain (and increase) the TDM's unique and critical marketplace function. Steps to further increase TDM's flexibility, enhance TETC members' collective purchasing power, and expand TDM's transparency benefits are included, any and all of which would further increase TDM's value to members going forward.
3. **Explore new and emerging CV datasets** – Vehicle trajectories, driving events, and on-vehicle imagery datasets all offer the potential to dramatically increase the value of CV data to public agencies. Yet all are nascent in their productization for agency use. TETC has the opportunity to take the lead in leveraging these datasets through multi-state demonstrations and provide a pathway for the development of future products and platforms. Ultrasonic sensors from select vehicles may also offer potential for roadside infrastructure, work zone, and parking detection.

This report outlines specific activities for stabilizing the CV data supply chain and exploring new and emerging CV data sets. The current base term of the TDM contract ends in mid-2026, so TETC and the TDM Steering Committee should consider how the TDM can be evolved to accommodate recent industry changes. This report lists several recommendations on how this evolution could be accomplished. These activities and recommendations are designed so that any agency can move forward independently of others to increase the impact of CV data, as this project has sought to do.

From the list of specific activities, four stand out as important and timely to build upon this SCOOP project:

1. **Automotive industry collaboration** – Engage with the Alliance for Automotive Innovation to establish direct communication channels between agencies and OEMs, who are the generators of most connected vehicle source data. Ascertain if collaboration is possible to

establish practices and principles to balance the needs of consumers and manufacturers with those of public agencies to derive maximum benefits while minimizing commercial, legal, public relations, and political risks.

2. **Trip data truncation best practices** – Collaborate with key current and potential source data generators to develop best practices for trip/waypoint truncation to make trip-based analytics as useful as possible while balancing privacy protection. Then, feed the output of this work into the next formal TDM procurement process to ensure the most usable trip/waypoint dataset offerings.
3. **Strategic road risk assessment** – Leverage licensed data waypoints and corresponding driving events (ideally both passenger and commercial vehicles) to establish a baseline of the relative risks on the limited access network, including speeds, queues, weaving, braking, etc. This effort would require licensing data and would presumably require resources from a coalition of willing member agencies to fund and direct.
4. **Real-time value-add simulator/translator** – create a module to run "any" historically licensed waypoint, driving event, and/or imagery dataset through a simulator to generate derivative datasets and capabilities (e.g., high resolution queues, congestion scans, event logs such as the remote detection of wrong way driving events, ramp spillback, lane blockage, high speed variations, work zone activations, vision-based incident and debris detection, evacuation route selections and detouring, etc.) to populate in RITIS and/or enhance the existing datasets already in RITIS as a pre-cursor to real-time applications. This effort could start with the data licensed for the Strategic Road Risk Assessment or the real-time visualization of moving vehicles on roadways for monitoring the response to severe incidents, detours, evacuations, etc.

Agency Feedback

Sixteen TETC member agencies participated in a survey for this project. Interest in new CV data uses was consistent and high. Of the dozen use cases surveyed, every agency showed interest in at least two, and all use cases attracted interest from between seven and fourteen agencies. The barriers we identified that would slow the implementation of these new use cases were not surprising, most of which would be best addressed via collective TETC action rather than individual members. Many agencies also expressed a desire to move forward with a multi-state procurement of data and region-wide demonstration of use cases both for visibility of results and for cost reduction.

Industry Input and Research

This report addresses industry interviews and research at length, including the ever-changing list of data companies and their market motivations over the past 20 years. Since the TDM started in 2022, free market capitalism's "creative destruction" has accelerated, creating opportunities and threats for public sector data users and their use cases for that data.

When TDM started in 2022, Wejo and Otonomo offered real-time GPS data directly from millions of connected cars, and Life360 offered the same from millions of mobile phones on the move. None of these companies make this data available today.

Fortunately, other vendors have stepped in to fill the void. StreetLight Data and Compass IoT collectively generate data from 20-30 million connected cars. Arity and CMT (Cambridge Mobile Telematics), both focused on driver risk for insurance companies, generate data from a combined 50-75 million mobile phones. Furthermore, nearly all of the roughly 10 million commercial motor vehicles in the US generate data as well, through several different service providers including Geotab, Samsara, Solera/Omnitracs, and Verizon Connect. Location and movement data (called "GPS waypoint data" throughout this report) is the foundational data element that each vendor captures. Many of these vendors also offer some form of “driving event” data as well (e.g., hard braking, skidding, etc.), of differing types, amounts, and detail. External-facing cameras are starting to generate imagery from vehicles in modest but still useful amounts, presently led by Vizzion.

Market Assessment

At a minimum, all CV data described in this report will likely be made available through a tool or platform as processed data products; however, approximately half of this data is available for direct licensing by agencies. GPS data is widely available, though not necessarily for complete trips. The market for event data is still being defined, and imagery from outward-facing dashcams—particularly from commercial vehicles—is emerging and exciting.

While the amount, quality and timeliness of CV data needed to support a desired use case can vary widely, a useful rule of thumb has been proffered by Dr. Darcy Bullock of Purdue (a leader in CV data application research): “Great results and actionable insights can be achieved with 3-5% penetration rates of Connected Vehicle (CV) data. Anything more means quicker recommendations and results.”[1]

The exciting fact is that now we have multiple source data generators that meet or exceed this 3-5% threshold. However, applications to use this ever-growing data in cost-effective and sustainable ways are still maturing. TETC can play a key role going forward by creating clarity on which applications and use cases would be the most fruitful for industry to develop—and do so at a scale that demands attention from both the industry and other key stakeholders.

[1] https://www.compassiot.com.au/us/media/trajectory-industry-insights-released

Introduction

Nearly 20 years ago, the leadership of The Eastern Transportation Coalition (TETC) decided to conduct an experiment to procure speed and travel time data for the entire I-95 Corridor and adjacent arterials with the purpose of:

1. Validating the quality of the data across all geographies.
2. Standardizing best practices for data sharing between agencies while creating the best possible data use agreements.
3. Demonstrating new and transformative use-cases.
4. Reducing the cost to agencies through a multi-state procurement.
5. Encouraging innovation and the development of new products in the private sector.

Through subsequent rebids of the original procurement, TETC's Traffic Data Marketplace (TDM) has expanded to include additional data sets beyond speed data. However, the world of connected vehicle data is rapidly evolving from both a technological and policy perspective. Some data companies have disappeared. Others have been acquired or have undergone significant changes to their business models and product offerings. New companies are starting up as we complete this report. And lawsuits and impending legislation surrounding privacy concerns threaten the long-term viability of other connected vehicle offerings.

With so much changing, and the desire to innovate and be transformative at a large scale once again, this project focuses on a simple question:

"What more (if anything) can TETC do to increase the benefits of new and emerging connected vehicle data for its members?"

This project aims to maintain the Coalition's leadership in cost-effective agency use of connected vehicle datasets by further leveraging, where appropriate, the Coalition's unique combination of technical and procurement innovation with the economies of scale of multi-agency initiatives.

TETC and many of its member agencies played THE key role in accelerating the use of connected vehicle data for public sector applications through its historical Vehicle Probe Project (VPP, the predecessor of the Transportation Data Marketplace), which went live in 2008.[2] Coalition agencies continue to be national leaders in using connected vehicle-derived services for operational, emergency management, safety, performance measurement, and planning applications.

PennDOT recently completed a "Connected Vehicle Data Study" that provides timely updates on use cases and the status, interest, and feedback of private industry. A summary presentation of the study is included in this report as Appendix A. This study leverages and builds upon the PennDOT study, focusing on the "Coalition angle"—how the Coalition could add value to advance the impact of CV data beyond what a single agency could do alone.

This study focuses on third-party CV applications that leverage the cloud and utilize CV technology, either built into vehicles by automakers, installed by fleet owners to manage their vehicles through fleet management software, or cell phone-based apps. Approaches using DSRC, C-V2X, or RSUs are not within the scope of this study.

[2] https://inrix.com/blog/vehicle-probe-project-celebrates-10-years/

What is connected vehicle data? A connected vehicle (CV) is simply a vehicle that can communicate wirelessly and bidirectionally with other systems (the cloud, other vehicles, etc.). Commercial CV applications began nearly 40 years ago, when Qualcomm introduced Omnitracs in the mid-1980s, the first commercial service that allowed logistics companies to locate, monitor, and manage fleets of long-haul trucks via satellite communications networks. Consumer CV applications began in 1996, with the introduction of OnStar by General Motors. Connected vehicle data is simply vehicle-related data that is shared outside the vehicle.

What generates connected vehicle data? Sensors, computers, and communications chips are required to generate CV data. Currently, combinations of embedded equipment, aftermarket devices, and smartphones are utilized to generate CV data:

- **Embedded:** Roughly 50%[3] of active vehicles in the US contain fully embedded systems capable of generating CV data, with OnStar as the original and most widely known example. (Note that just because data is *capable* of being generated doesn't mean data is *actually* being generated.)
- **Aftermarket:** Navigation devices and usage-based insurance "tags" are examples of equipment drivers may choose to install in their vehicle that can generate CV data. These devices may have built-in communications chips or rely on "tethering" to a smartphone to export CV data from the vehicle.
- **Smartphone:** Smartphone applications (sometimes without the phone owner's knowledge) leverage on-phone capabilities to determine location, movement, and sudden changes, such as rapid deceleration, and utilize the existing communications channel to share data outside the device. When the smartphone is inside a vehicle, this is essentially CV data and is nearly ubiquitous in passenger vehicles[4].

Does CV data vary based on how it is generated? Yes. As a general rule, embedded solutions can yield more types of data and are likely to have more accuracy and higher update rates, given their access to continuous power and purpose-built sensors. Smartphones typically have fewer and less sensitive sensors, and they must manage battery and data utilization more carefully than embedded solutions. Other aftermarket solutions (such as usage-based insurance tags that are purpose-built to generate data to improve insurance market efficiencies) can typically generate high-quality CV data.

This project examines all known combinations of in-vehicle, aftermarket, and mobile device approaches.

[3] 52% is estimated by Statista in 2025, continuing to increase: https://www.statista.com/statistics/1276407/connected-car-fleet-as-a-share-of-total-car-parc-in-united-states/

[4] https://www.statista.com/topics/2711/us-smartphone-market/#topicOverview

TETC Then and Now: Leaders in the Use of Connected Vehicle Data

Simply put, The Eastern Transportation Coalition established a precedent for agencies to use connected vehicle data in mission-critical applications through its groundbreaking Vehicle Probe Project (VPP). It's impossible to exaggerate how innovative the decision to proceed with the VPP was at the time; for context, when the decision was made in 2006 to release the RFP, the iPhone did not exist, and Google Maps did not yet include real-time traffic conditions.

This section highlights the key role the Coalition has played in advancing connected vehicle data-based applications. It is important to note for this study that, while unquestionably significant and unique, the Coalition's role to date has largely been supportive and enabling, creating an environment where its member agencies can tap into connected vehicle data and implement innovative use cases.[5]

THE WALL STREET JOURNAL.

Latest World Business U.S. Politics Economy Tech Markets & Finance Opinion Arts Lifestyle Rea

I-95 Panel Plans Satellite Traffic-Report Network

By Christopher Conkey

Jan. 29, 2008 11:59 pm ET

Share Resize

In one of the biggest rollouts yet for technology designed to help motorists avoid traffic jams, the I-95 Corridor Coalition will announce plans today to disseminate real-time data on traffic flows and accidents along the East Coast using a satellite network by Inrix Inc.

The rollout this summer will cover 2,500 miles, including many major roads and a continuous stretch of heavily congested Interstate Highway 95 between New Jersey

Figure 1: The Coalition's launch of VPP was featured in a Jan. 29, 2008, article in The Wall Street Journal.

The VPP went live in 2008 with the primary goal of providing Coalition members with the ability to acquire reliable, real-time travel time and speed data for their entire roadway network without the need for sensors and other hardware. The VPP broke new ground in many ways, as described in the following sections.

Road Network Monitoring

VPP demonstrated the ability for agencies to monitor speed and congestion along an entire roadway at a fraction of the cost of traditional infrastructure-based detection. Before the deployment of VPP, transportation agencies were almost exclusively reliant on infrastructure-based solutions to monitor traffic conditions in real time. Roadside sensors are expensive to install and maintain. A significant part of this installation cost involves bringing power and communications to each sensor location. The cost of power and communications infrastructure per site increases substantially outside urban areas, which have easier access to power and fiber optic networks. These economic realities resulted in roadside speed detection hardware often being concentrated in urban areas. For the first time, VPP allowed agencies to see real-time traffic conditions in rural areas where the installation of roadside infrastructure was cost-prohibitive.

[5] Later in this study, we make a case for more active Coalition involvement in shaping the data market itself.

VPP also created consistent and seamless speed and congestion datasets across and between states, regions, and agencies. The traditional infrastructure-based approach can result in different vendors and detection technologies being deployed at the city, county, or state level. This can make data sharing difficult across jurisdictional boundaries, even along major corridors. VPP demonstrated how consistent CV data across jurisdictions can provide a single source of truth for travel times along multi-regional corridors.

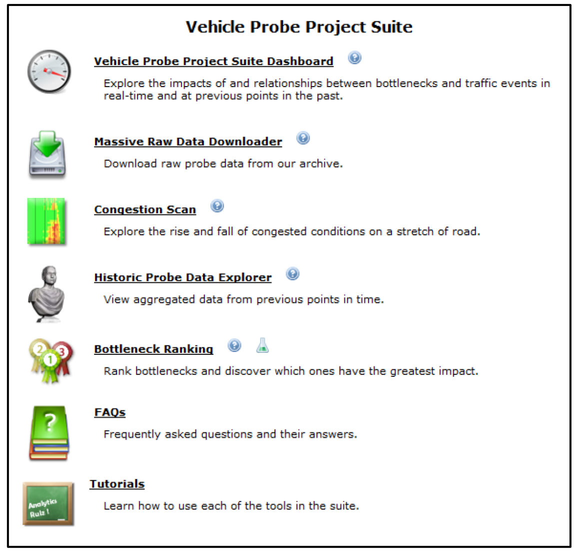


Figure 2: Example of tools available in an early version of the Vehicle Probe Project Suite.

As part of VPP, the Coalition introduced a revolutionary, multi-state traffic monitoring website for agency staff use, requiring only a browser and credentials for access. The Vehicle Probe Project Suite, developed by the University of Maryland CATT Lab, allowed agencies to easily generate performance charts and visualize congestion across state boundaries. Today, this tool has expanded nationally as the Probe Data Analytics Suite. It continues to introduce new data sets and visualizations as the industry evolves, for the benefit of Coalition partners.

Contracting and Project Management

Another major accomplishment of the Coalition was establishing comprehensive requirements for privately sourced real-time traffic data, including conducting a world-leading validation program to test and ensure data quality. It's not enough to have CV data—agencies need to know they can trust the quality of the data. The Coalition implemented a rigorous data validation program to help agencies build confidence in using CV data and understand the accuracy of data being purchased. This data validation effort involved deploying portable Bluetooth readers in the field to collect ground truth travel time data. The ground truth data was then statistically analyzed and compared with the CV probe data made available through the VPP contract. This was the first systematic data validation program to ensure CV datasets met predefined quality specifications.

The Coalition used this ground-truth data to implement "pay-for-performance" data contracts with each data provider. Payments to each vendor were tied to validated data quality and availability. This ensured that Coalition customers received the data quality and availability they were paying for, which can be challenging to achieve with new and emerging technologies.

Another issue the Coalition recognized early on was that data vendors sold their data to agencies under inconsistent usage terms and conditions. Agencies had to negotiate on their own for the best possible data use agreement. When data use agreements are overly restrictive, agencies may be unable to share data with their partners or use the dataset to its full capability. Under VPP, the Coalition established a consistent and publicly-available data use agreement that provided agencies with maximum flexibility to use, store, and reuse data for their purposes.

Traveler Information Applications

Using commercial CV data sets, states could provide traveler information on a scale that wasn't previously possible. Some of the resulting applications included the following:

- Using commercial CV data to generate travel times on dynamic message signs for an entire state network.

- Using commercial CV data to publish travel times on dynamic message signs at key roadway junctions where viable parallel routes exist.
- Publishing a statewide real-time traffic map on a state DOT's 511 travel information website and in a state DOT-branded 511 mobile app.
- Powering high-definition displays of real-time traffic at shopping mall exits.
- Publicly communicating holiday historic travel trends from previous years to aid travelers in travel planning to avoid congestion.
- Providing queue warning alerts on dynamic message signs statewide, even in rural areas with no camera coverage for verification.

Analytics Applications

Another tremendous foresight from the Coalition was its funding of platforms to store, retrieve, and analyze CV probe data. After it is stored, real-time data quickly becomes historical data. When managed and visualized properly, data initially intended for real-time operations can later be used for planning purposes. This "buy once, use many times" approach significantly increases the value of data while reducing costs, as it reduces the need for special-case data sets that would otherwise require separate purchases. Some of the significant accomplishments of the VPP Suite (currently PDA Suite) include the following:

- Creating online performance monitoring and analysis tools, enabling large-scale analysis with only a browser and credentials required for access.
- Creating region-wide real-time and historic monitoring of signalized arterial corridors to detect and address abnormal delay patterns.
- Generating project assessment summaries showing the impact of transportation improvements by their effects on traffic flow using private data (before/after studies).

Over nearly two decades since initiating the VPP, TETC has continued to advance both the state of the practice and the state of the art in connected vehicle data applications. It is still the only multi-agency effort to leverage CV data outside the federal government.

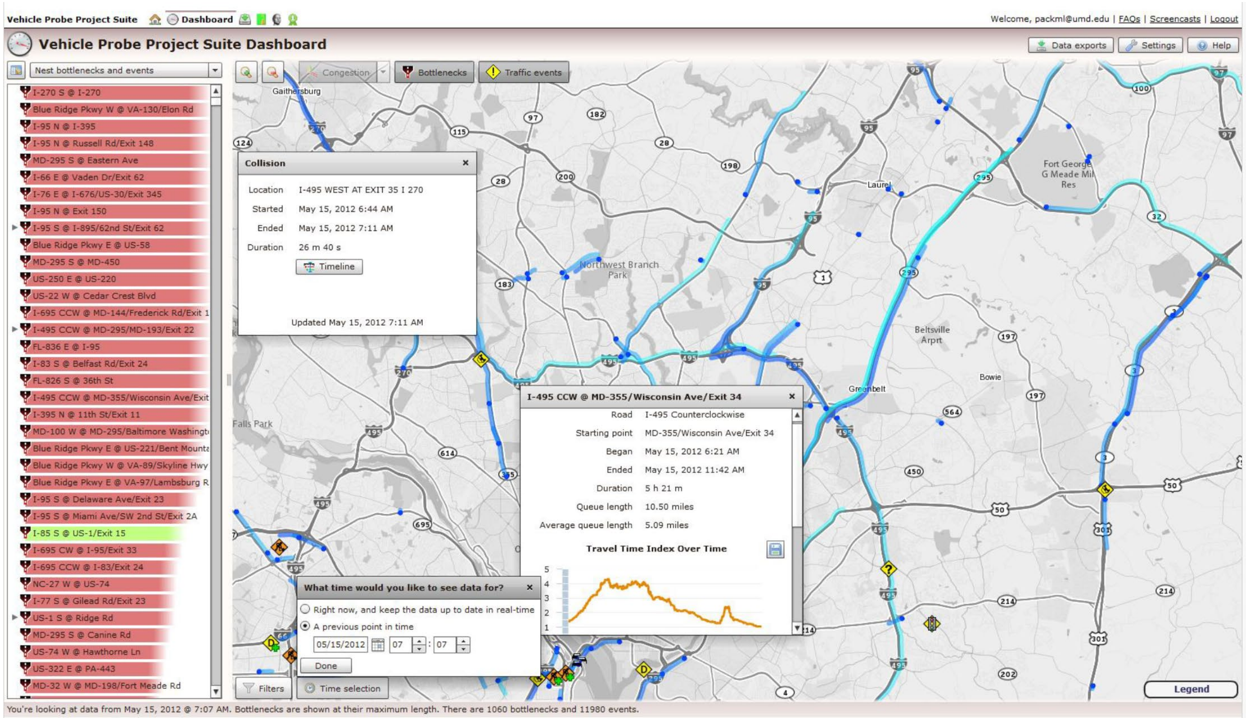


Figure 3: Example of an early version of the Vehicle Probe Project Suite Dashboard.

The Coalition's current effort, the Transportation Data Marketplace (TDM), offers members the opportunity to select from a host of prequalified vendors to provide data in six different categories, including Travel Time and Speed, Origin-Destination, Freight, Waypoint, Volume, and Conflation. And the Coalition continues to lead in application development via UMD CATT Lab's RITIS tools,[6] as well as efforts to validate origin-destination and volume datasets.[7]

[6] https://ritis.org/handbook.pdf

[7] Latest validation studies for origin-destination datasets is here and for volume datasets is here

Connected Vehicle Data Market

Public agencies are only one of many users of connected vehicle data, and in most cases, they are not the primary drivers of commercial product decisions. To position the Coalition and its members to benefit from CV data over the long term, it is essential to understand the broader market and its technical and commercial ecosystems.

A Brief History of the CV Data Market

Since the Coalition became involved with CV data as part of the Vehicle Probe Project, the CV data market has progressed through four distinct waves and continues to evolve. Understanding how the market evolved in the past can help us identify challenges and opportunities in the years ahead.

Wave 1: Genesis (2008 – 2013)

The CV data market began by leveraging tools built for in-vehicle navigation and routing. CV data aggregators primarily sourced data from mobile resource management (MRM) providers, such as truck logistics companies. In this wave, CV data was exclusively GPS-based data for deriving speed and congestion on roadways. We began to see agencies using CV data to replace speed detectors. This was also when the public started to see live traffic on a map from companies like INRIX, Google Maps, Apple Maps, and Waze—while these capabilities are now commonplace, it can be hard to remember that they were novel not long ago. Data from smartphones also slowly began to enter the market.

Major public procurements, tools, and involvement: TETC VPP, Probe Data Analytics Suite

Wave 2: Expansion (2014 – 2018)

In the second wave, the primary driver of the market was from mobile navigation and in-vehicle navigation sources. Data sets were expanded to include trips and origin-destination data. We saw tremendous growth in data sourced from smartphones, including from MapQuest and later Life360. With the widespread adoption of CV data for speed and congestion analysis, the Federal Highway Administration moved forward with including congestion reporting and target setting as part of MAP-21 implementation. We also began to see data vendors experimenting with estimating volume data and early traffic signal performance measures applications.

Major public procurements, tools, and involvement: TETC VPP2, National Performance Management Research Data Set (NPMRDS)

Wave 3: Exuberance (2019 – 2023)

By 2019, the primary market drivers had shifted to finding new ways to monetize CV data from anyone willing to sell. Companies like Wejo and Otonomo dominated the market, providing CV data aggregators with new capabilities. We also began to see the introduction of safety event data into the market, such as data on hard braking or cell phone handling. Signal analytics applications went live, and further steps were taken to bring estimated volume products to market. During this wave, we also saw the first significant public blowback when it was revealed that Life360 was selling location data to data aggregators. This prompted Life360 to change how it shared data with third parties, which reduced some of this data’s usefulness.

Major public procurements, tools, and involvement: TETC TDM, AASHTO Performance Measures Technical Service Program

Wave 4: Regrouping (2023 – present)

In the middle of 2023, the CV data market experienced significant disruptions with the bankruptcy and subsequent sale of both Wejo and Otonomo. With these large and stabilizing data sets off the market, it also became clear that some location-based services (LBS) data sources had significant quality issues (either due to a lack of care or, in some cases, outright fraud). As part of the current wave, we see some of the underlying data sources used by Wejo and Otonomo returning to the market, but with additional restrictions placed upon this data in the name of privacy. Data aggregators are working diligently to diversify their list of data sources, thereby reducing the risk of relying too heavily on a single data source provider.

Major public procurements, tools, and involvement: Upcoming TETC TDM2

The following tables show the evolution of source data providers, market drivers, and derivative data sets.

2008: VPP Start

- **Fleet Data**
  - Omnitracs

2014: VPP 2 Start

- **Fleet Data**
  - Omnitracs
  - Fleetmatics
  - Verizon
- **Mobile Phones**
  - MapQuest
  - Apple (TomTom only)*
  - Google (not shared)
- **Connected Vehicles**
  - Starting but limited

2022: TDM Start

- **Fleet Data**
  - Geotab (not shared)
  - Verizon (Fleetmatics)
  - Solera (Omnitracs)
- **Mobile Phones**
  - Life360
  - Apple (TomTom only)*
  - Google (not shared)
- **Connected Vehicles**
  - Wejo
  - Otonomo

2025: Now

- **Fleet Data**
  - Geotab (not shared)
  - Samsara
  - Verizon (Fleetmatics)
  - Solera (Omnitracs)
- **Mobile Phones**
  - Arity
  - CMT (not shared)
  - Apple (TomTom only)*
  - Google (not shared)
- **Connected Vehicles**
  - StreetLight Data
  - Compass IoT

* - Assumed: Neither Apple nor TomTom has publicly confirmed

Figure 4: Key GPS source data creators over time.

2008: VPP Start

- **Consumers**
  - In-vehicle navigation and routing
- **Fleets**
  - Asset management
  - Route optimization
- **Public Sector**
  - Freeway speed detector alternative

2014: VPP 2 Start

- **Consumers**
  - Mobile navigation
  - CarPlay/Android Auto
- **Fleets**
  - In-vehicle navigation and routing
  - Electronic logs
- **Public Sector**
  - Network wide monitoring
  - Network wide performance assessment
  - Travel times on DMS

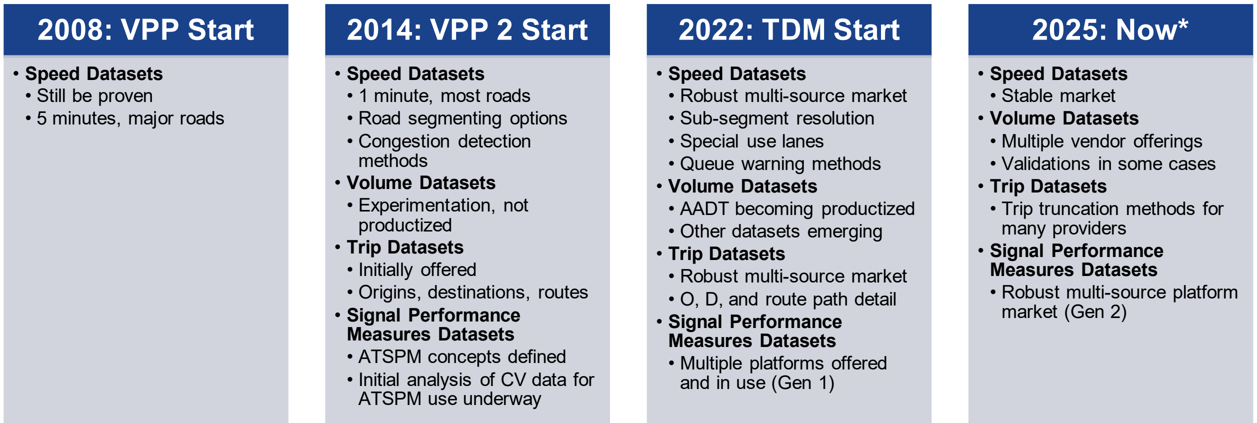
2022: TDM Start

- **Consumers**
  - Monetizing CV data
- **Fleets**
  - Performance optimization
  - Cameras/liability protection
- **Public Sector**
  - Federal reporting requirements
  - Network wide travel demand analysis
    - O-D + routes
    - Volume weighting
  - Signal performance

2025: Now

- **Consumers**
  - AV/EV dev support
- **Fleets**
  - Return on investment (ROI)
  - Driver training/retention
- **Public Sector**
  - Safety
  - Critical event assistance
  - Reduce effort to gain benefits from CV data

Figure 5: CV derivative dataset market drivers over time



* - Based on adequate GPS waypoint data continuing to be available

Figure 6: CV derivative dataset evolution

CV Data Taxonomy

"CV data" is an umbrella term applied to a wide array of data types and product offerings. To help the Coalition better assess the market, we created the following CV data market taxonomy.

Primary Types of CV Data

There are three types of CV data currently being generated in quantity:

- **Telemetry data:** Measurements from in-vehicle sensors that continuously sample a metric of interest, such as location, heading, speed, gyroscope, ambient temperature, fuel use rate, etc. Telemetry datasets can grow large and be costly to share outside vehicles. GPS waypoint data and O-D data are simple applications of telemetry data.
- **Event data:** Discrete events detected within the vehicle indicating "out of the ordinary" situations, such as airbag deployment, anti-lock braking activation, rapid acceleration or deceleration, and trip metadata (e.g., vehicle type, engine type, number of occupants). Event data is more economical to share outside of vehicles, but is also less consistent and "combinable" between possible data providers, due to differences between vehicle manufacturers in the type, quantity, and format of event data collected.
- **Imagery data:** Images or video from cameras on vehicles, coming from inward or outward-looking dashcams or cameras embedded for level 1 advanced driver assistance system (ADAS1) applications. The data size is large, depending on the resolution and image frequency, with video sizes getting very large very fast.

It is important to note that most data generated within vehicles is not shared outside the vehicle, principally for privacy and cost reasons (communications, data storage, etc.). What is possible and what is realistic often diverge in the CV data market.

CV Data and Public Sector Solutions

CV data alone is necessary but insufficient to create usable solutions for agencies. As a result, the market has evolved to generate data, make it usable, and further facilitate its use by agencies that lack the time and/or expertise to manipulate data in its original form. At a high level, there are three levels of CV data-based solutions, which can also be mixed and matched:

- **Source datasets:** Original data collected from vehicles, often called "raw data." It can be geospatially and or temporally filtered. It is the most basic CV data collected, providing the greatest flexibility for analysis and interpretation. However, it requires data aggregators or agencies to perform the complex steps of processing the raw data and turning it into useful information. GPS waypoints are the most obvious example of source data.
- **Derivative datasets:** Any dataset that is derived from the source data. The speed data that most member agencies have utilized for many years is an example of a derivative dataset, where a vendor interprets GPS breadcrumbs to derive useful speed data mapped to the road network. Hard braking is also a derivative dataset type, as are most event data datasets (source data interpretation can occur within the vehicle, outside the vehicle, or in some cases, both). Derivative datasets are often more cost-effective for specific use cases, but they are also, by default, "black box" in some way, requiring transparency from a vendor and/or third-party validation to establish confidence in dataset quality and suitability. Derivative datasets still require steps by agencies to derive value, but typically, less effort and cost are needed compared to raw source data.
- **Tools and platforms:** A combination of source or derived dataset(s) and software that enables one or more use cases with minimal to no additional efforts by agencies. UMD CATT Lab's Probe Data Analytics Suite, INRIX's Signal Analytics, and many safety and trip analytics solutions are examples of datasets and software combined and acquired as a package. Enables agencies to derive value from CV data with minimal additional steps, with the tradeoff typically being less transparency regarding the underlying CV data used with the tool or platform. (Some platforms, like the Probe Data Analytics Suite, facilitate transparent data use by permitting users to download detailed source data for offline analysis; others may severely limit users' ability to use data outside the vendor's platform.)

Figure 6 illustrates a graphical example of how derivative data providers serve as data aggregators—collecting data from multiple source data providers, transforming source data into derivative datasets, and then making those datasets available either through their own platforms, via standalone datasets, or through third-party platforms, often via an API.

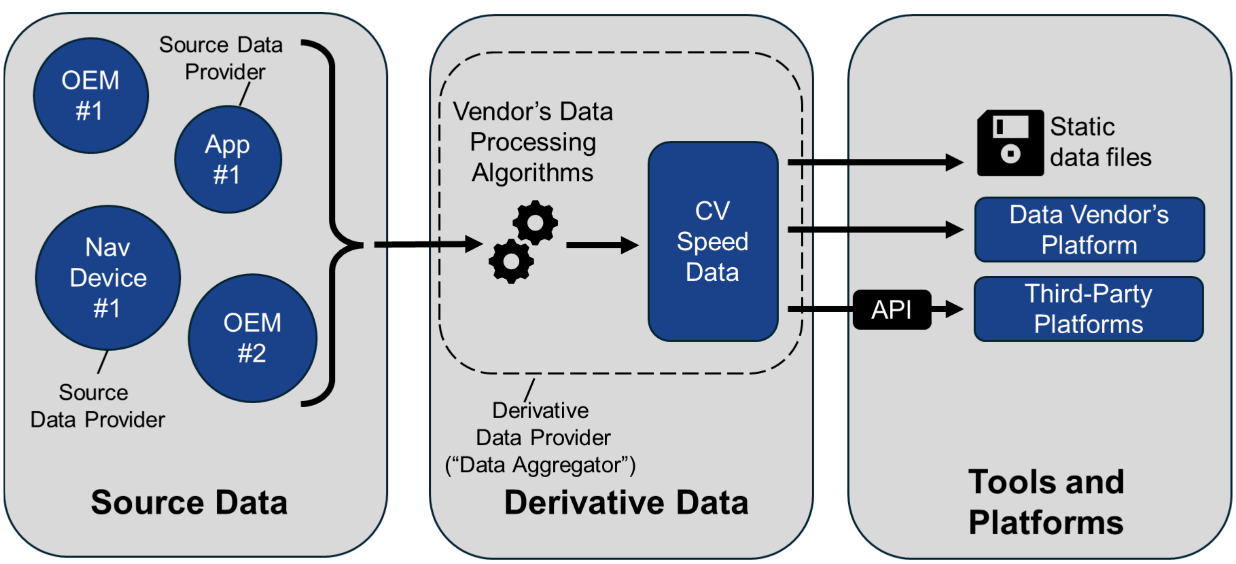


Figure 7: Illustration of how source data becomes derivative data, which is made available through files and platforms

This study encompasses three types of CV data and three levels of CV data-based solutions within the SCOOP study scope.

Products and Use Cases for CV Data

The CV data market can be summarized using the CV Data Marketplace taxonomy described in the section above, with a matrix of data and solution types. Example products and use cases, categorized by data type and solution sophistication, are presented in the table below. The following table provides examples of companies currently involved (or who could potentially be involved) in each data area. Appendix A also contains a presentation summarizing sources and potential use cases for CV data.

Table 1: Examples of how source data becomes derivative data, which is then used in various tools and platforms

	Solution type by class of data		
	Source Data	**Derivative Data**	**Tools and Platforms**
Telemetry	• GPS waypoints • Accelerometer/ gyroscope • Precipitation/ temperature • Engine data (oil, fuel, transmission, electrical, emissions, etc.)	• Speeds and travel times (real-time, archived, average, free flow, min, max, etc.) • Bottlenecks, slowdowns, queues • Trips: origin, destination, route • Traffic counts/ volumes (AADT, dayparts, etc.) • Vehicle-intersection interaction logging • Wrong-way driving logging • Datasets filtered for specific vehicle types (e.g., EVs or CMVs)	• Real-time traffic monitoring, maps, alerts • Historical traffic - speeds, bottlenecks, volumes • Traffic signal analytics • Trip/OD analytics • Safety analytics • Work zone performance analytics
Events	• Diagnostic trouble codes • Warning light activation • Trip metadata (start/end, vehicle info, number of passengers, etc.) • Park (e.g., truck parking) • User-generated incident reports (e.g., Waze)	• Aggregated/filtered datasets: ○ Crash and safety ○ Weather ○ Asset management ○ Truck parking event logs ○ Incident logs	• Safety analytics • Truck parking analytics
Imagery	• Vehicle/device-generated picture/video	• Rough pavement locations • Bad striping locations • Street furniture anomalies • Work zone characterization (e.g., barrels, cones, signs)	• Asset management • Work zone MOT analytics

Table 2: Examples of how source data from data originators becomes derivative data and is then made available through tools and platforms

	Data vendors by class of data		
	Source Data Providers (source data)	**Data Aggregators (derivative data)**	**Tools and Platforms**
Telemetry	• StreetLight/Jacobs • Compass IoT/ Stellantis • CMT • Arity	• INRIX • HERE • TomTom/ Timmons	• INRIX • Iteris • StreetLight • Replica • Geotab Altitude • Cambridge Systematics LOCUS • Flow Labs • Miovision Traffop • Google RM Insights • Others...
Events	• Compass IoT/Stellantis♦ • CMT♦	• Arity (not available, only through Michelin) • CMT♦ • Geotab♦ (fleet only) • HERE♦ • Others♦	• Michelin Mobility/Arity • INRIX/GM • Nira Dynamics/Audi • i-Probe/Honda • Geotab (fleet only)
Imagery		• Vizzion (fleets) • Netradyne (fleets)♦ • Nexar♦	• Michelin Road Condition Analytics • Blyncsy Road Conditions

♦ - It's assumed that these companies have the capabilities to produce these types of data, but it's uncertain whether they will sell this data to others or make it available for public sector use.

Summary of Industry Interviews

Over a dozen companies have been interviewed in support of this project, ranging from long-term players supporting public agency use of CV data to newer players in the market. Appendix B includes key takeaways from the collective discussions.

Overall impressions from the interviews are:

- The number of CV data-capable vehicles continues to grow, with tens of millions of passenger vehicles and many millions of commercial motor vehicles connected-capable; most new vehicles enter the nation's fleet connected-capable.
- Most potential CV data never leaves the vehicle, due to cost. Even for the data that does leave the vehicle, automakers still only release a small fraction of it (not enough revenue to offset public relations (PR) concerns over user privacy in many cases).
- We are only scratching the surface of the true benefits of CV data to support public agency services and objectives. Many, many opportunities exist to enhance the utilization of CV data.

- The costs to convert source data into usable derivative datasets or use them in tools and applications have likely stopped decreasing and are bottoming out. Fees could potentially rise in the coming years.
- While the number of vehicles capable of providing CV data continues to grow, the ability for agencies to access unfiltered source telemetry and event data may have peaked due to privacy and PR concerns.
- Data aggregators will continue to be the most likely provider of derivative datasets, and potentially even source datasets, for several reasons:
    - Often require multiple sources to achieve a critical mass of source data.
    - Automotive OEMs have not historically contracted directly with public agencies.
    - A significant amount of work is required to make source data useful. Most OEMs don't view the creation of derivative data sets as a worthwhile investment to do themselves (they outsource this to others).
- Data aggregators historically desire to offer sophisticated tools and platforms, potentially diminishing their interest in providing source or derivative datasets over time.
    - Subscriptions (as opposed to one-time project-based data purchases) have the most value to companies. Software can have higher profit margins, too.
    - Data can be less transparent in tools or platforms, which requires more validation efforts.
    - It can be more challenging to cross-pollinate data siloed within a platform.

It is important to remember that public agencies are secondary stakeholders in the CV data marketplace. CV data exists principally to address the needs of vehicle manufacturers and dealers, insurance companies, fleet operators, and the drivers and users of individual and fleet vehicles. There are benefits to being a secondary stakeholder in the CV data ecosystem, most notably because primary stakeholders cover the vast majority of the costs associated with generating the CV datasets. There are also drawbacks, including that secondary stakeholders are not the primary concern of those designing connected vehicle services. Changes in broader market forces and conditions can dramatically impact the CV data market for public agencies.

CV Data Evaluation Factors

There is no standard product definition for CV data. This market has evolved to offer “what is possible” based on the density of connected vehicles, the types of data that can be generated from those vehicles, the willingness to share, the cost of providing, and other factors. As such, evaluating CV datasets in terms of meeting agency needs and overall cost-effectiveness has become a crucial step for agencies considering the acquisition of CV datasets, and an area in which the Coalition has provided unique leadership for its members via the TDM. As documented below, evaluation has both technical and commercial factors.

Technical Evaluation Factors

Source or derivative datasets can vary in quality in many ways. This section highlights several important technical categories of data quality to consider for any CV dataset.

- **Density** – The percent of overall vehicles (or similar metrics like trips or vehicle miles traveled [VMT]) included in the dataset.
- **Frequency** – How often new data is created in the dataset (e.g., every second or every 60 seconds). The update frequency is meaningful for telemetry and imagery data, but probably less so for event data, assuming all events are captured.
- **Latency** – How long from source data generation to availability for agency use. Particularly important in real-time operations use cases.
- **Completeness** – How much data is gathered and shared for a given vehicle or device? For example, is data available for the full trip or just a truncated portion of it? Is data available on all roads, or are some roads omitted?
- **Consistency** – Given that datasets of any meaningful scale will be generated by different vehicle brands, types, model years, etc., how consistent is the combined data (e.g., are all harsh braking events triggered by the same deceleration rate)?
- **Accessibility** – How simple is the dataset to retrieve and interpret? Are application programming interfaces (APIs) available and documented? Is the data in the dataset understandable and well-documented by a data dictionary?
- **Filtering** – How efficiently does the dataset size match the desired region(s), time period(s), etc.? Data structured to be efficiently filtered can minimize agency efforts and costs as they use it in their own applications.
- **Reliability** – How consistent is the dataset’s presentation to the agency(ies) across the factors above, plus the overall availability of the dataset (“the dataset is what was agreed upon licensing, and it is reliably being provided when, where, and how it is expected to be provided”).
- **Representation** – Is the data a representative sample of vehicles and devices, providing data representing the full range of vehicles of interest for the given use case? Can results be extrapolated or scaled as necessary?

The data quality required to achieve a use case can vary significantly depending on the use and even within the same general use case.

Example: Wrong Way Driving

Identifying a wrong-way driving (WWD) incursion on grade-separated roads is possible by analyzing GPS waypoint data to determine if, where, and when a vehicle is traveling in the wrong direction on a highway or interchange ramp. Systematic real-time detection of wrong-way driving from via connected vehicles is unlikely to be achievable for many years (if ever) for reasons noted below. However, identifying WWD hotspots and tracking their trends over time may be possible using CV data, as the requirements are more achievable. In each case, high-frequency data is required. If agencies have a limited budget for storing and analyzing data, this high-frequency data can be filtered to provide information only on or near grade-separated roads and their interchanges. Some other key factors to evaluate data suitability vary widely between real-time detection, hotspot identification, and trending:

- **WWD real-time detection**: Requires
  - Very high data density, ideally 100% connected vehicle penetration.
  - High-frequency updates to allow for quick and reliable detection and confirmation of wrong-way travel.
  - Zero latency to enable immediate detection. At most, a single vendor can provide 5%–15% of vehicles that meet the frequency and latency requirements, which is nowhere near the required data penetration rate.
- **WWD analysis and trending**: Hotspot identification and trending may be addressed by non-real-time dataset(s) that combine a sufficient combination of data density, adequate temporal length of the dataset (e.g., a year) to generate enough wrong way incursions or near incursions (where a vehicle turns onto a ramp the wrong way but realizes the mistake and turns around before merging into traffic) to establish statistical confidence in the dataset. Filtering to specific roads would be important (e.g., limited access road network and associated interchanges).

The technical viability for CV datasets is a moving target. Experience has shown that discrepancies can occur between marketing claims and the actual composition and quality of datasets that are delivered. Samples, pilots, and validations (before purchase, part of the early stages of the contract, and/or ongoing throughout the contract, depending upon need and budget) are highly recommended whenever possible.

Commercial Evaluation Factors

Datasets that pass technical requirements for specific use case(s) must also be commercially viable for purchase. Even if the data meets technical needs, successful use of CV data is often determined by commercial factors such as:

- **Data use and licensing terms** – In nearly all cases, vendors will not transfer ownership of datasets to agencies; instead, they license the right to use a given dataset under documented terms for the contracted period. Terms typically include stated limitations on approved uses and users of the dataset, restrictions on sharing datasets, and in some cases derivative datasets created using the licensed dataset, outside of authorized users, copying or downloading all or parts of the dataset to other systems, and disposition of the dataset upon contract completion (e.g., must it be deleted immediately, after a specific time period, or is usage perpetual after the end of the contract and if so under what terms). **The Coalition has created the most robust data use and licensing model in the VPP and TDM programs – this model should be considered one of the crowning achievements of the VPP/TDM program and serve as the baseline for any and all future licensing under the Coalition banner.**
- **Payment terms** – The manner and timing of vendor payment for a given dataset have important implications for the vendor, the Coalition, and the funding member agency. Prepayments aid vendor cash flow and simplify contract management for all parties; however, they may also reduce vendor incentives to deliver on their promises. Ideally, payment terms balance the following factors:
  - Minimizing complexity for all parties.
  - Ensuring the given dataset meets agreed performance requirements.
  - Providing clear guidance on what happens if or when performance requirements are not met.

    - Supporting vendor financial stability to continue dataset delivery.
    - Conforming to agency procurement policies and procedures.
- **Transparency of sources and algorithms** – When agencies acquire CV data, they are basically outsourcing data creation. Thus, it is important to understand what data is provided, how it is created, and from what sources, to the maximum extent possible. With derivative datasets, understanding the algorithms used to process source data is necessary to establish high confidence. In terms of tools and platforms, details about the underlying datasets within are required to gain trust in published outputs. In reality, there is a push-pull between the user's desire for complete transparency of sources and methods and the vendor's desire to maintain the confidentiality of intellectual property, as well as, in some cases, upstream confidentiality agreements that the vendor must execute in order to obtain the source data in the first place. For this reason, the next topic—dataset validation—has become an important part of CV dataset acquisition. Looking forward, given both the history of some source data providers (more details below) and the risks around privacy perceptions, it is becoming increasingly important for acquirers of CV datasets to seek *provable permissions* from source data generators to ensure that datasets purchased are indeed legally legitimate and at low risk of being turned off during the contract term.
- **Dataset validation** – While the act of validating datasets is technical, vendor willingness to participate in fair and open validation processes is a commercial issue. Participation is even more important if transparency over sources and methods is lacking. Sometimes validation is the only way to test if the black box process of generating the data meets expectations. Vendors with an unwillingness or tepidness to participate in validation and a lack of transparency regarding data sources and methods have historically been red flags.
- **Pricing simplicity and economies of scale** – Ideally, pricing models for datasets are straightforward, with acquirers benefiting from the economies of scale associated with larger acquisitions (a more extensive road network, a multi-year commitment, etc.). Acquirers should seek simple pricing models from vendors and exercise caution when considering opaque pricing models.
- **Documentation, support, and training** – Like every service contract, data delivery is only a part of the overall effort. The ability for users to understand the data requires suitable documentation, training materials, and initial and ongoing support. Also, in the rare occasion that dataset delivery is interrupted, demonstrated vendor commitment to address and rectify issues cannot be understated, especially when datasets support mission-critical applications. Vendor willingness and a history of supporting public sector customers are key elements in evaluating the likelihood of success for a given contract.
- **Business stability** – The CV data market is a technology business. As such, vendors offering datasets and services can evolve quickly. Start-ups arise with new datasets, and ongoing businesses sell, merge, or sometimes go out of business. Vendors can change their strategy without warning, such as transitioning to only providing datasets within tools and platforms, rather than licensing the underlying datasets. Considering the health of a vendor has always been important, but the market dynamics of the past few years have emphasized this vividly. If this project were conducted in 2021, the principal vendors discussed would have surely been Wejo and Otonomo, emerging companies that recently went public and raised hundreds of millions of dollars each, specifically to provide CV data and services. Neither provides CV data today. Wejo went bankrupt in May 2023, abruptly stopping data delivery to all customers without warning, leaving many customers and data aggregators in the lurch and with unfulfilled contracts. Also in 2023, Otonomo merged with another company and transitioned out of the CV data business, publicly stating that it was highly unprofitable as a CV data company and was redeploying its remaining cash in

another business area. Companies like INRIX, TomTom, HERE, and Iteris have all been in existence for at least two decades, demonstrating that business stability in this industry is indeed possible. Wejo and Otonomo demonstrate that stability is not a given, even with access to data and resources. Each vendor should be examined on its own merits. The more mission-critical the use case is, the more critical vendor stability becomes for the contract term and beyond.

CV Data Perspectives and Considerations

Connected Vehicle Data – a Brewing PR Problem?

The geolocation data industry lives in a delicate balance between exciting new applications and emerging privacy concerns. CV data is a subset of the broader geolocation data universe (which includes advertising and retail sales tracking) and thus shares this same nexus and tradeoff of privacy, permissions, and benefits.

Until recently, in the United States, concerns about privacy-protecting regulations or consumer sentiment that forces CV data providers to pull back from supplying location data have not materially impacted the growth of public sector uses for CV data. It is likely this could change and should be carefully monitored. As discussed in the Priority Actions section, proactive steps could be taken to reduce the possibility of CV data availability being further limited.

While it is unclear how legal and regulatory processes will proceed in the coming months and years, a more likely outcome is that some, many, or even all possible CV source data providers may discontinue offering data to the public sector due to the risks of negative press and subsequent potential lawsuits.

A single reporter, Kashmir Hill, with the *New York Times,* has written several stories in the past year that raise awareness and concerns regarding the use of connected vehicle data (subscription required to read articles):

- March 11, 2024: “Automakers Are Sharing Consumer Driving Behavior With Insurance Companies,” https://www.nytimes.com/2024/03/11/technology/carmakers-driver-tracking-insurance.html
- March 14, 2024: “Florida Man Sues G.M. and LexisNexis Over Sale of His Cadillac Data,” https://www.nytimes.com/2024/03/14/technology/gm-lexis-nexis-driving-data.html
- March 18, 2024: “Your Car May Be Spying on You,” https://www.nytimes.com/2024/03/18/podcasts/the-daily/car-gm-insurance-spying.html
- March 22, 2024: “General Motors Quits Sharing Driving Behavior With Data Brokers,” https://www.nytimes.com/2024/03/22/technology/gm-onstar-driver-data.html
- April 23, 2024: “How G.M. Tricked Millions of Drivers Into Being Spied On (Including Me),” https://www.nytimes.com/2024/04/23/technology/general-motors-spying-driver-data-consent.html
- April 30, 2024: “‘Smartphones on Wheels’ Draw Attention From Regulators,” https://www.nytimes.com/2024/04/30/technology/regulators-investigate-carmakers-driver-tracking.html
- June 9, 2024: “Is Your Driving Being Secretly Scored?” https://www.nytimes.com/2024/06/09/technology/driver-scores-insurance-data-apps.html
- June 26, 2024: “Automakers Sold Driver Data for Pennies, Senators Say,” https://www.nytimes.com/2024/07/26/technology/driver-data-sold-for-pennies.html
- December 20, 2024: “How Your Car Might Be Making Roads Safer,” https://www.nytimes.com/2024/12/20/technology/connected-cars-roads-data.html
- January 13, 2025: “Texas Sues Allstate Over Its Collection of Driver Data,” https://www.nytimes.com/2025/01/13/technology/texas-allstate-driver-data-lawsuit.html
- January 16, 2025: “General Motors is Banned From Selling Driving Behavior Data for 5 Years,” https://www.nytimes.com/2025/01/16/technology/general-motors-driving-data-settlement.html

While this reporter has been the most prolific, others also publish similar stories. Reporting such as this has led to multiple lawsuits filed against companies by the State of Texas, including against

General Motors and, as noted in an article above, Allstate/Arity (even as Texas DOT heavily leverages CV data). Similar lawsuits have also been filed in Indiana and Arkansas.

The Coalition should continue to monitor the industry's response to negative publicity regarding CV data. The industry's response to privacy filtering or data availability could materially impact the usefulness of certain CV data types in the future.

GM’s History in Connected Vehicle Data – An Illustrative History

General Motors’ history with connected vehicle data is instructive when trying to understand how source data creators view the overall market, and where they may or may not go next. Below is a narrative timeline of notable milestones over the past 20 years for GM:

1. **1996:** General Motors becomes the first automaker to introduce connected car features to the market with OnStar.
2. **2011:** OnStar revised its terms and conditions to permit the sale of vehicle location and speed data to third parties.
3. **2019:** Invests $25M for 35% of Wejo, while agreeing to a long-term source data supply agreement.
4. **2021 (June):** The OnStar Guardian App is launched, expanding OnStar's capabilities beyond GM vehicles.
5. **2021 (October):** GM Future Roads business unit established, leveraging GM data to improve the safety and efficiency of road networks.
6. **2021 (November):** Wejo goes public, raising over $225M in cash with a total market capitalization exceeding $1B.
7. **2022 (May):** GM Future Roads partners with INRIX, launches Safety View.
8. **2023 (Summer):** Wejo declares bankruptcy, shuts down operations, and its assets are sold to Streetlight/Jacobs.
9. **2024 (March):** The *New York Times* reports GM sells data for insurance industry use without customer knowledge (additional related articles throughout 2024).
10. **2025 (January):** GM settles with the Federal Trade Commission, banned from selling driver behavior data for 5 years.

The arc of GM and Connected Vehicle data can be arrayed to fit the famous Gartner Hype Cycle curve, with inflated expectations giving way to reality, which leads to disillusionment (see Figure 8 below, with numbers on the chart corresponding to the list above). A concern is that over time, GM will decide that sharing CV data with third parties is not worth the consequences. They (and others like them) may need assistance from organizations like the Coalition to not give up on CV data and reach the plateau of productivity.

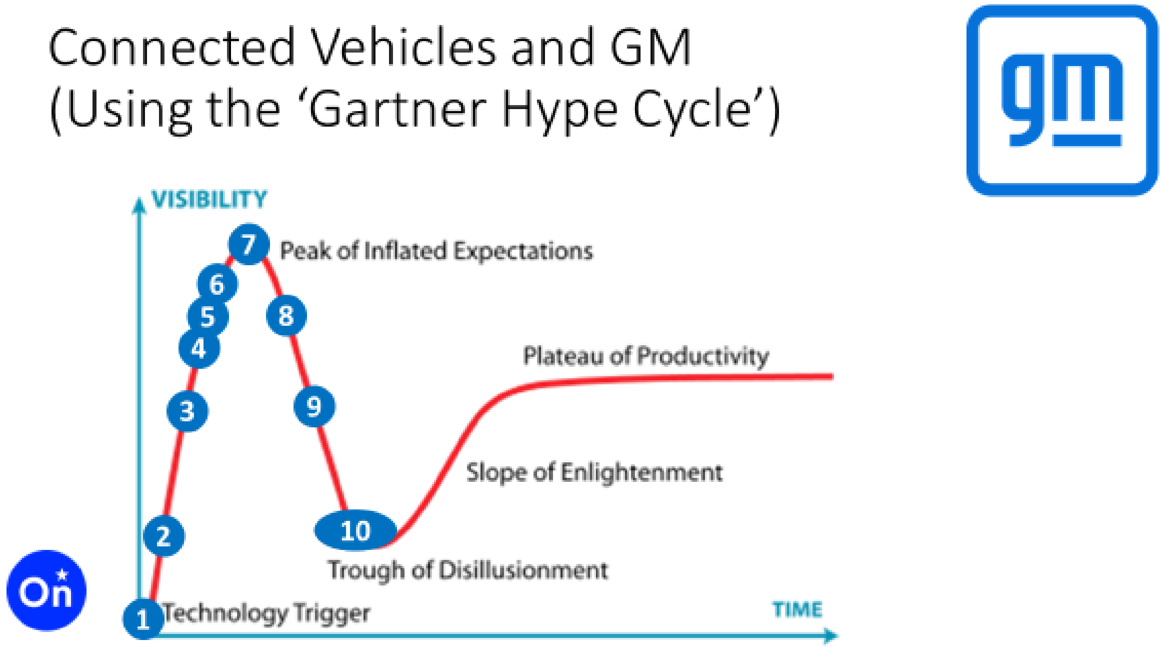


Figure 8: How GM's experience with connected vehicle data follows the Gartner Hype Cycle.

In October 2024, GM announced at its annual investor day that its OnStar unit generates approximately $2 billion in annual revenue, with "high margins." While details of how that revenue is derived are not publicly available, it is safe to assume that most revenue comes from vehicle owner subscriptions. Its primary purpose is to provide vehicle owners and drivers with safety, security, and convenience services.

One has to assume that within General Motors Executive Offices, questions have been and will continue to be asked, "Is it worth the reputational and potential financial damage to continue licensing ANY data to third parties, even those doing public good with anonymous data?" In 2023, before the negative press, GM had transitioned its data broker contract from Wejo, which had gone out of business, to StreetLight/Jacobs. During this transition, GM implemented more stringent privacy protections, which have reduced the sharing of GPS waypoint data from the starting locations of many vehicle trips. Overall, the available data is still massive and useful, but it is now less useful for origin-destination and trip analytics applications in particular. Over time, it is possible that this measure will remain in place, additional restrictions could be implemented, or data dissemination could cease altogether. The Coalition must continue to monitor this situation, as GM data is critical to many vendors' products in the existing TDM and also serves as an example of the same issues that every other source data creator is facing.

Connected Vehicle Use Cases and Agency Feedback

Today, nearly every Coalition member agency leverages CV data to support operations, planning, and/or performance measurement. A survey of Coalition members was conducted in the summer of 2024 to gain a deeper understanding of the adoption and desire for CV data solutions. Twenty responses were received from operations and planning leaders from 16 state agencies across the Coalition, yielding representative results from a diverse array of states.

More detailed results and comments are included in Appendix C. Key conclusions from the member survey include the following:

- EVERY agency was interested in more than one CV data use case.
- Uniform interest in Coalition collaboration on CV data for cost reduction reasons, budgets, and procurement policies.
- The top five types of CV data that agencies were interested in receiving from providers included (see Figure 9):
    - Locations of hard braking events.
    - Locations of debris on the roadway.
    - Waypoint data for trips.
    - Data about road weather conditions (like snow or rain).
    - Locations of crashes detected from airbag deployment or other sensors.
- The most common barriers agencies cited to purchasing CV data were a lack of funds to purchase the data itself or a lack of funds to integrate the data with their ATMS or other existing systems (see Figure 10).

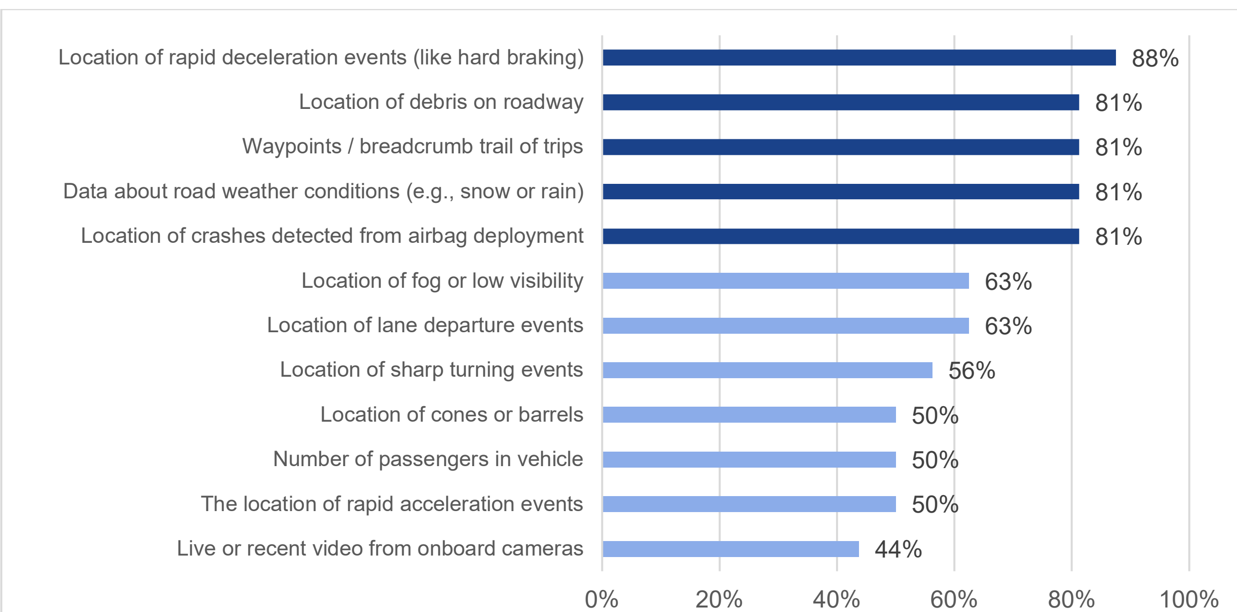


Figure 9: What types of data would you like to receive from CV data providers?

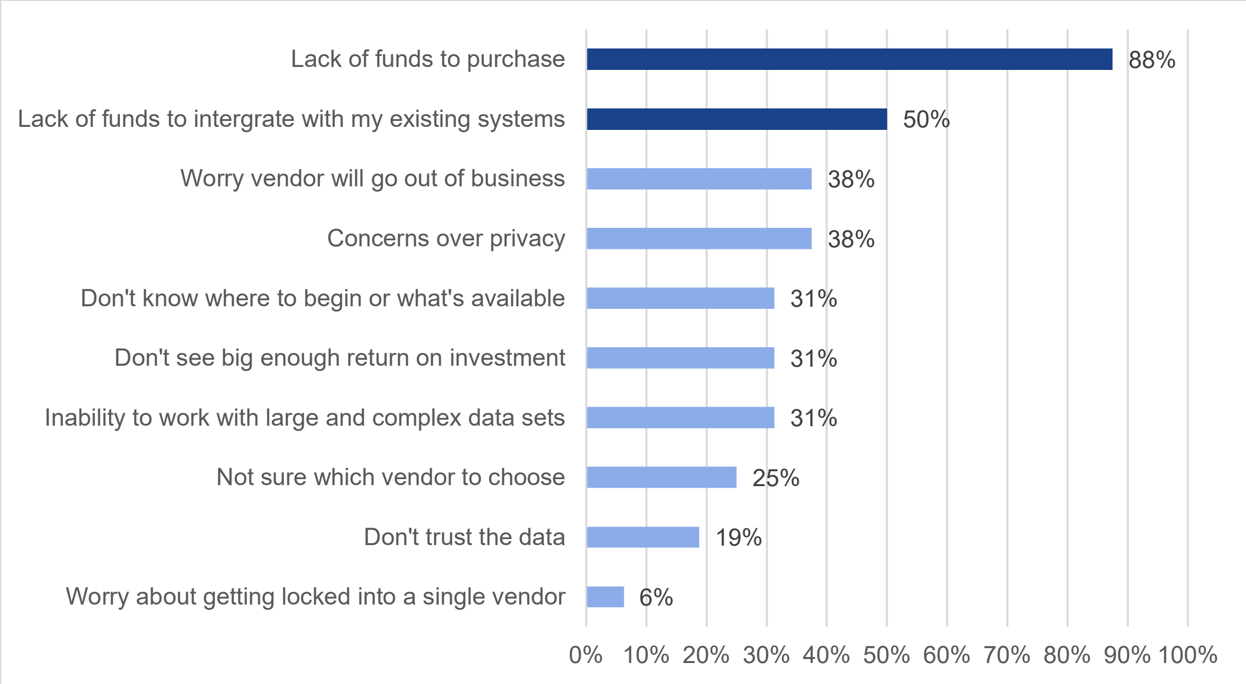


Figure 10: What might prevent you from purchasing CV data?

TETC Future Data Opportunities: Moving from an Enabling Role to an Active Champion

A previous section highlighted the key role the Coalition has played in advancing connected vehicle data-based applications. The Coalition's role to date has largely been supportive, creating an environment that enables its member agencies to tap into connected vehicle data and implement innovative use cases. While this role has been incredibly significant and unique, the Coalition has an opportunity to further strengthen its leadership position by taking a more active role in the CV data environment to advocate for the needs of public sector agencies.

The Coalition has repeatedly demonstrated its unique ability to allow member agencies to pool resources for contracts and projects. The Transportation Data Marketplace is a clear example. However, it is more of a framework agreement that allows members to procure services from vendors at pre-approved terms and fees that each vendor attests are their best possible pricing.

This SCOOP study is focused on a deeper level of coordination, where CV-data-based service(s) could be acquired in more of a pooled-fund fashion.

The Coalition brings multiple unique capabilities to an agency's use of CV data regarding economic and technical benefits.

- **Economics**
  - Economies of scale: By offering prospective vendors a larger region to provide datasets and/or solutions, each member's costs will likely be less than if they acquired data on their own, potentially significantly less.
  - Contracting ability: The TDM marketplace illustrates the Coalition's ability to establish a multi-agency contract vehicle. The effort to create such contract mechanisms is significant, but not much more than a single agency creating a single contract for its own use.
  - Resource pooling: As an example, this SCOOP project is funded by investments pooled from multiple Coalition member agencies.
- **Expertise**
  - Coalition and university staff have developed almost two decades of domain and process expertise that is available to assist agencies in assessing, procuring, and implementing CV data solutions. The Coalition can use this expertise to advise and support members as they continue the TDM into the future.
  - The Coalition has a long history of sharing expertise between agency staff, effectively expanding the knowledge base of each member agency. Countless cases exist where one agency has created a new application based on CV data, shared its experience, and this has led to broader, multi-agency applications (e.g., network-wide implementation of travel times on DMS).
- **Flexibility**
  - The TDM provides member agencies with complete flexibility to decide whether they wish to procure products through the TDM, specifying what products and for how long, thereby allowing the agency to align data purchases with its needs and budgets.
  - The Coalition's committee structure and outreach efforts allow agencies to participate as much or as little as they desire and time allows. This structure supports participation ranging from leading and driving Coalition efforts in a given area with committee leadership to educational and training opportunities that allow individuals to learn about current developments and available resources.

Recommended Actions

In many ways, the use of connected vehicle data by public agencies is at a crossroads. With nearly every vehicle now a potential data generator, opportunities abound. But the business side of CV data is perhaps in the most flux it's ever been. The Coalition can take action to protect the benefits of using CV data both today and in the near term. Assuming success, the Coalition can seed the next wave of growth to further increase the impact and cost-effectiveness of CV data use.

Who can generate CV data, how it is created, and what it can be used for is a vast, constantly changing topic that defies simple summarization. For this effort, we have chosen to focus on datasets and applications of interest to public agencies, assess their current status, and identify actions the Coalition could take to continue improving CV data utilization and impact.

Existing Datasets

The four primary dataset types available in the market today are assessed below: speeds, volume, turning movement counts, trips, and signal performance measures. The tables below describe market health, pros, cons, risks, and actions for each dataset type.

Speed Datasets

Speed datasets represent the original use of CV data by public agencies and are available to agencies in a healthy, functioning marketplace. Some of the strengths of CV speed data sets include healthy price competition, a wide range of vendors for agencies to choose from, and robust methods available for data validation. Challenges include non-standardized roadway segmentation among vendors and a black-box calculation methodology that can make it difficult for users to understand how data quality changes when a vendor's mix of data suppliers changes. While speed datasets are likely the safest from the perspective of market disruption, the risk of major data suppliers leaving the market is not zero. The Coalition should continue to use its influence to advocate for the need for speed data sets and work to ensure their continued availability.

Market Health	• Healthy, functioning market.
	• Speed data is the original, most mature, and most successful use of CV data by public agencies since its first enabling in 2008.
Pros	• Widely available via multiple vendors.
	• Price competition has been healthy, with the Coalition's VPP and TDM programs facilitating meaningful economy-of-scale pricing benefits for its members.
	• Industry innovation has enabled expanded road coverage, better data precision and granularity, and (relatively) quick additions to key newly built roadways to datasets.
	• Data validation provides assurances of data quality and flags any noteworthy changes.
	• Many platforms are available to archive and analyze speed datasets, create additional derivative datasets (e.g., bottlenecks and queues), and generate actionable information across agency TSMO and planning functions.
Cons	• The lack of road segment standardization beyond the lowest common denominator of TMC segments has made interchangeability between vendors difficult for agencies once systems are implemented. Still, this is a relatively minor negative to an otherwise healthy market.
	• Lack of transparency regarding the source data used to create speed datasets, particularly if those changes negatively impact data quality or result in output changes that artificially create trend changes when none exist, muddying analyses.
Risks	• Vendors must acquire access to streaming GPS waypoint source data of sufficient quality and quantity to generate speed datasets.
	• Multiple underlying GPS source data providers license data to multiple companies that then generate speed datasets in what is a generally healthy functioning market.
	• While it might seem unlikely that these sources, which have been in place in some form for nearly 20 years, would be interrupted, the impact of such a "black swan" event[8] cannot be overstated.
	• If the number of source data providers is reduced, end-user (agency) pricing and product stability risk increase significantly.
	• The worst case is for a critical mass of GPS data sources to dry up, and the market for speed datasets becomes irrational in terms of pricing and/or ceases to exist altogether.
Actions	• It is in the public agencies' best interest to ensure a healthy and stable market for speed data.
	• The risk of disruption of GPS source data is low but not zero, and has increased over the past few years.
	• The Coalition should use all tools in its toolbox to promote and ensure that the maximum flow of GPS source data continues.

[8] Theory of black swan events is a metaphor that describes an event that comes as a surprise, has a major effect, and is often inappropriately rationalized after the fact with the benefit of hindsight. Source: https://en.wikipedia.org/wiki/Black_swan_theory

Volumes and Turning Movement Count Datasets

The marketplace for volume and turning movement datasets is still considered to be emerging, but it is becoming healthier over time. However, the array of offered data sets remains inconsistent and confusing. In terms of strengths, the quality of data is improving (with some approaching FHWA's AADT accuracy and precision benchmarks). Data is also available from multiple vendors. However, these data products can inherently be more complicated than traditional speed data products. There is also concern about the stability of the underlying data sources that feed these datasets, which could make trend analysis more difficult.

Market Health	• Emerging, relatively promising market.
	• A confusing and overlapping array of datasets is being offered.
Pros	• Multiple volume source vendors.
	• Data quality is improving, as evidenced by the Coalition's soon-to-be-published most recent validation report.
	• Several vendors are on the verge of meeting FHWA's accuracy and precision AADT benchmarks.
	• This dataset benefits from investments previously made by vendors to generate speed datasets. So, in many cases, the costs to generate are amortized as compared to speed datasets.
Cons	• Inherently more complicated dataset(s) than speeds—to illustrate, the Coalition validates FIVE different volume dataset types (AADT/ADT/AHDT and specific daily/hourly archives).
	• Limited validation of turning movement counts to date.
	• Since trending is a key element of volume analyses, data stability is important—it is unclear how stable the underlying sources and expansion factors used to generate datasets have been and will be in the future.
	• Reliance on count stations for calibration prevents vendors from establishing complete control of the quality of their datasets.
Risks	• The lack of alignment on desired core datasets among agencies may not end the "wild west" of vendor and dataset proliferation anytime soon.
	• Continued source data supplier turmoil could slow down the improvement of dataset quality by vendors.
	• The dependency on count station data is a risk vendors cannot mitigate alone, unless novel scaling and calibration approaches are developed that don't require agency data.
Actions	• GPS source data market health and stability, described in the Speeds Dataset Analysis, also applies to volumes and turning movement counts.
	• Count station data stability will also directly affect dataset quality and utility, and the Coalition should consider how it can best coordinate with its member agencies to define and maximize calibration data stability.

Trip Datasets

The health of the trip dataset market is promising. Still, risks to market stability are beginning to emerge, including haphazard restrictions on source data intended to improve user privacy that may reduce the value and utility of trip data for agencies. This is an area where the Coalition can use its strength to advocate in the interest of agencies to ensure data is sufficiently anonymized while preserving its value for route analysis.

Market Health	• Promising, but the risk of instability is increasing.
	• There is a flux of GPS source data supply and different, changing, and seemingly haphazard limitations on complete trip data provision that increase the risk of the utility of downstream trip and OD analytics being degraded.
Pros	• Demonstrated functioning market where quality trip data can be supplied, tools and platforms can be created, and agencies can fund, leading to a potentially healthy market.
	• Multiple existing Trip Analytics tools and platforms are available for use.
Cons	• Quality trip data, as described in “Pros,” peaked before Wejo’s bankruptcy in 2023.
	• Unclear if any trip dataset of scale is currently “complete” and available to support credible analysis, and not lacking in sufficient trip visibility (origin, trip path, and destination), and/or filtering some trips out completely.
	• Many trip datasets are only available within a vendor's proprietary platform, with a lack of transparency leading to significant uncertainty in analysis results.
	• Results from different platforms cannot be combined or compared.
Risks	• Incomplete trip datasets render analyses useless or inaccurate to such an extent that applications leveraging trip datasets to characterize road network demand and travel patterns are no longer utilized.
Actions	• Lead a collaborative effort between data providers, vendors, and users to proactively define trip dataset content and uses that balance agency needs for dataset completeness with industry desire to avoid media, public perception, and/or regulatory scrutiny.
	• Trip dataset definitions should ensure all possible trips are included (not filtered), origins and destinations are provided in sufficient detail (some “blurring” perhaps via agreed zonal structures), and trip paths or waypoints along the journey to ensure route choice and trip performance can be characterized while protecting origin.
	• Permitted uses of trip datasets should be established to clearly define uses that can withstand scrutiny of the media, political leaders, and the general public as clearly being in the public interest and not overstepping the data generator's privacy.
	• Trip dataset vendors should attest to “provable permission” from source data providers to make their dataset available to agencies.

Signal Performance Measures (SPM) Datasets

The health of the market for signal performance measures datasets is progressively improving. As the acceptance of probe-based signal performance measures begins to increase, the potential applications of this dataset are game-changing, similar to how probe speed data expanded the landscape beyond roadside speed detection with the Vehicle Probe Project.

Market Health	• Healthy, if nascent, market.
	• Proven approaches being deployed and improved, with game-changing potential.
Pros	• Multiple vendors offering "signal analytics" solutions based on rigorous academic research and broadly agreed-upon performance metrics.
	• Sufficient waypoint data and methods to translate source data into SPMs.
	• Much like speed datasets, results can be compared against "ground truth" to determine if results are based on sufficient data and sound algorithms.
Cons	• No derivative datasets are available; they are only embedded in tools or platforms, limiting transparency and flexibility.
	• Balancing the scale potential of CV data-based SPMs with the need to account for the complexity of real-world intersections to generate accurate SPMs is a work in progress.
	• Performance and accuracy improve with higher waypoint refresh rates (5 seconds or less, ideally), but most waypoint data has lower refresh rates (15 seconds or more).
	• SPM metrics mimic detector-based solutions, including an upstream detector that has limitations when queue lengths are longer than the upstream virtual detection point (note this is likely only a temporary limitation and one that hardware solutions will continue to have).
Risks	• Requires GPS waypoints in sufficient quantities, accuracy, and update frequency; if waypoint availability is disrupted, market development for CV data-based SPM could be disrupted.
Actions	• In terms of GPS waypoint source data, there is nothing additional needed that hasn't already been described for previous datasets to stabilize and expand availability.
	• Consider defining a desired export format for signal and movement metric results to enable combinations and comparisons across SPM vendor datasets.

Emerging Datasets

Based on market assessment, agency feedback, and the report authors' cumulative experience, out of the multitude of possible additional CV datasets, the Coalition would be best served to focus its limited attention on three specific emerging dataset types: waypoints/trajectories, driving events, and images.

GPS Waypoints and Vehicle Trajectories Datasets

GPS waypoints are instrumental in vendors generating speed data and have been available in some capacity for nearly twenty years. However, for various reasons—licensing costs, data integration costs, complexity, and use cases—direct access to waypoints by agencies in a systemic and substantial manner has yet to occur.

In 2020-2021, the Coalition led what is perhaps the most extensive examination of trajectory data to date, conducting a pilot project to investigate direct, real-time access to Wejo waypoint data for operational use cases, including hurricane evacuation monitoring.[9] The project generated positive feedback from nearly all agency participants, who saw the promise of understanding the location, volume, and movements of vehicles in key locations and situations in essentially real-time. However, the complexity of establishing ongoing programs to ingest and utilize massive datasets, combined with Wejo's bankruptcy, has slowed meaningful progress since the project concluded.

With at least three known (and willing) sources of real-time waypoint data for passenger vehicles and several others for commercial motor vehicle data available for licensing, the timing is good to refocus on leveraging waypoints and resulting vehicle trajectories. Proof-of-concept projects and applications that leverage historical archived and real-time data are candidates for consideration and are discussed in more detail in the Priority Actions summary. Furthermore, any traction gained from the direct licensing of waypoint data can help stabilize the overall CV data market, providing an added benefit to Coalition efforts to protect the health of the CV data supply chain.

Things to Learn:

- What is the value of a bulk historical dataset to assess network performance in new, innovative ways, such as limited access road network risk assessment?
- Can vehicle trajectory datasets be used for generative AI/ML training, such as high-resolution 3-D trajectory plots for queues and bottlenecks?
- Is it possible for vendors to establish interactive real-time APIs to limit agency data ingestion and processing costs and complexity?
- What are the effective penetration rates as compared to total vehicle volumes?
- How close to the presentation of real-time conditions is possible?
- Is lane-level detail possible, or at least multi-channel flows on larger 3+ lane highways?
- What is the value of visualizing waypoints and trajectories for operations staff interpretation versus decision support analysis tools?

Driving Events Datasets

Datasets that include hard braking events have been available for evaluation and licensing to some degree for nearly a decade. However, a solid framework for using driving event datasets in public sector applications (what to license, what it’s worth, and how to use driving event data, including but not limited to hard braking data) has not yet crystallized. The value of driving event data increases when viewed alongside other data types; thus, it may make good sense to evaluate

[9] Report: https://tetcoalition.org/wp-content/uploads/2021/03/MT2008_Wejo_HurricaneReport_2021.pdf, webinar presentation: https://tetcoalition.org/wp-content/uploads/2019/12/TSMO-Hurricane-Final-Slides-01-28-21-Final-ver2.pdf

it alongside and as a complement to waypoint data. Since multiple vendors license both waypoints and event data, seeking both in the same contract and evaluating uses in the same proof of concept project(s) makes logical sense.

Things to Learn:

- How do we assess the true value of hard braking events? Does it make sense to determine network-wide risk values, or are benefits better suited to certain roadway types over others?
- What "weather" event data is available and is truly useful for public agency applications?
- If waypoint data is available, what events can be derived from waypoints versus a specific event data stream type (e.g., rapid deceleration detected in waypoints as opposed to a hard braking event)?
- What is the ability to obtain datasets in real-time versus bulk data transfers?
- How transparent is the logic of event generation? Can we develop the ability to compare or combine event data of the same type from different vendors?

Imagery Datasets

Dashcams are proliferating across all types of vehicles. In cases with forward-facing cameras, they offer visual insights for agencies theoretically everywhere of interest on the road network. While it is unclear if any vendor will offer passenger-deployed dashcams in the near future, those deployed in commercial vehicles and/or shared mobility providers offer viable near-term options for mobile images. While Vizzion is the only vendor currently licensing imagery from outward-facing in-vehicle dashcams, the promise of this new data stream warrants attention and potentially support to cultivate. The ability to obtain imagery in a cost- and system-scalable manner offers an intriguing complement to other CV-based datasets, with the promise of supporting use cases across TSMO, asset management, and several other potential agency functions.

The North Carolina Department of Transportation (NCDOT) has piloted and is licensing imagery from Vizzion.[10] RITIS is also building an interface to Vizzion mobile imagery for inclusion across many real-time and historical analytics modules. Vehicle-based imagery could be an ideal area for the Coalition to focus on, accelerating the availability of more data from more vehicles, as well as application development, learning, peer exchange, and adoption, if the promise of initial pilots continues in operational scenarios.

Things to Learn:

- What is the current penetration rate of imagery datasets? Can enough imagery be provided to support a critical mass of use cases and make it worth the investment?
- If or when the penetration rate of imagery datasets grows, what applications could be enabled?
- Is this dataset suitable for real-time use?
- How do we combine this data with other datasets, such as bottlenecks, queues, work zones, and the other emerging datasets (trajectories and events)?

[10] https://vizzion.com/Iteris-NCDOT.html

Priority Actions for Consideration by TETC

When considering the landscape related to the above CV-based datasets, and where the Coalition's unique status and collective experience lies, we recommend actions be considered in three overlapping and mutually reinforcing areas:

1. Industry engagement and collaboration to leverage the Coalition's gravitas and experience in promoting and accelerating industry stabilization, health, and growth.
2. Leverage the timing of the end of the base term of the current TDM contracts to optimize the impact of the next generation of the program, keeping the Coalition as the lead related to CV data use.
3. Enabling the next generation of CV data applications through pilots, proof of concepts, and, where appropriate, application development, especially as it relates to safety and core TSMO needs.

Action 1: Stabilize the CV Data Supply Chain

As the largest collective market for public sector use of CV data, the Coalition and its members have the loudest voice to impact the CV data supply chain for agency uses. This market clout provides both an opportunity and an obligation to educate source data generators, aggregators, and application developers on the available market opportunities. Possible actions include:

- Establish ongoing contact with all possible source data creators to maintain and expand the pool of available CV data for public sector use. Consider a Request for Information process to expedite the process. In addition to direct communications with known creators and contacts, engage with the Alliance for Automotive Innovation[11] (the national association of automotive OEMs).
- Head off possible negative consequences of current media stories regarding CV data use and sharing by developing case studies (or encouraging members to do so) to demonstrate the value of CV data to agencies, and how data is not misused or abused.
- Focus on trip datasets and waypoints to identify dataset models that align with the user community and source data generators. The purpose is to enable robust trip analyses while satisfying privacy concerns.
- Develop approaches to encourage more source data availability and improve overall industry stability:
    - Investigate, possibly pilot, data "clean room" solutions.
    - Requirements development in emerging CV data areas ("if a supplier can supply 'this,' then we can procure it since it supports functions x, y, z").
    - Consider approaches that accommodate integrating source data from multiple vendors to improve penetration rates needed for specific applications.

Initiatives for Consideration

The Coalition may wish to consider formalizing the action items above into more specific initiatives, as described below:

1. **Automotive industry outreach** – Engage with the Alliance for Automotive Innovation to establish direct communication channels between agencies and automakers, who are the primary generators of most connected vehicle source data. Ascertain if collaboration is possible to establish practices and principles to balance the needs of consumers and manufacturers with those of public agencies to derive maximum benefits while minimizing commercial, legal, public relations, and political risks.

[11] https://www.autosinnovate.org/

2. **Trip data truncation best practices** – Collaborate with key source data generators to develop recommended best practices for trip/waypoint truncation to obtain maximum benefit.
3. **Data RFI** – Conduct a Source Data Generator RFI to obtain up-to-date and "on the record" information from potential data vendors such as trips/waypoints, event data, and imagery.
4. **Case studies** – Publish a synthesis of TETC member case studies demonstrating the value of CV data-based applications that have no privacy implications.

Action 2: Optimize TDM for the Next Generation of the Program

Since the initial 2007 TDM RFP, the Coalition's Transportation Data Marketplace has evolved to accommodate and capitalize on the growth and evolution of the overall CV data market. The initial period of the current Transportation Data Marketplace master contracts ends June 30, 2026. If the Coalition determines it is in the membership's best interest to initiate a new RFP process, there are several areas for consideration to further expand the scale and benefits of the TDM. Potential updates are listed below, grouped by thematic areas:

Increase Program Flexibility

Each update of TDM, from VPP1 to VPP2 in 2014 and to the current TDM in 2021, allowed the Coalition to adapt the program to market conditions at the time. Between procurement cycles, however, accommodations were needed to allow for the introduction of emerging datasets and services under "optional services." Given past experience and the expected flux in the CV data supply chain over the next few years, the Coalition can build flexibility into the core procurement process. While Federal, state, and university procurement laws and guidelines govern and may prohibit some actions upon review, actions to consider include:

- Shift to more of a Qualified Products List or GSA Multiple Award Schedule model, establishing simple baseline requirements for known datasets and services, then establishing master contracts for any vendor that offers products that meet or exceed baseline requirements (and agree to the foundational terms and conditions that make TDM the successful program that it is).
- Shift to a "call for products" cycle every one or two years, allowing new entrants to join and current vendors to revise and update their products and pricing list to maintain the most up-to-date portfolio in the TDM (while continuing to allow multi-year task orders, based on the products and pricing of the master contract in effect at the time of task order completion). Include in each solicitation cycle a statement of interest for any new types of source data, applications, tools, and platforms that the Coalition and its members would like to see added to the TDM.

Increase Purchasing Power

With roughly a dozen member agencies currently contracting over $10 million annually through the TDM, this contract vehicle is by far the largest single ongoing mechanism for licensing CV data and associated services for public agency use in the US today. As a result, the Coalition has market clout. It should work to positively leverage and expand that clout where possible. Possible actions that would increase the collective purchasing power of participating agencies include:

- Require all qualified products to have understandable pricing models and tables, including documentation of prices and discounts (if any) for multi-state and multi-year acquisitions.
- Explicitly encourage pricing models for each product that improve unit pricing as the size of requested services increases, both for larger geographies and multi-year contract terms (i.e., leverage volume pricing and economies of scale).

- Request or require "Most Favored Customer" pricing declarations for proposed or accepted products. Note that this is required for GSA Multiple Award Schedule (MAS) contractors.
- Expand the contract vehicle to willing public agencies outside the Coalition, thereby increasing vendor opportunities and, in turn, enhancing Coalition purchasing power.
- Encourage agencies to synchronize and pool contracting resources in cases where the same product is desired. This could potentially yield further savings than single agency contracting.
- Clarify the pay-for-performance and clawback elements of contracts to balance the needs of vendors and members for cash flow simplicity with the necessary incentives to meet performance requirements, while also easing the complexity of contract management and interpretation for contract administrators.

Maximize Transparency of Products Offered

With an increasing number of vendors and products available within the TDM, confusion over exactly what is offered and delivered can increase. The Coalition should continually seek to enhance the transparency of what is offered to ensure optimal agency decision-making and informed investments. Examples of actions include:

- Continue the validation program under agency guidance—identify qualified products as validated, pre-validated (to be validated but not yet), or not validated to offer transparency for agencies.
- Require all qualified product vendors to demonstrate "provable permissions" of their right to provide the dataset or service they offer and provide transparency regarding the stability of these permissions over the full term of the master contract.
- Ensure that each vendor or product adheres COMPLETELY AND EXPLICITLY to the TDM data use and licensing terms and model that has been critical to the TDM's long-term success. Any deviation from the terms for a specific vendor or product must be well-documented and agreed upon in advance by the Coalition, its members, and the contracting team.

Action 3: Explore New and Emerging CV Data Sets

Vehicle trajectories, driving events, and on-vehicle imagery datasets all offer the potential to significantly enhance the utility of CV data for public agencies. Yet all are nascent in their productization for agency use. The Coalition has the opportunity to lead in leveraging these datasets and combining them with existing datasets and platforms to make the potential future a reality. For reasons explained in Appendix E, initial efforts should focus primarily on limited-access road networks.[12] Collaboration between Coalition staff and its members, who have the interest, experience, and expertise, can refine an action plan. We include the following strawman actions for initial consideration, debate, and refinement:

- Target process and application development to increase the use of CV data in safety planning and emergency incident management, particularly on the limited-access road network ("strategic road network").
- Define an approach to creating a "strategic road risk assessment" using licensed historical CV data and digital mapping. The goal is to establish a repeatable baseline analysis method that can inform state or agency safety planning and programming processes.
- At minimum pilot, and ideally conduct for all TDM participating states, an initial "strategic road risk assessment" process as part of the definition process—pavement roughness,

[12] See Appendix E for background on the rationale for focusing on these roads

- weaving, ramp spillbacks, lane-by-lane speed variability, wrong-way driving, secondary crash risk, etc.
- Focus on real-time application development to increase CV data use across the myriad of strategic network high-risk use cases—incidents, work zones, evacuations, weather, planned events, asset management, etc.
- Encourage, lead, and amplify efforts to test, evaluate, and utilize CV imagery data across TSMO applications.

Case Studies: Real-Time Trajectory Data

To date, research related to the use of trajectory data has been performed on historical data. Opportunities are increasing for agencies to directly license "live" GPS waypoints, as occurs today with data aggregators. As highlighted, several technical and commercial questions remain unanswered. But there is little doubt real-time feeds could be a game-changing enhancement to traffic operations and safety research. Several use cases for real-time trajectory data are described below.

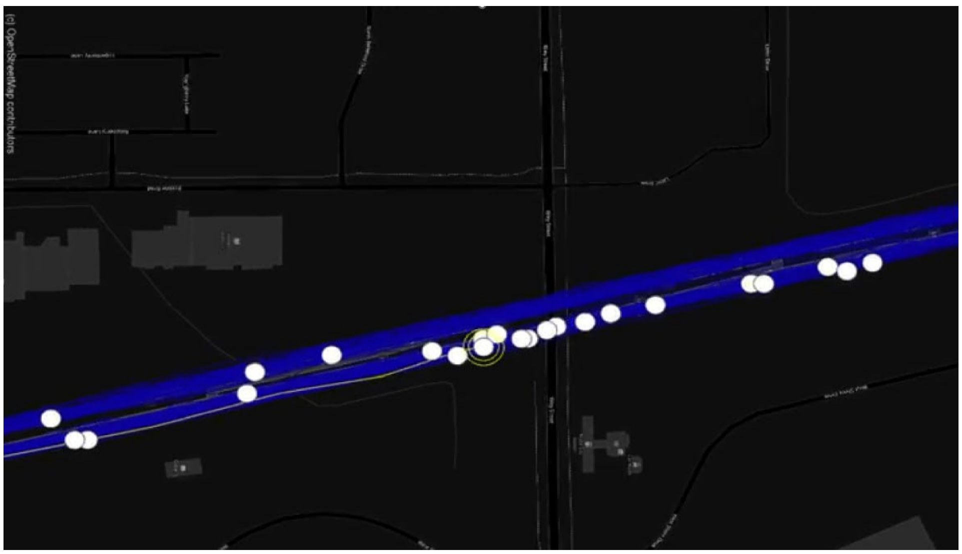


Figure 11: Representation of real-time vehicle trajectories on a limited-access highway.

Real-Time Movement Visualizations

- *Real-time detouring:* With real-time trajectory data, traffic management centers could visualize granular traffic movement on a map in real time. This would allow TMC operators to observe real-time detouring around crashes, work zones, and police activity. Traffic detouring from a limited-access highway is often diverted onto signalized arterial roads, which can be overwhelmed by the influx of traffic. If agencies can see the detour routes cars are actually taking in real time, they can activate the appropriate traffic signal timing plans to better accommodate the detouring traffic.
- *Evacuation monitoring:* During evacuations, agencies could more accurately estimate the number of vehicles taking each evacuation route and understand how many vehicles are taking official vs. non-official routes. Agencies could observe highly specific locations of traffic congestion where additional on-site responders may be beneficial. Officials could also gain insight into how far people travel from home when they evacuate—are most people heading a few miles inland away from the storm, or are they leaving the state altogether? This data could also be used in an after-action report to understand how quickly people leave when evacuation orders are given, what percentage of people in non-mandatory evacuation zones decide to evacuate anyway, and the general timing of when people leave vs. how far they travel.

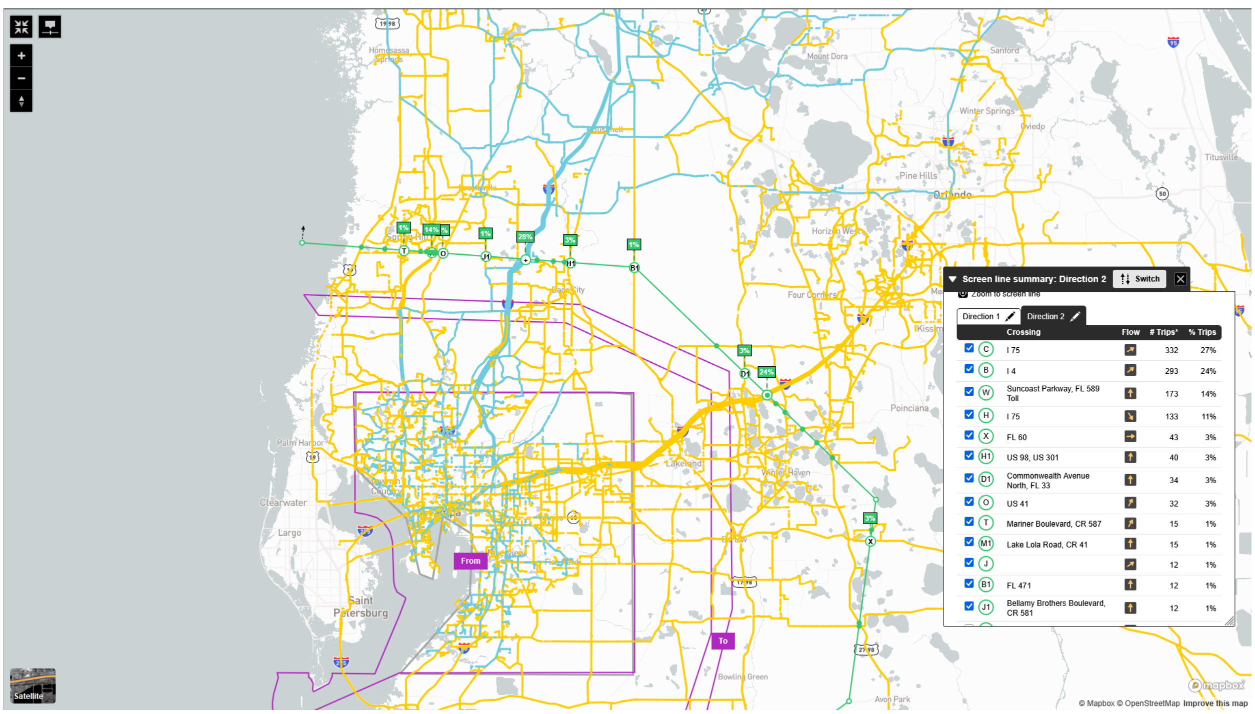


Figure 12: Traffic exiting Hillsborough County, FL, in the two days prior to landfall of Hurricane Milton (2024). This analysis could be improved with real-time trajectory information.

- *Virtual traffic helicopter:* These visualizations could serve as a virtual traffic helicopter that can view conditions across entire regions, which would be especially valuable in areas without roadside cameras or other ITS infrastructure. With vehicle-level trajectory data, agencies could see the real-time length of queues upstream of incidents with high levels of accuracy. Current segment-based data provides an approximate queue length in terms of segment boundaries, but trajectory data would enable queue visualization within a few yards. Law enforcement could also understand how many people were making illegal U-turns and wrong-way movements to get around a road-closing incident.

Figure 13: "Virtual helicopter" applications could enable agencies to view highly granular traffic patterns.

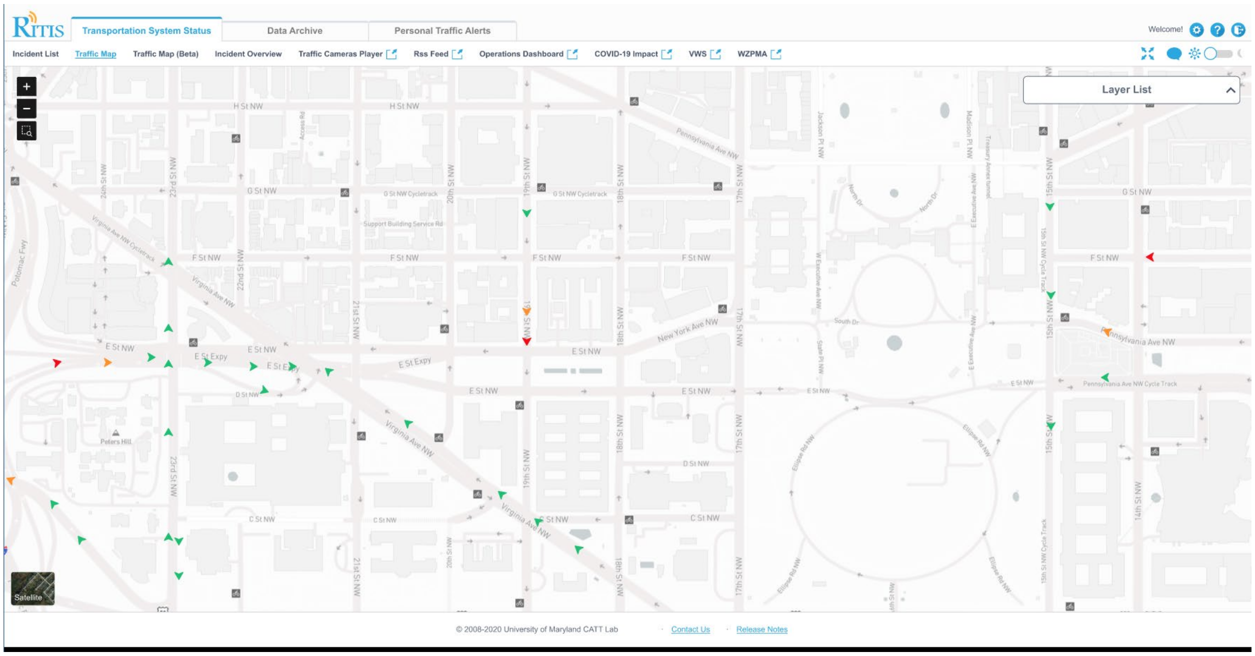


Figure 14: Real-time animation of vehicles moving through DC, color-coded by their speeds. Locations are provided every 3-15 seconds.

By using real-time trajectory data, agencies could also visualize speed differentials between vehicles on the road. Experience has shown that vehicles traveling much faster or slower than the prevailing traffic speed can increase the probability of a crash, particularly in dense traffic conditions. Agencies and researchers could observe these speed differentials in real time and generate risk scores that traffic management centers could use to take proactive steps to reduce this risk. These steps could include triggering active traffic management scenarios, such as activating variable speed limit signs or ramp metering when a risk score exceeds a specific value in a particular location.

Non-Infrastructure-Based Wrong-Way Driving Detection

Wrong-way driving events on limited-access highways can be extremely dangerous. When vehicles are closing at high rates of speed, it can leave very little time for other vehicles to take evasive action. Impaired drivers can often travel on a limited-access highway for many miles before stopping or turning around, exposing many other drivers to the danger of a head-on collision.

Real-time trajectory data could be used to identify wrong-way driving activity as it happens, particularly on limited-access roadways. While CV data is generally not accurate enough to be used on a per-lane basis, it is accurate enough in most cases to identify vehicles traveling in the wrong direction on median-separated limited access highways. Another advantage of this technology would be the ability to pinpoint the location of the wrong-way driver, which can be very difficult to do in practice. Often, when law enforcement receives a report of a wrong-way driver, the vehicle can be miles up the road before police arrive on the scene. By knowing the driver's exact location, law enforcement can get ahead of the problem, and traffic management centers can post targeted traffic alerts for drivers in harm’s way.

Even though real-time trajectory data would not be able to detect every single wrong-way driver, each wrong-way driving event carries potentially fatal risks. Developing tools that mitigate even a small percentage of the events could still improve public safety.

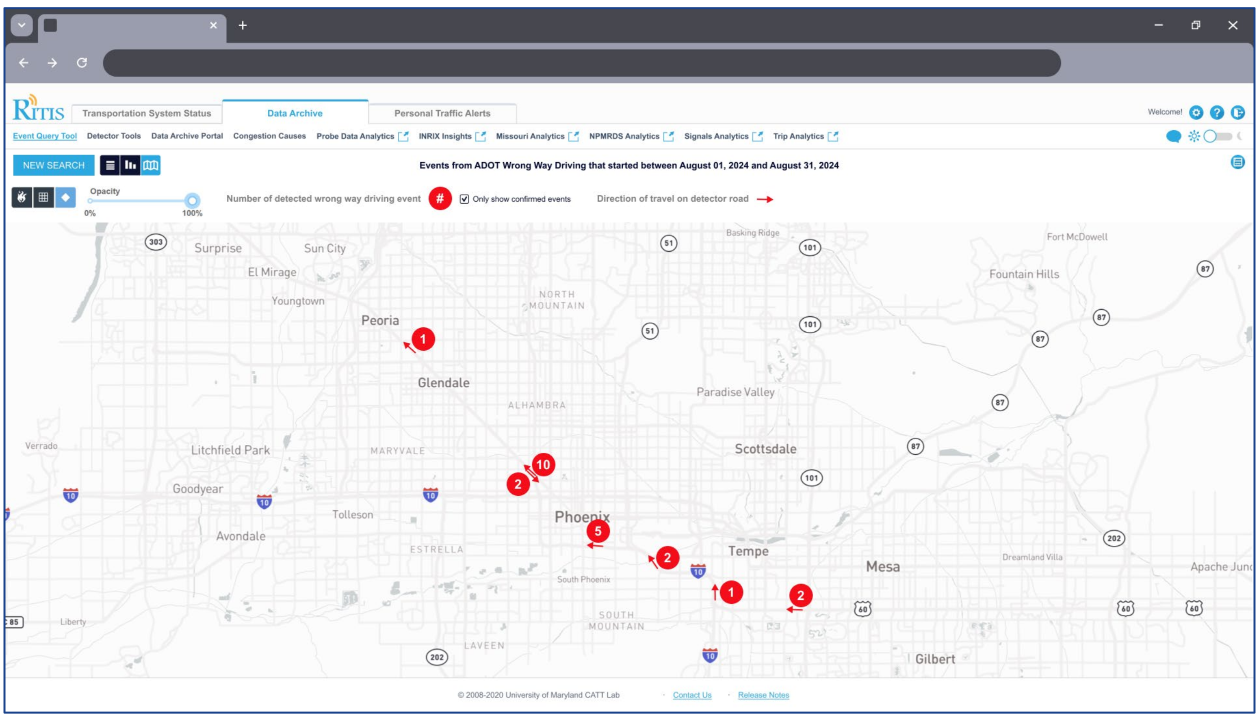


Figure 15: Example of how trajectory-based wrong-way driving alerts could be integrated into RITIS.

Real-Time Data with Vehicle-Specific Information

The trajectory data case study scenarios above could be further enhanced with the addition of vehicle-specific event data. For instance, if fuel consumption information was available during evacuations, agencies could better understand where vehicles were running low on gas and where to direct fuel trucks to interstate gas stations. If vehicle-specific emissions information were provided, agencies could build much better estimates of emissions caused by recurring congestion, poor signal timing, and incidents.

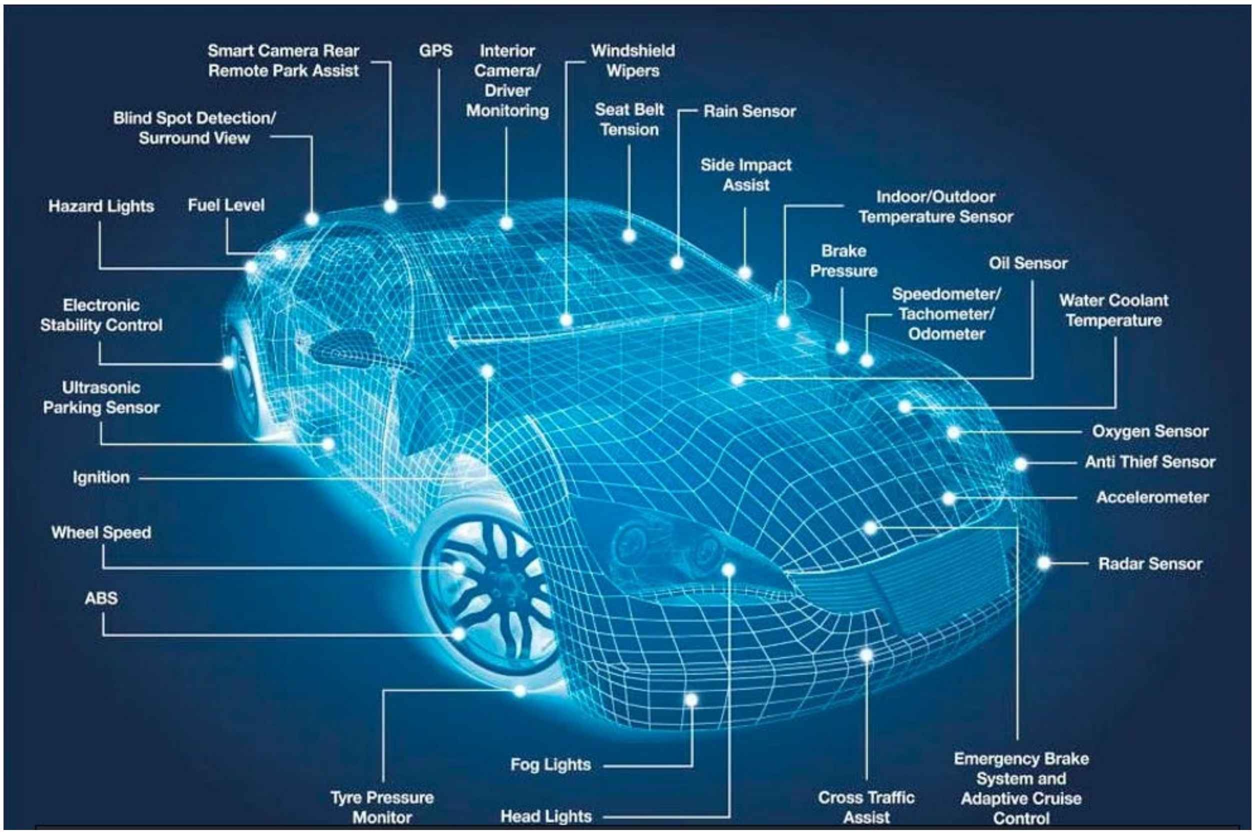


Figure 16: This figure shows many vehicle-specific data elements collected by modern vehicles.

Safe/Unsafe Truck Parking Locations

The lack of safe and adequate truck parking is a major concern for trucking companies nationwide. Federal regulations limit the time commercial drivers can spend behind the wheel, so once their daily allotment of time expires, they are required to park for rest breaks. Even though truck drivers time their routes to account for hours of service regulations, there are often not enough empty parking spaces at many rest areas and truck stops to accommodate the demand. This can lead truck drivers to park in undesignated spaces (like highway ramps) that can block roadways and put other drivers at risk. Some truck drivers park overnight in retail or abandoned parking lots that may not have adequate security, which can put drivers at personal risk.

Real-time trajectory data could enable agencies to identify hotspots for commercial vehicle parking, particularly in areas where trucks are parked in unauthorized locations. Outcomes could include improved enforcement or data to support the expansion of designated parking facilities for commercial vehicles, which could help alleviate the problem.

Figure 17: Examples of truck parking at rest areas, showing trucks parked inside and outside designated parking areas.

Case Studies: Imagery Data

For many years, agencies have relied on roadside closed-circuit television cameras (CCTV) for video in their traffic management centers. CCTVs are extremely useful for TMC operators as they manage responses to incidents, as they can see roadway conditions and the location of responders in real time. Operators also have complete control of the cameras, allowing them to view the scene that is most useful to them. However, roadside CCTV is not without its downsides. Cameras are expensive to install and maintain, so non-urban areas often have incomplete coverage. Also, pole-mounted cameras are fixed to a permanent location, so they can't be moved to a different location on the road to get a better view. (Zooming in on an incident scene doesn't help if a tree or building blocks the view of a crash.) Drone-based cameras can alleviate some of these downsides, but they require responders with the necessary equipment to be on-site, which means drones cannot be relied upon during the very early stages of an incident.

Commercial imagery data could be used to fill these gaps and increase public safety, with a few examples discussed in the following use cases.

<u>Real-Time Incident Detection and Confirmation</u>

Data vendors are beginning to provide real-time imagery to agencies from dashcams installed in commercial vehicles. By integrating these images into traffic management centers, operators could obtain snapshots from areas around crash sites, especially in locations without existing CCTV coverage. University researchers are also investigating how machine learning can analyze images in real time to proactively identify crashes and disabled vehicles, which could then be presented to traffic management center operators as alerts for follow-up.

Figure 18: Example of how dashcam imagery could be shown in RITIS.

<u>Confirmation of Waze Events</u>

For the past few years, Waze has partnered with transportation agencies to improve the data quality on its platform. Waze can receive official incident and construction notices from agencies and provide them to their millions of users. At the same time, Waze has made the real-time, crowdsourced incident information provided by its users available to agencies. These alerts can be a great supplement to official DOT incident feeds, since they can provide alerts for roads not currently covered by DOT cameras. The downside is that crowdsourced information can be duplicative and isn't always as accurate as information provided by trained DOT staff. Real-time dashcam imagery data could be used to validate and confirm user-reported incidents and allow traffic management center staff to report them as official incidents.

Road Condition Monitoring

Operators could also use these images to confirm poor weather conditions, the current status of work zones, and current traffic conditions during special events and evacuations. Current segment-based probe data can be used successfully to identify locations of slowdowns and congestion. Still, without some sort of video confirmation, it can sometimes be difficult to determine the precise cause for that congestion. By pairing in-vehicle imaging with other real-time CV datasets, operators could understand the "what" and the "why" behind specific congestion events, enabling them to respond more effectively.

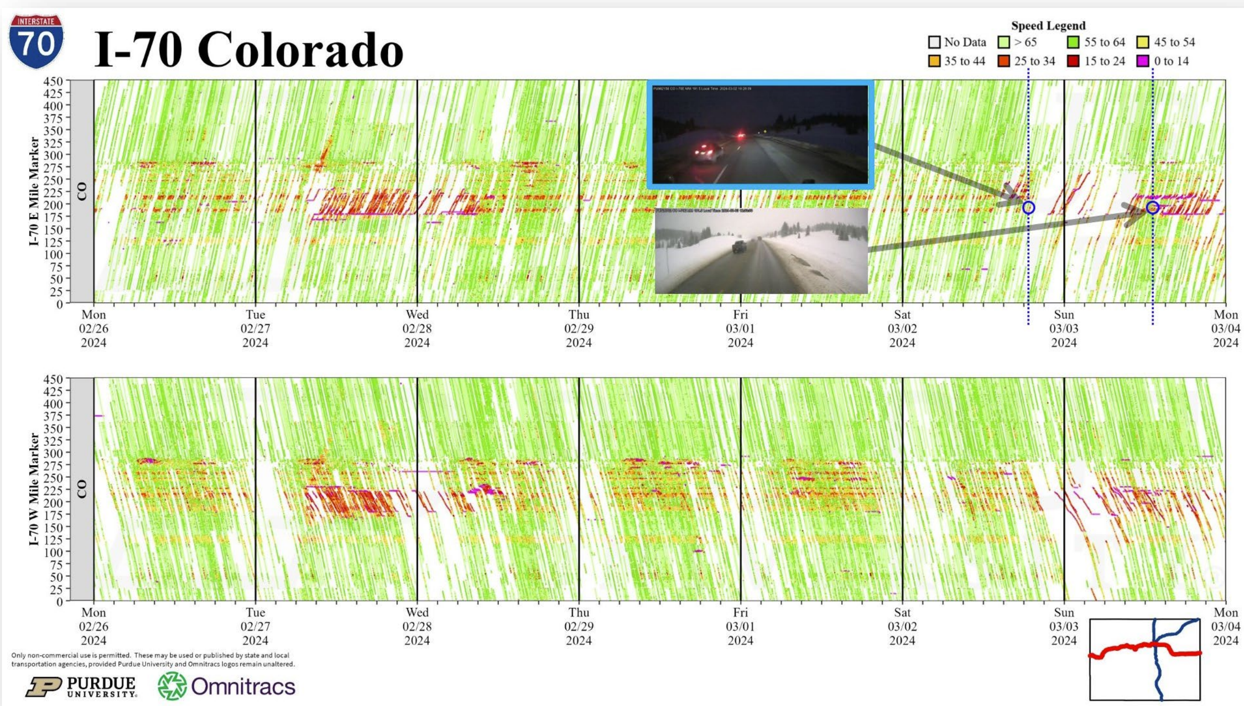


Figure 19: Example of how dashcam imagery can provide context to trajectory speed data.

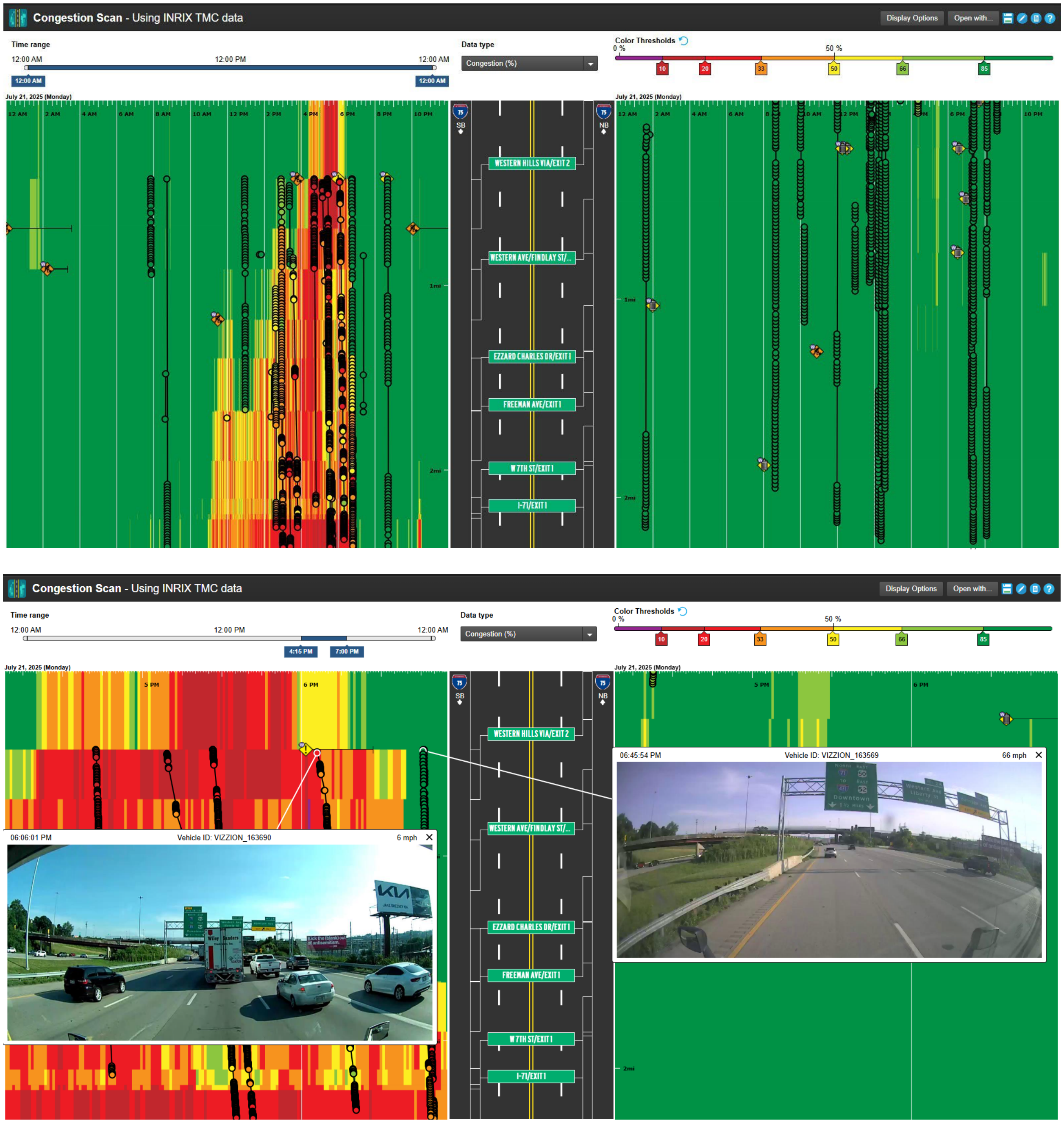


Figure 20: Example of how dashcam images and trajectories could be incorporated into the Probe Data Analytic Suite's Congestion Scan tool.

Work Zone Audits

Highway work zones can be dynamic environments. Lane configurations often change on a regular basis. Some work zones are temporary, appearing for a few hours then disappearing. Dashcam video, paired with AI and machine learning algorithms, can be used to analyze the current state of work zones and identify anomalies. Is the number of lanes closed correct? Are there any cones knocked over into the travel lanes? Does the contractor have permission to have a work zone here in the first place? Automated vehicle-based video feeds have the potential to spot-check work zones multiple times per day to identify issues that construction engineering inspectors might not be on site to catch.

Issues can also occur when work zones are closed for holiday travel. Traffic cones and barricades are left in place during the holiday work pauses, but construction workers are often not on site to monitor if a construction barrel has been struck or if a sign message is out of date. Real-time imagery data from vehicles could be analyzed and used to provide alerts to traffic management center operators in the event of a work zone issue.

Case Studies: Driving Event Data

Today's vehicles generate tremendous amounts of data on safety, vehicle state, and performance. While most of this data remains stored locally on the vehicle, more and more datasets are being transmitted from vehicles either in real time or when a vehicle reaches its destination. By leveraging these datasets and encouraging data aggregators to release even more detailed data, agencies could gain unprecedented insight into safety risks on their roadways and a better understanding of how they could help combat them.

Road Risk Scores

Today, many agencies look at crash rates in their road safety evaluations. While crash statistics are valuable metrics, they are inherently backward-looking and reactive in nature. Driving event data provides agencies with the opportunity to quantify risky driving behavior before a crash occurs. Most risky driving events don't result in a crash, but they do increase the probability of a crash. Areas with large concentrations of risky driving behavior may be at increased risk of future crashes. Driving event data and near-miss statistics could be useful inputs into road safety audits that work to proactively improve the safety of road infrastructure.

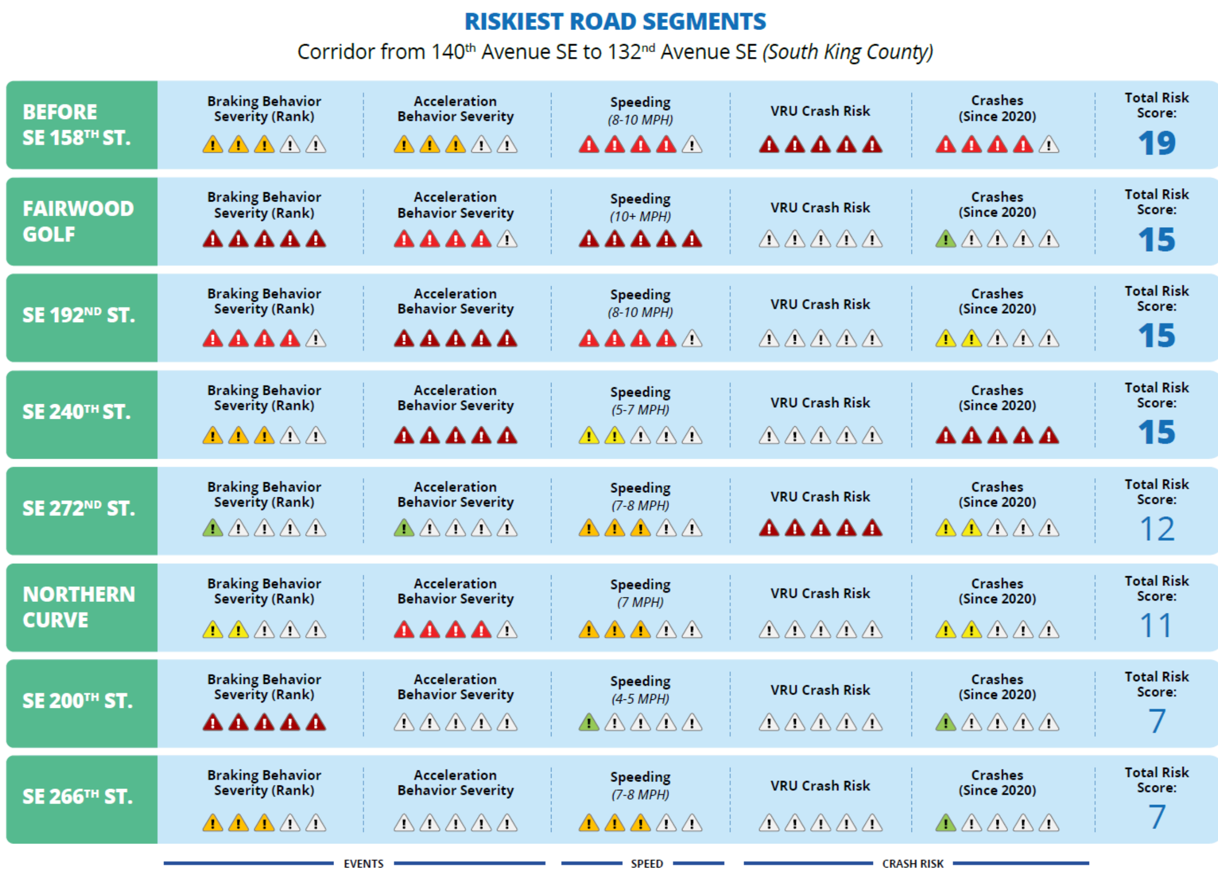


Figure 21: Example of driver event data used to generate a total risk score for road segments

Freight-Only or Passenger-Only Risk Assessments

Commercial and passenger vehicles can have differing levels of risk on the same road segment, due to differences in size, weight, and operational characteristics. Roads with tight turns and steep grades can be treacherous for semitrucks, but easier to navigate for passenger cars. Data aggregators can separate event data from vehicles based on vehicle type, which could allow agencies to develop specific risk scores for different classes of vehicles. This could, in turn, lead to more targeted safety improvements to those facilities. In addition, agencies could develop resources to encourage commercial vehicles to avoid less safe roads and use safer facilities instead.

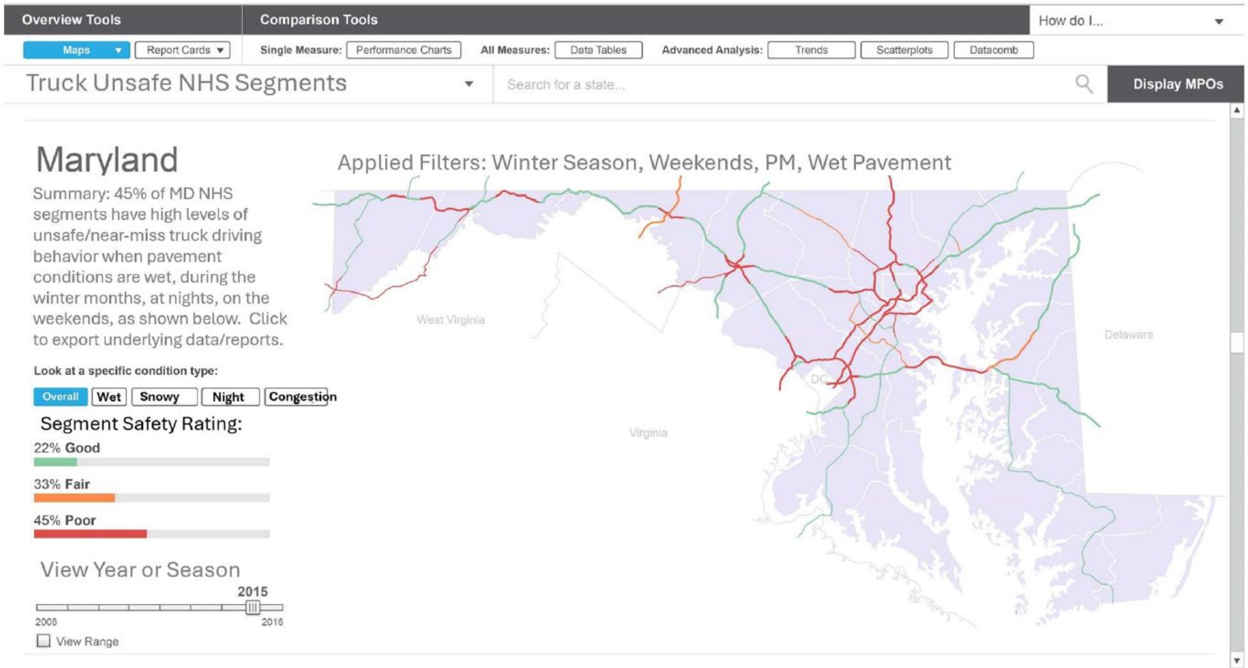


Figure 22: Example of truck-specific safety ratings for highways.

Initiatives for Consideration

The Coalition may wish to consider formalizing the action items above into more specific initiatives, as described below:

1. **Conduct a strategic road risk assessment** – leveraging licensed data waypoints and corresponding events (ideally both passenger and commercial vehicles) to establish a consistent baseline of the relative dangers of the limited access network, ranging from speeds, queues, weaving, braking, etc.
2. **Real-time simulator/translator for RITIS** – create a module to run "any" historically licensed waypoint, driving event, and/or imagery dataset through the simulator to generate derivative datasets (e.g., high-res queues, event logs (WWD, ramp spillback, lane blockage, high speed variations, work zone activations, etc. to populate in RITIS and/or enhance the existing datasets already in RITIS) as a precursor to real-time applications.
3. **Real-time geofencing algorithm for RITIS** – enable generation of real-time geofence-specific requests for source data (e.g., images, waypoints, etc.) based on triggers from other data in RITIS, like weather, work zones, or crashes.
4. **Real-time data visualizer** – enhance the RITIS map and other RITIS modules to display licensed real-time waypoint data in areas and times of interest, including snapping to the road network and creating and showing vehicle trajectories from the waypoints.

Appendix A – Summary of PennDOT's 2024 Connected Vehicle Data Study

This appendix contains slides from a presentation given by Gunnar Rhone, PE, from the Pennsylvania Department of Transportation on November 12, 2024. This report presents the findings of a comprehensive study conducted by Michael Baker International (MBI) for the Pennsylvania Department of Transportation (PennDOT) on the feasibility and potential of utilizing connected vehicle (CV) data for various transportation applications. The primary objectives of this study were to identify and analyze different types and sources of vehicle data, assess the potential use of this data for transportation system performance analysis, measurement, and monitoring, and evaluate the viability of obtaining data from original equipment manufacturers (OEMs) and third-party providers.

THE EASTERN TRANSPORTATION COALITION

CONNECTED VEHICLE DATA STUDY

GUNNAR RHONE, PE – PENNDOT

NOVEMBER 12TH, 2024

pennsylvania
DEPARTMENT OF TRANSPORTATION

AGENDA

- Study Purpose
- Discovery
- Findings
- Vehicle Data Viability
- Recommendations/Opportunities
- Transportation Pooled Fund Study Framework

STUDY PURPOSE

Determine feasibility and potential for CV data

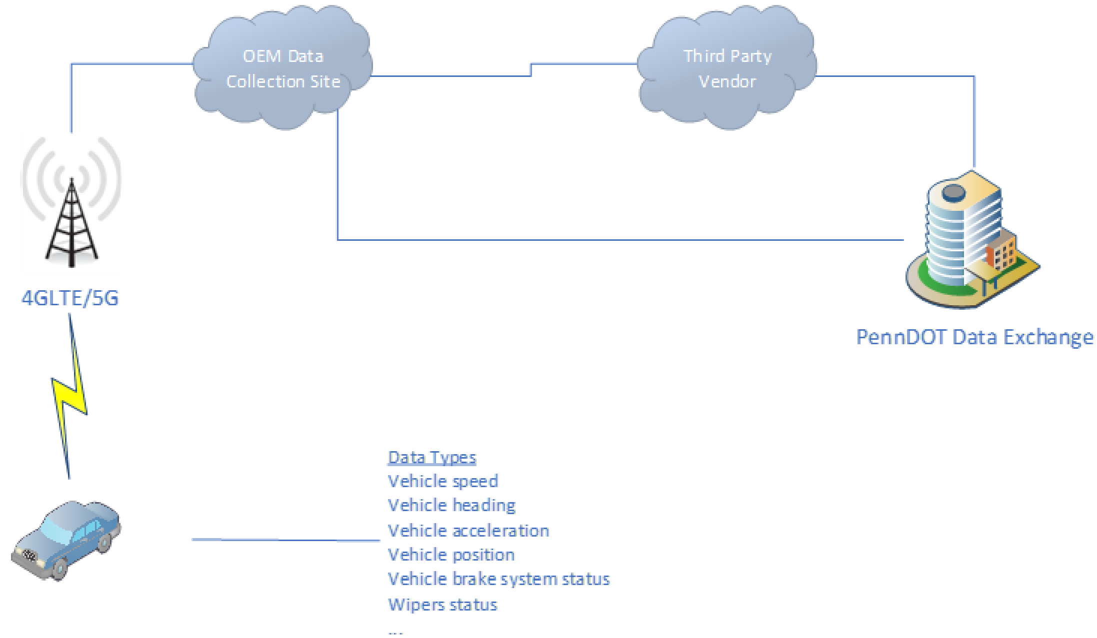
- Identify and analyze different sources of vehicle data
- Assess potential of data for TSMO use cases
- Evaluate viability of obtaining data from OEMs and 3rd Party data providers

USE CASES

Traffic Operations	Safety	Maintenance/Work Zones	Planning
Real-Time Traffic Management Systems	Accident Prediction and Prevention	Work Zone Traffic Flow Monitoring	Traffic Pattern Analysis
Traveler Information	Wrong Way Driving Prevention	Dynamic Work Zone Signage	
Traffic Incident Detection and Response	Work Zone Warning/Notification	Predictive Queue Warning	
Road Weather Information/Weather Impact Analysis	Queue Warnings		
Active Traffic Management	Driver Behavior Analysis		
Traffic Signal Timing Optimization			

TYPES OF DATA ASSESSED

- Traditional V2I (5.9 GHz) Communication
- Fleet Data – OEM Directly
- Fleet Data – Third Party



Example of Vehicle Data Through OEM Directly or Through Third Party Vendor

DATA SETS

SAE J2735 Basic Safety Message (Comparative Data Set)

Part I

- Vehicle Size
- Vehicle Position
  - Latitude
  - Longitude
  - Elevation
- Speed
- Heading
- Acceleration/Deceleration (negative acceleration)
- Brake System/Traction Control Status

Part II

- Vehicle path history and prediction
- Vehicle exterior light status
- Emergency Vehicle Status
- Trailer hauling descriptions
- Vehicle Classification Data
- Weather Probe Data

Challenges and Opportunities

- Requires ubiquitous RSU deployment
- Currently very small vehicle penetration
- Can support safety-of-life (low latency) applications

DATA SETS

Basic Vehicle Telematics

- **ECU** - Central Control Device that manages vehicle systems and functions
- **Inertial Measurement Unit** (IMU) - Measures the vehicle's movement by using accelerometers and gyroscopes to detect changes in acceleration and angular velocity

Typical Data Available	Other Functional Data
• Vehicle Position • Latitude • Longitude • Elevation • Speed • Heading • Acceleration/Deceleration Brake System/Traction Control Status	• Wiper Status • Headlight Status • Seat Belt Status • Air Bag Deployment Status

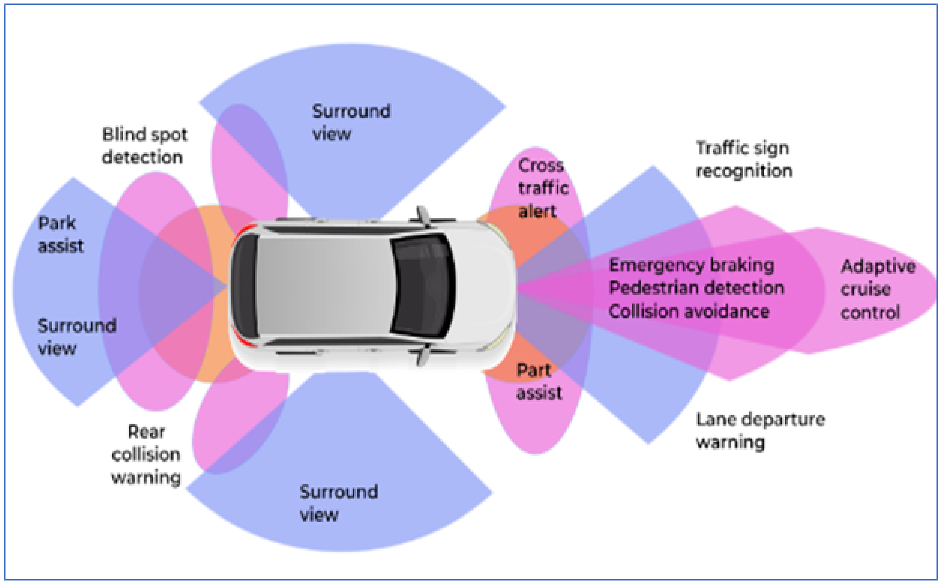
Challenges and Opportunities

- All new vehicles provide this data feed, but richness of data is manufacturer dependent
- Data transmission is dependent on supporting cellular connectivity
- Data frequency vary by manufacture and use case
- Near real-time data – data collection cost increases as latency decreases

DATA SETS

ADAS Sensor Suite

- Additional sensors provide data rich output
  - Emergency Braking
  - Traffic Sign Recognition
  - Lane Departure Warning



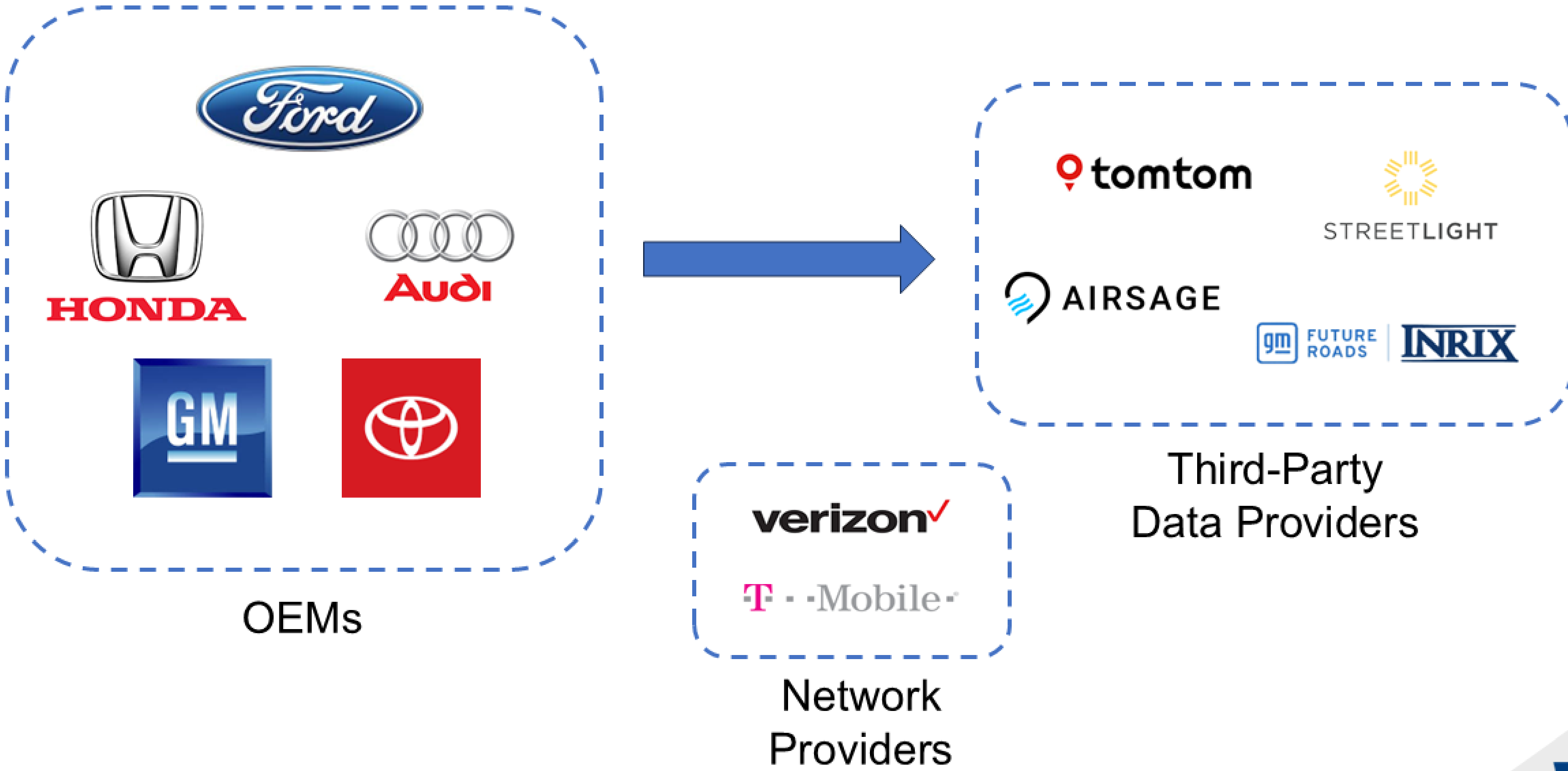
<u>Challenges and Opportunities</u>

- Data standardization and interoperability not there yet
- No known use cases for IOO traffic operations or safety applications
- Some promise for maintenance use cases

DATA CHARACTERISTICS

Characteristic	Description
Data Density	Essential for predicting traffic patterns or travel conditions. Requires significant vehicle data in the target area. Large data volumes needed for algorithm improvement.
Data Consistency	Must be normalized across different vehicle sources. Ensures reliable application use.
Data Latency	Low latency needed for real-time traffic management. Historical data can be older and still useful.
Data Frequency	High frequency increases costs but improves data resolution. Lower frequency sufficient for general mobility or historical analysis.
Data Richness	Detailed datasets valuable for real-time traffic and short-term delays. High-level datasets useful for origin and destination studies.
Data Integration	Significant effort required for integrating traffic datasets. Important to specify needed data based on use cases.
Data Cleaning	Ensures data validity and proper range. Corrects GPS signal drift and correlates position to the roadway.
Geofencing	Filters data to show areas of interest. Creates virtual boundaries for relevant data analysis. Triggers alerts for specific events within geofenced zones.

OUTREACH

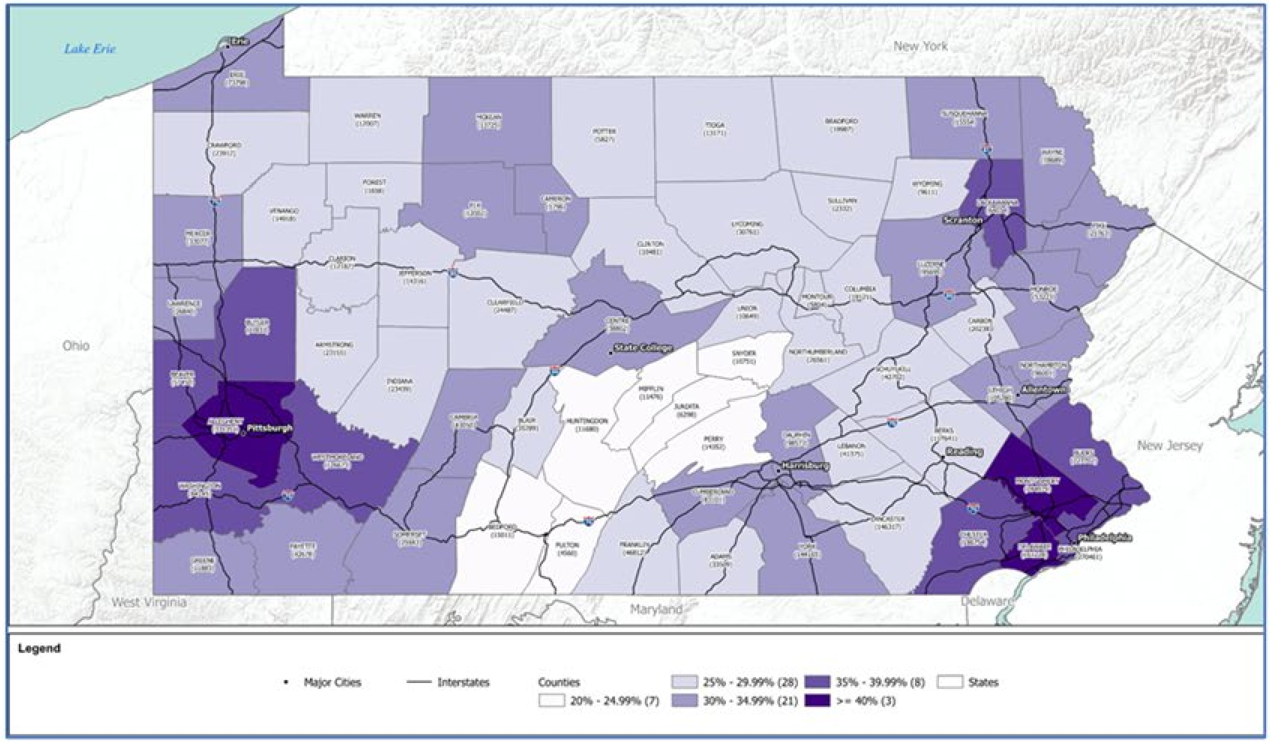

Ford
HONDA
Audi
GM
OEMs
tomtom
STREETLIGHT
AIRSAGE
FUTURE ROADS
INRIX
Third-Party Data Providers
verizon
T-Mobile
Network Providers

VEHICLE DATA VIABILITY

Pennsylvania Vehicle Makeup

- As a % of total registered vehicles

1. Ford – 6%
2. Toyota – 4%
3. Honda – 3%
4. Chevrolet – 3%
5. Jeep – 3%



STUDY FINDINGS

- Wealth of data is collected from vehicles and can apply to various use cases both near real-time and historical
- Most OEMs are reluctant to share data directly with DOTs, preferring instead to work through third-party providers
- The market is dominated by third-party data resellers who aggregate, clean, and analyze data from various sources
- Data privacy and security is a concern, including sharing data to insurance companies
- Litigation concerns with data that might be discoverable in a court case
- Data management and processing costs

STUDY FINDINGS

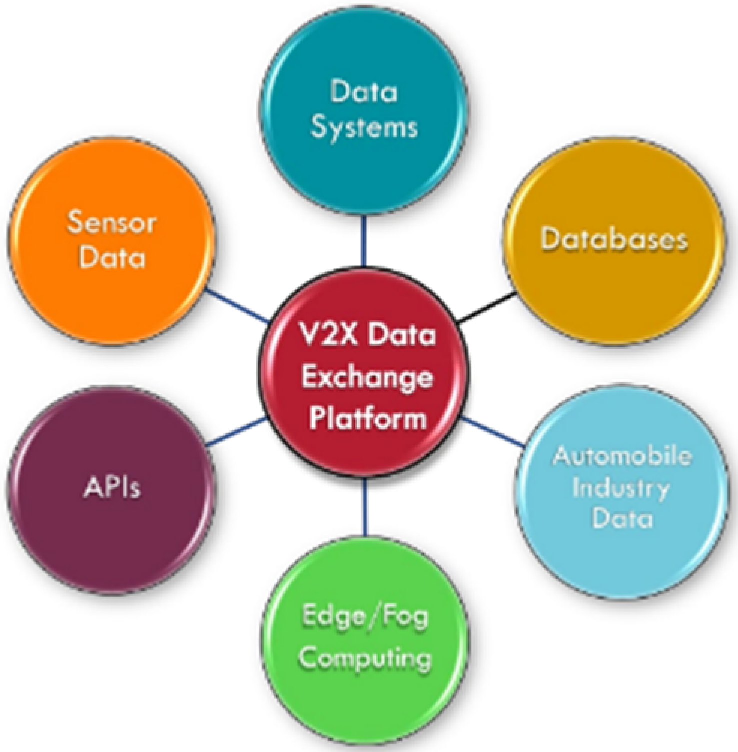
European Data Experience

- EU Data Act aims to democratize data access and foster innovation
- Vehicle manufacturers are required to share data
- Safeguards on IP
- Transparency on data use

Takeaways for US
Balancing Interests
Interoperability and Standardization
Legal and Regulatory Alignment
Transparency and Accountability
Innovation and Competitiveness

OPPORTUNITIES

- Prioritize relationships with third-party data providers for near-term data acquisition
- Conduct pilot use case through the V2X Data Exchange to evaluate the effectiveness of CV data
- Invest in robust data management infrastructure to handle large-scale CV data effectively
- Engage in collaborative efforts like TPFS to share costs and knowledge with other DOTs
- Explore grant opportunities
- National engagement



PFS FRAMEWORK

TPF-5(539) “Establishment of a Public-Private Transportation Data Exchange Center”

- Led by Missouri DOT
- Nine members (HI, CT, KS, MO, NJ, PA, TX, WI, CA)
- $1.6 M
- Goal – develop secure computing, data analytics and storage infrastructure with data repository that will collect relevant vehicle data and share data

CALL TO ACTION

How can TETC members help?

- Engage in the new Data Exchange PFS
- Look to develop use cases for CV data
- Engage in conversations on value

QUESTIONS?

Gunnar Rhone, PE
Engineering Specialist – Transformational Technology
Pennsylvania Department of Transportation

717.214.2997
grhone@pa.gov

Appendix B – Vendor Interview High-level Summary

In late 2024 and early 2025, we conducted interviews with a representative sample of major CV data vendors to gather their perspective on the state of the industry and identify any future trends to be aware of. Due to the proprietary nature of the shared information, we have chosen to present the interview notes in terms of general observations and trends. These observations are grouped by topical area below.

General Observations

- Vehicle OEMs are more reticent than ever to share source data with speed data aggregators.
- The ability for agencies to view and stream raw GPS location data could be transformational. However, getting access to this level of data could be difficult. Companies that share raw GPS data are subject to significant anonymization rules imposed by company lawyers (to reduce risk to the company) and by outside entities, notably the European Union (EU) General Data Protection Regulation (GDPR).
- Some data vendors provide their data on a subscription basis, while others offer it on a per-project basis.

Speed Data Sets

- Some vendors traditionally known for speed data are branching out to include other data products (like trip data). Others have no intention of releasing breadcrumb or waypoint data.

Vehicle Event Data

- Events can include a disabled vehicle, a slippery road, fog, a crash, heavy rain, and road construction.
- The industry is inconsistent regarding how data is priced and provided. Some vendors make event (e.g., safety) data available through a data analytics platform on a subscription basis. Others provide a static dataset file, but not the analytical tools needed to examine the data.
- Depending on the vendor, vehicle events can be captured at a very high frequency, often at or exceeding 15 Hz. However, data is not often sent to the cloud in real time, unless a crash is detected.
- It is common for event data to be cached during a trip and uploaded to the cloud after a vehicle's trip is completed. This batch uploading approach could limit the ability to do real-time monitoring of low-severity events that aren't uploaded immediately.
- Right now, the primary purpose of vendors collecting vehicle event data is to improve safety and provide incentives to reduce driver distractions. Any public sector use of this data or derivative products is currently secondary to its primary use.
- While vehicle event data is not ubiquitous across all vehicles in the US, the number of vehicles supplying such data is impressive. One data provider claims to receive event data from 3.5 million connected vehicles. Another claims to receive data from 140 million vehicle trips per day.
- OEMs' onboard equipment for recording event data is constantly improving. Some new vehicles can report data every second, while older vehicles report data every 15 seconds.

Trip Data

- It can be common for trip data providers to receive vehicle location data from their sources every 15 seconds. Some vendors cache one-second GPS data in the vehicle and dump it into the cloud every 15 seconds. Others wait until a trip is complete before uploading.
- Some trip data vendors receive their source GPS data from vehicle OEMs. Others receive it from third-party hardware (or even their own hardware) installed in fleet or commercial vehicles.

Volume Data

- The Coalition can help move volume products forward by making more continuous count stations available for calibration and training and ensuring timely data publication (every six months or less, for example).
- Common segmentation that is independent of map providers would be helpful.

Imagery Data

- Imagery data providers are beginning to enter the market. One provider has dashcam imagery from between 30,000 and 60,000 vehicles.
- Imagery data providers need the following (at a minimum) from dashcam partners:
  - Permission to share.
  - Reasonable pricing.
  - Real-time capability.
- This is an emerging marketplace—the number of current vendors is small, but could grow.

Appendix C – Agency Survey Results

This appendix contains the results of a survey that was issued to all member agencies of The Eastern Transportation Coalition. We received twenty (20) individual responses from a total of sixteen (16) agencies. **This section presents the results based on the twenty individual responses.** In the section entitled "Connected Vehicle Use Cases and Agency Feedback," we aggregated those twenty responses and reported them based on the sixteen agencies that responded. Because of this, the percentages shown in this appendix may not match those shown in the body of the report.

---

1. **Does your agency purchase, receive for free, or otherwise leverage any type of connected vehicle data?**

   (responses = 20)

Value	Percent	Responses
Not sure	10%	2
No	25%	5
Yes	65%	13
	Totals	**20**

---

2. **If your agency does have access to CV data, what type of data do you have access to? Please explain what type of data you get and the vendor that sells it to you.**

(responses = 16)

1.	Historic Wejo data.
2.	Safety data (hard braking, etc.) to analyze crash timelines.
3.	Speed data from HERE and INRIX as well as CV data from StreetLight.
4.	Probe vehicle data from INRIX, including their Signal Analytics Platform. The Signal Analytics Platform aggregates individual probe vehicle trajectory data into signal/corridor performance measures.
5.	We use HERE, Replica, and TomTom for many purposes, including safety and traffic operational analysis.
6.	INRIX/Replica purchased through the TETC Contract. Data includes latitude and longitude, speed, speed threshold, heading, origin/destination, date/time, weather, wiper status, hard braking, near collision, seat belt, and vehicle headway.
7.	Origin-destination data via INRIX.
8.	One-time purchase of Wejo waypoint and event data under the Coalition TDM. The Wejo data followed the standard Wejo schema for events and waypoints.
9.	INRIX/Drivewyze, Replica, StreetLight. Also reviewing SMATS. TMC is primarily focused on traffic operations data and is most interested in real-time or near-real-time data.
10.	TETC TDM - INRIX travel time and speed, volume, O-D, and dangerous slowdown alerts.
11.	Travel time and Speed (INRIX), O-D and Waypoint (INRIX), Freight (Quetica), and previously Volume (INRIX).
12.	RITIS/INRIX platform, StreetLight Insight (bike, ped, and vehicle traffic, corridor studies, etc.), WAZE user data.
13.	Vehicle probe data through RITIS and INRIX.
14.	StreetLight: O/D volume data; INRIX: speed data and signal analytics
15.	Probe vehicle data, vendor not mentioned.

---

3. **If you buy CV data today, how much do you pay for it? Type N/A if your agency is not purchasing data today.**

(responses = 14)

Responses redacted to protect confidential pricing information.

---

4. **Are you allowed to share the data you purchase with other agencies inside or outside of your state?**

(responses = 13)

Value	Percent	Responses
We are limited in sharing	54%	7
We can freely share the data with other people at our DOT, with partner agencies, etc.	46%	6
	Totals	**13**

We are limited in sharing, and here's how:

1.	Access can be requested by staff and consultants that work for our agency.
2.	Can share with in-state universities based on data use agreement.
3.	Data needs to be manipulated in some way before sharing. For example, we can add two links or two time periods.
4.	Partner agencies must sign an INRIX Data Use Agreement before having access to the Signal Analytics Platform.
5.	We share via the RITIS platform, so, anyone who has access to RITIS can access the data.
6.	In-state sharing only.

We can freely share the data with other people at our DOT, with partner agencies, etc.

1.	In-state MPOs and consultants who are working on our agency's projects.
2.	Only with those with which our agency has an MOU.

5. What does your agency use these data for?

(responses = 13)

1.	Currently exploring applications for safety analysis, speeding analysis, weather detection, and planning model O-D data.
2.	NPMRDS reporting requirements, traffic modeling, and project-based speed data requests.
3.	Performance monitoring of arterial signals and corridors.
4.	Safety and traffic analysis.
5.	We are testing application use. The main intention is using it for safety project justification.
6.	We use this for planning level analysis and for operational after-actions, or at times during a long-term event to improve detour and evacuation plans by examining where traffic actually traveled during an event.
7.	We conflated that data to our internal linear referencing system (LRS). We have a few pilot projects: Wejo weather data, safety analytics (crashes, speeding), various speed studies.
8.	Planning only, TMC determining how to best use alongside our existing data. We have thoroughly discussed integrating into our traffic operation systems, but have not found sufficient yet for near-real-time.
9.	Dangerous Slow-down Alerts, TMC real-time info, emergency operations knowledge (e.g., during Eclipse), work zone planning and hopefully in the future alerts, corridor planning, other planning.
10.	Planning studies, after-action reviews, detour/work zone planning, performance management and reporting, road diet planning, incident response, congestion management, travel demand modeling, public communications/press inquiries, etc.
11.	Grants, traffic studies, corridor studies, prioritizing investments
12.	For traffic planning purposes, incident after-action reviews, holiday planning, and understanding recurring and non-recurring congestion.
13.	Travel time

6. Are you limited on what you can do with the data? Any usage restrictions?

(responses = 13)

Value	Percent	Responses
No restrictions.	69%	9
Yes, there are restrictions on what we can do with the data, and here's what those restrictions are:	31%	4
	Totals	13

Yes, there are restrictions on what we can do with the data, and here's what those restrictions are:

1.	DUA, but this is minimally restrictive.
2.	Data must be manipulated before being provided to others.
3.	We cannot publish unaggregated data.

7. What types of data would you like to receive from CV data providers?

(responses = 13)

Value	Percent	Responses
The location of rapid deceleration events (a.k.a. heavy braking).	85%	17
The location of crashes detected from airbag deployments or other sensors.	85%	17
Waypoints / breadcrumb trails of trips.	75%	15
The location of debris on the roadway.	75%	15
Data about road weather conditions (snow or rain,)	70%	14
The location of lane departure events.	60%	12
The location of fog. or similar low visibility.	60%	12
The location of sharp turning events.	55%	11
The location of rapid acceleration events.	50%	10
The number of passengers in vehicles.	50%	10
The location of cones or barrels.	50%	10
Live or recent images/video from dashcams or other on-vehicle cameras.	40%	8
Other - Please explain	35%	7
	Totals	**13**

Other – Please explain:

1.	BSM like data (similar to breadcrumbs)
2.	Location of hard braking events
3.	Near miss for network screening
4.	Road and TCD conditions
5.	We would need the data to be on more roads than just limited access highways. Would want the Federal Aid System.
6.	Dashcam images to supply the location of active work zones, as evidenced by video of the cones, barrels, workers, and construction vehicles; disabled and abandoned vehicles along the road, as evidenced by the dashcam images.
7.	Friction data is an interesting one. Seat belt and distracted events are another.

8. What types of data would you like to receive from CV data providers?

(responses = 19)

1.	Planning and incident response
2.	Safety surrogate measures, road weather information, real time asset condition and incident information
3.	This could be useful in determining real-time conditions on our roadways, allowing for faster response.
4.	We would use this data for safety purposes. It would be invaluable in learning more about potential safety hazards.
5.	O-D planning, incident notification, occupancy,
6.	Signal/corridor performance measures can be derived from the breadcrumb trails of trips (as used in our current contract with INRIX). The rapid deceleration events can be used to determine near-miss events, a key safety data. Airbag deployment data can be used to identify crashes
7.	Safety screening and justification for projects.
8.	This data would provide an additional layer of safety-related data to locate and prioritize safety improvement projects based around specific driver behaviors.
9.	Waypoints/breadcrumbs would continue to be used for the above, however, the more real-time we can get it the better we can adjust signal timing or make other detour adjustments. Location of crash detections would allow for another detection tool to locate incidents in our area of responsibility. Road weather conditions would give us mobile road weather sensors (in a sense) without needing to deploy a lot of devices on our own fleet. Location of debris would also serve as a detection tool. I do worry about "oversaturation" of data into live-operations despite the benefits.
10.	Track crashes if they were to happen in a CAV. Understand the growing pains of CAVs. Use data to create better tailored policies and laws to help CAVs operate safely in our state. Understand CAV traffic flows in our state. Create policies that promote the safe use of CAVs and use laws to mitigate the adverse impacts of CAVs if present.
11.	Incident detection and mitigation, origin-destination, deploying maintenance resources to clear hazards, track active work zones and confirm allowable work zone hours.
12.	Our state likely has less on-road ITS deployed than states with heavier traffic, so the data would be used to supplement reports from state police and other sources. A specific use is getting messages up on changeable message boards to alert or detour travelers.
13.	Safety countermeasures, weather related events etc.
14.	Safety planning, travel demand modeling, modal planning, safety target setting, "zero vision" type activities.

15.	Work-zone crash forensics, saefty measures
16.	It is mainly used to inform safety initiatives, traffic planning, and projects, as well as live incident detection.
17.	Dashcam images to identify and provide info about crashes, work zones in the TMC coverage area.
18.	Event Detection
19.	If I could get BSM like data at low latency, incident detection, safety analysis (offline), travel time, OD, combine with ATSPM for system volumes, and system efficiency evaluation

9. What types of data would you like to receive from CV data providers?

(responses = 11)

Value	Percent	Responses
Lack of funds to purchase	80%	16
Lack of funds to integrate with my ATMS or other systems	40%	8
Don't see a big enough benefit or return on investment	25%	5
Concerns over privacy	30%	6
Don't trust the data	15%	3
Worry we'll get locked into a single vendor.	10%	2
Worry the vendor will go out of business or the data will disappear (market instability)	35%	7
Not sure which vendor to choose	20%	4
Inability to work with large and complex datasets.	25%	5
Don't know where to begin or what's available.	25%	5
Other - Please explain	15%	3
	Totals	**11**

Other – Please explain:

1.	Overcleaning from vendor
2.	Uncertain B/C - need to define business case
3.	Worry about over-saturation of data into operations

---

10. If there were significant cost savings (or other benefits) resulting from a multi-state or multi-agency procurement of CV data by the Eastern Transportation Coalition, would you be interested in participating in the procurement?

(responses = 19)

Value	Percent	Responses
Yes	79%	15
Other - tell us what you're thinking	21%	4
	Totals	**19**

Other – tell us what you're thinking:

1.	Depends on cost, coverage, etc.
2.	Depends on my funding situation
3.	My answer would be yes, my agency likely wouldn't support the expenditure as exhibited in the past.
4.	Possible - would depend on price point

Appendix D – Example of Data Available from Connected Vehicles

The following list[13] contains examples of the wide variety of onboard vehicle data currently being collected by OEMs. While this list is not exhaustive (and can vary between automakers), it offers valuable insights into potential capabilities and can serve as a starting point for brainstorming new use cases.

Data Category	Specific data elements (examples)
ADAS: Advanced driver assistance system	• State of the automated parking brake • Blind spot warning system active • Blind spot warning system turned on • Lane keeping assist system actively engaged • Lane keeping assist system turned on • Parking assist status. Has the driver muted warnings? • Rear cross-traffic system status • Driver assistance system status • Object detected in side-fired ultrasonics
Charging	• Battery capacity • Energy content of the battery • Battery level percentage • Battery charge limit • Charge rate • Charger voltage • Estimated range • Time until charging is complete
Climate	• State of the vehicle's climate control system • Outside air temperature
Crash	• Automatic emergency call enabled state • Crash incident • Vehicle tipped state • Crash status
Cruise Control	• Set target speed
Dashboard Lights	• Warning lights, such as: ○ Low tire pressure ○ Antilock brake system failure ○ Low engine oil ○ Charging system fault ○ Engine coolant temperature

[13] High Mobility's APIs illustrating the range of telemetry and event data potentially available (https://www.high-mobility.com/car-data)

	• Status indication lights, such as: ○ Doors open ○ Low fuel ○ Seat belts unbuckled
Diagnostics	• Battery level • Battery voltage • Engine coolant temperature • Engine RPM • Timer for next vehicle service • Airbag status • Tire pressure
Doors	• Door lock status • Door position
Engine	• Engine on/off status
Fueling	• Gas door position
Honk Horn – Flash Light	• Hazard light status
Hood	• Hood lock status
Ignition	• Ignition state
Lights	• Parking light status • Fog light status • Interior light status
Maintenance	• Brake servicing status • Remaining distance to next oil service
Offroad	• Route elevation incline
Parking Brake	• Parking brake status
Race	• Acceleration • Electronic stability program • Selected gear
Rooftop Control	• Sunroof position • Sunroof tilt • Convertible roof state
Seats	• Seatbelt state
Theft Alarm	• Tow protection sensor triggered • Interior protection sensor triggered
Trips	• Trip distance • Start and end coordinates of the trip • Harsh braking • Harsh acceleration • Maximum trip speed • Fuel consumption during the trip

Trunk	• Trunk lock status • Trunk lid position
Usage	• Average fuel consumption in the last trip • Current fuel consumption • Average speed since last reset • Average weekly distance • Eco score rating • Distance traveled with electricity in the last trip • Duration of last trip
Vehicle Location	• Altitude • Coordinates • GPS signal strength • GPS source type • Heading angle
Windows	• Window position

Appendix E – Limited Access Road Network Mileage and Volume Statistics

When it comes to piloting new use cases for CV data, it makes sense to focus these efforts on the limited access road network. Here are some of the most prominent reasons we recommend focusing initial efforts on limited access highways:

- **Scale:** Well under 2% of the nation's road mileage accounts for 35% of the nation's total vehicle miles travelled. Thus, limited access roads have both the highest underlying volume, yielding the most CV data, and the highest benefits accrue to the highest density of travelers.
- **Operating characteristics:** These facilities are designed so that, under normal operating conditions, vehicles flow smoothly and don't stop. Thus, cases where speeds drop suddenly are indicative of anomalous conditions.
- **Favorable geometric design:** These facilities tend to have the widest lane widths and significant additional cross-sectional right-of-way. They are often surrounded by low-density development or no development at all. Further, they are grade-separated with clear delineation in directional flow (save for reversible lanes). This is by far the best road environment for geolocation, precision, lane-by-lane differentiation potential, and anomaly detection/filtering of CV data.
- **Least possible potential for privacy concerns:** These facilities have no homes, driveways, dwelling addresses, or trip origins or destinations. Thus, there are minimal privacy concerns with CV data used strictly on a limited access highway being used to impute a user's complete trip, or to uncover a driver's identity.

Table 3: Limited access highways carry a significant portion of a state's VMT, even though they represent a small fraction of a state's roadway mileage

State	Total Limited Access Network Miles	% of Total Miles	% of Total VMT
Alabama	1,037	1.1%	23.3%
Connecticut	625	2.9%	48.8%
Delaware	109	1.7%	26.1%
District of Columbia	28	1.9%	25.0%
Florida	2,273	1.8%	27.9%
Georgia	1,430	1.1%	29.5%
Kentucky	1,406	1.8%	37.1%
Maine	385	1.7%	24.2%
Maryland	858	2.6%	42.0%
Massachusetts	903	2.5%	38.4%
New Hampshire	308	1.9%	34.6%
New Jersey	920	2.4%	38.9%
New York	2,737	2.4%	37.3%
North Carolina	2,250	2.1%	29.9%
Pennsylvania	2,794	2.3%	35.0%
Rhode Island	162	2.5%	43.5%
South Carolina	979	1.3%	30.8%
Tennessee	1,407	1.5%	33.4%
Vermont	336	2.4%	25.2%
Virginia	1,577	2.1%	39.0%
West Virginia	571	1.5%	27.8%
TETC Total	**23,095**	**1.8%**	**33.1%**
U.S. Total	**67,409**	**1.6%**	**35.2%**

Source: FHWA Highway Statistics, 2022, Published January 2024 (https://www.fhwa.dot.gov/policyinformation/statistics/2022/)