Connected and Automated Vehicle Data Issues and Opportunities

Texas CAV Task Force White Paper Subcommittee on Data, Connectivity, Cybersecurity, and Privacy

Authors:
Data, Connectivity, Cybersecurity, and Privacy Subcommittee of the
Texas Connected and Autonomous Vehicles Task Force
James Hubbard, Texas A&M Transportation Institute

August 2023

This Page Intentionally Left Blank

Table of Contents

List of Figures

Acknowledgments

The Texas Connected and Autonomous Vehicles (CAV) Task Force would like to acknowledge and thank all its voting and participating membership and the members of this subcommittee for their hard work and many hours dedicated to developing this white paper. The task force would especially like to thank the contributing authors of this paper for taking the thoughts of so many and combining them into one well-written document. In addition, special thanks go to Beverly Kuhn, Ed Seymour, and Robert Brydia of the Texas A&M Transportation Institute for their management and creativity in assisting the CAV Task Force Texas Department of Transportation team and subcommittee chairs. Finally, the CAV Task Force would like to thank Texas Governor Greg Abbott and his staff for their guidance and vision in creating this task force. Texas is better prepared for CAVs due to their leadership.

Disclaimer

The contents of this white paper reflect the views of the Texas CAV Task Force members, who are responsible for the information presented herein. The contents do not necessarily reflect the official views or policies of the State of Texas or any Texas state agencies. The white paper does not constitute a standard, specification, or regulation, nor does it endorse standards, specifications, or regulations. This white paper does not endorse practices, products, or procedures from any private-sector entity and is presented as a consensus broad opinion document for supporting and enhancing the CAV ecosystem within Texas.

Texas CAV Task Force Charter

The Texas CAV Task Force was created at the request of Texas Governor Greg Abbott in January 2019. The task force is responsible for preparing Texas for the safe and efficient rollout of CAVs on all forms of transportation infrastructure.

The primary functions are:

1. Coordinating and providing information on CAV technology use and testing in Texas.
2. Informing the public and leaders on current and future CAV advancements and what they mean in Texas. This process includes reporting on the current status, future concerns, and how these technologies are changing future quality of life and well-being.
3. Making Texas a leader in understanding how to best prepare and wisely integrate CAV technologies in a positive, safe way, as well as promoting positive development and experiences for the state.

The CAV Task Force is composed of a voting group of no more than 25 members and represents the full spectrum of CAV stakeholders.

Terminology Note

The Texas CAV Task Force addresses the full spectrum of connected, automated, and autonomous vehicles. An *automated vehicle* refers to a vehicle that may perform a subset of driving tasks and requires a driver to perform the remainder of the driving tasks and supervise each feature's performance while engaged. The performance capabilities of automated and autonomous vehicles consist of levels 0–5 with level 0 having no driving automation and level 5 having full automation,

with automation increasing at each progressive level. A fully autonomous vehicle can perform all driving tasks on a sustained basis without the need for a driver to intervene.

These definitions are still blurred in common discussions and language. Currently, the industry is developing automated vehicle capability while pursuing fully autonomous vehicles. The white papers generally use the term *autonomous* to refer to vehicles with fully autonomous capabilities and the term *CAV* to refer to the grouping of connected, automated, and autonomous vehicles. Please see the 2021 terminology white paper for a full listing of terms and definitions used in this developing technology ecosystem.

List of Terms and Acronyms

CAV	connected and autonomous vehicle; also, connected and automated vehicle
FDOT	Florida Department of Transportation
IOO	infrastructure owner/operator
JSON	JavaScript Object Notation
OEM	original equipment manufacturer
SME	subject matter expert
TxDOT	Texas Department of Transportation
UDF	user-defined function
V2X	vehicle to everything
WZDx	Work Zone Data Exchange

This Page Intentionally Left Blank

Executive Summary

The connected vehicle market is growing at a fast pace, and cars are becoming more connected than ever before. This trend is expected to continue; a Counterpoint Connected Car study predicts that more than 70 percent of the cars sold will be connected by the year 2025. The data generated by the vehicle typically includes information about the vehicle's status and driver's behavior, as well as location-based data. With the continued emergence of advanced driving assistance systems (ADAS) features, vehicles that are connected back to the original equipment manufacturers (OEMs) become more relevant as a critical data source and insight into the traffic stream. Many newer vehicles also have an electric vehicle component, from which additional data may be available. Overall, the interest in connected and autonomous vehicle data comes from OEMs, suppliers, insurers, mobility providers, infrastructure owners/operators (IOOs), fleet owners, and others.

The various applications for vehicle data include but are not limited to the following:

Predictive maintenance,
Fleet management,
Roadside assistance,
In-car payments,
Usage-based Insurance, and
Traffic management.

In particular, IOOs can analyze the data in a multitude of ways to gain insight into roadway conditions. For example, hard-braking events are being examined in conjunction with other roadway characteristics to highlight areas where roadway improvements may need to be prioritized. As another example, video from dashcams or forward-looking cameras can be analyzed to detect missing roadside features (i.e., the feature was there on the previous video but is not there now), vegetation encroaching on roadway signs, striping and pavement conditions, and numerous other use cases.

Both raw and processed vehicle and roadside data add value to all parties involved. The sharing of these data among the multiple interest groups provides widespread opportunity to examine and improve work zones, provide insight into crashes, examine operational conditions, and more. This data sharing can be accomplished via a data exchange—essentially a data portal where participants can both send and receive data.

Data exchanges can also play an important role in piecing together disparate data. Combining these data sources will enable users to analyze road conditions in real time and communicate important travel information to the traveling public, state/local government entities, private-sector partners, and other stakeholders. In addition to real-time analysis, exchanges could also support analysis of long-term historic data, enabling data-driven infrastructure investments and various research initiatives. Public agencies and/or private companies will likely need to see a business case value to participate in a data exchange community. This value may increase as the number of partners grows and the types and amount of data exchanged become more substantial.

Before this is possible, issues such as the following must be addressed and clarified to the satisfaction of all participants:

Standards,

Privacy,
Governance,
Security, and
Use cases.

Tackling these concerns will help Texas continue to be an innovation leader in this emerging sector and benefit all Texans.

Introduction

The data economy is growing at a meteoric rate. The world has never been so comprehensively recorded. More data are produced every year, and the conclusions that can be learned by examining these data also increase, from geographical positioning systems that guide us to our destination to smartwatches that monitor our heart rate. Digital devices all produce data. These data can be used to track various aspects of not only the device but also its surroundings.

Companies collecting these data can extract enormous amounts of value from it. However, processing and comprehending such a large volume of data can be a monumental task. The big data revolution has sparked renewed interest and innovation in areas such as artificial intelligence and machine learning. These data can help guide the creation of new products and services or predict current customers' preferences.

Data exchanges are developing as a key component of the data economy. By connecting data suppliers and consumers through an intuitive experience, data exchanges have created a modern-day gold rush to help modern digital organizations address their needs for data. Furthermore, by eliminating the friction from finding, acquiring, and integrating data, these exchanges make it easier to monetize data assets and create new revenue streams. Public agencies and/or private companies will likely need to see a business case value to participate in a data exchange community. However, one business use case that has become readily apparent is the two-way exchange of data. For example, an infrastructure owner/operator (IOO) may provide information about the location of work zones, and a connected vehicle might provide data from traveling through the work zone, which can be examined to ensure that roadway infrastructure is being set up and managed in a way to enhance machine understanding of the environment. The value to an IOO may increase as the number of partners grows and the types and amount of data exchanged become more substantial.

What Is a Data Exchange?

At the simplest mechanism, a data exchange is simply a methodology to exchange data between a source and a receiver. Data exchanges may have different levels of complexity, requirements, and policies.

Type of Data Exchanges

Some of the many types of data exchanges are briefly described as follows:

- **Peer-to-peer data exchange**—A peer-to-peer exchange enables data sharing directly between two companies or departments inside the same company that want to share data on an ad-hoc basis. These types of exchanges consist of a small or large network with each user operating as a node. Peer-to-peer networks include a user platform that allows individual participants to exchange content and information. For example, a use case for a peer-to-peer database would be a large university with two data warehouses, one for business operations and one for research, which might use a peer-to-peer exchange to share data back and forth.

- **Private data exchange**—A private data exchange can come in many shapes and sizes. For example, it could be a consortium, which collects and standardizes insurance data and distributes it to participating members and regulators. Also, many large companies use a private data exchange to share inventory data with their suppliers and collect shipment data in return. Alternatively, a service provider, such as a marketing firm, might use a private data exchange to share account activity with individual customers. Some providers provide platforms specifically to facilitate private data exchanges.
- **Data marketplace**—A data marketplace is a public data exchange open to any company that wants to supply or consume data. This platform allows users to buy or sell different data sets from various sources. These marketplaces usually use cloud services where companies and/or individuals can upload data (*1*).
- **Public/cooperative exchange**—A public/cooperative exchange provides data for mutual benefit such as the Work Zone Data Exchange (WZDx) and General Transit Feed Specification, which was developed by the Federal Highway Administration and is increasingly being supported by agencies across the United States.
- **Public/private partnership data exchanges**—A partnership where government and private companies work together to secure and share data for people, the public good and commercial operations. Data Privacy, Data Security, Data Trust and Data Sovereignty are foundational to these Data Exchanges. These data exchanges will allow secure, private transaction sharing of valuable data assets with trusted and approved partners and may become a new economic platform for the communities they serve. One of the most successful early examples of this type of exchange is the Human Genome Project. Another example of the variation in how these partnerships may exist is the Texas Strategic Mapping Program.
-

Figure 1 shows an example of a simple data exchange incentive model that connects data providers and data consumers.

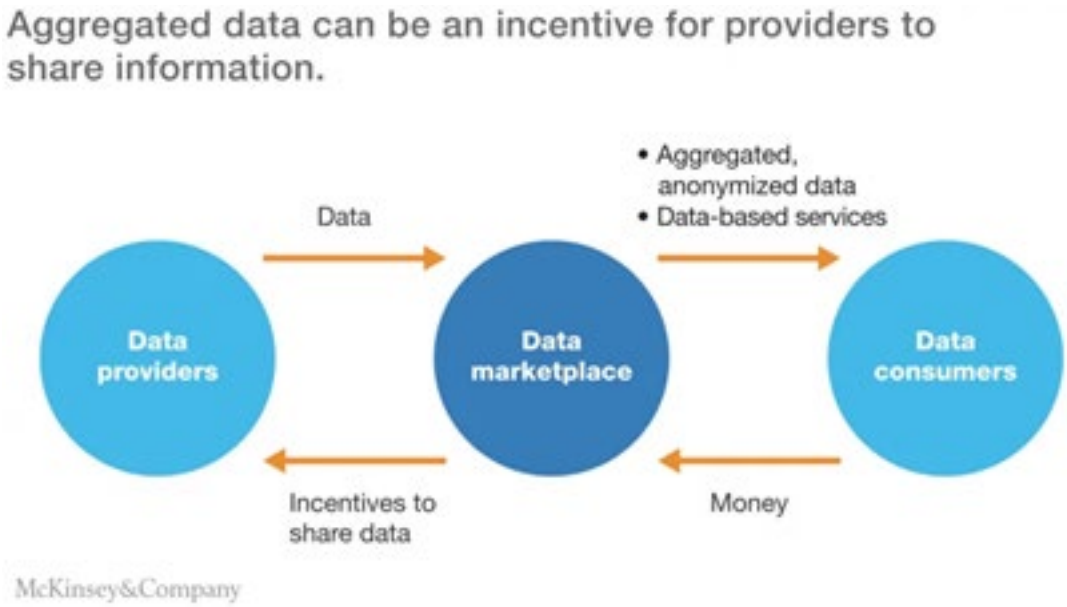

Figure 1: Data Marketplace

Baseline Functionality

For a data exchange to benefit all parties involved, it must enable data suppliers and consumers to share data securely and efficiently. As data exchanges develop and proliferate, the definitions, functionality, and overall usefulness will continue to be refined. For this to happen, the exchange must offer a baseline functionality that makes it easy for suppliers to publish data assets and for data consumers to find them. A data exchange will typically have the following baseline functions (*1*):

- **Searchable catalog**—Data consumers should be able to search a catalog of data assets in Structured Query Language (SQL) query or searchable tags. Each entry in the catalog contains a description of the data set and relevant details, such as the number of records, file type, statistics, and ratings.
- **Asset management**—A data exchange must make it easy for suppliers to upload, describe, manage, publish, and update data assets. It also allows suppliers to define licensing, access rights, and other terms and conditions and to manage their inventory.
- **Access control**—An exchange allows data suppliers to control who can access or purchase a data asset, in part or whole, and restrict access until an agreement or payment is complete. The four primary examples of data access control are:
    - **Discretionary access control**—The data owner grants access.
    - **Mandatory access control**—People are permitted access based on information clearance designed using a nondiscretionary method.
    - **Role-based access control**—Access is given to users based on a role assigned to them.
    - **Attribute-based access control**—Each resource and user receives attributes. This approach judges resource access by comparing the user's features, such as time of day, position, and location.

- **Data transfer**—A data exchange should support one or more ways for data suppliers to transfer data to consumers, including file transfer, application programming interface, and multi-tenant data sharing. Cloud-based file transfer uses object storage to hold files, simplifying access.
- **Transaction management**—Data exchanges should facilitate monetary transactions and offer payment terms, including bank transfers, credit cards, and account billing. Data consumers track their history of data purchases and subscriptions, see renewal terms, and cancel subscriptions. Likewise, data suppliers see a data set history of consumer transactions and activity.
- **Account management**—A data exchange lets consumers and suppliers create and manage accounts that contain information about users, administrators, authorized buyers and sellers, billing information, account activity, and payment mechanisms. Large organizations with multiple parties interested in acquiring external data can easily manage permissions and expenditures.
- **Authentication**—Data exchanges should have a way to authenticate that the supplier is sharing data it is entitled to share.

Currently, the industry has vast differences in the features and capabilities that are supported in data exchanges. For example, some of the developing cooperative exchanges between agencies may not have commercial features such as transaction and account management.

Value-Added Features

The following features and functions represent the evolution of the data exchange caused by competition. Most data exchange platform providers will soon require these features to do business (*1*).

- **Data enrichment services**—These data quality services include deduplication (removing duplicate items), address correction, standardization, data cleansing, file merging, and validation.
- **Selective sharing**—Some data exchanges let data suppliers selectively share data by filtering data for select consumers or by selecting consumers with whom to share data.
- **Data mapping**—Certain data exchanges analyze consumer data and then recommend supplier data within the exchange that best supplement or enrich those data.
- **Connector software developer kit**—Data exchanges can offer a software development kit that allows data exchange operators or suppliers to create a custom connector to less common data platforms.
- **Derived aggregate data**—Instead of giving consumers access to raw data that may contain sensitive data, data suppliers can instead offer user-defined functions (UDFs), such as analysis routines. Consumers run the UDFs against a described data set and receive the aggregated analytical output.
- **Enhanced onboarding**—This includes offering workflows and wizards to simplify the onboarding process for data suppliers. The workflows evaluate supplier data, assess compliance, and manage data ingestion.
- **Alerts**—Some data exchanges alert consumers when a newly published data set matches their interests. Some also alert data suppliers when potential customers join the exchange.

- **Enhanced reporting**—Data suppliers want a clear picture of their sales performance in the exchange. These reports can include top buyers to allow the suppliers to focus their efforts on the right consumers.
- **Custom data products**—Some data exchanges enable providers to create a data product by blending, segmenting, or engineering data to suit use cases and then publish that product to the exchange.

Data Governance

The Data Governance Institute defines *data governance* as "a system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods" (*2*). The Data Management Association sees data management as a wheel, with data governance as the hub from which the following 10 data management knowledge areas come from (*2*):

- **Data architecture**—the overall structure of data and data-related resources as an integral part of the enterprise architecture.
- **Data modeling and design**—analysis, design, building, testing, and maintenance.
- **Data storage and operations**—structured physical data assets storage, deployment, and management.
- **Data security**—ensuring privacy, confidentiality, and appropriate access.
- **Data integration and interoperability**—acquisition, extraction, transformation, movement, delivery, replication, federation, virtualization, and operational support.
- **Documents and content**—storing, protecting, indexing, and enabling access to data found in unstructured sources and making these data available for integration and interoperability with structured data.
- **Reference and master data**—managing shared data to reduce redundancy and ensure better data quality through standardized definition and use of data values.
- **Data warehousing and business intelligence**—managing analytical data processing and enabling access to decision support data for reporting and analysis.
- **Metadata**—collecting, categorizing, maintaining, integrating, controlling, managing, and delivering metadata (data that provide information about other data).
- **Data quality**—defining, monitoring, maintaining data integrity, and improving data quality.

Data governance has three roles:

- Steering committee,
- Data owner, and
- Data steward.

Steering Committee

Steering committees often consist of senior management, C-level individuals (high-ranking executives in charge of various departments in a company, such as a chief financial officer), or individuals accountable for lines of business. Steering committee members' responsibilities include

setting the overall governance strategy with specific outcomes, championing the work of data stewards, and holding the governance organization accountable for timelines and outcomes.

Data Owner

Data owners are individuals responsible for ensuring that information within a specific data domain is governed across systems and lines of business. Data owners are generally members of the steering committee though they may not be voting members. Data owners are responsible for the following (*2*):

- Approving data glossaries and other data definitions;
- Ensuring the accuracy of information across the enterprise;
- Directing data quality activities;
- Reviewing and approving master data management approaches, outcomes, and activities;
- Working with other data owners to resolve data issues;
- Performing second-level review on matters identified by data stewards; and
- Providing the steering committee with input on software solutions, policies, or regulatory requirements of their data domain.

Data Steward

Data stewards are accountable for the day-to-day management of data. Templar says data stewards are subject matter experts (SMEs) who understand and communicate the meaning and use of information. Data stewards work with other data stewards across the organization as the governing body for most data decisions. Data stewards are responsible for the following:

- Being SMEs for their data domain,
- Identifying data issues and working with other data stewards to resolve them,
- Acting as a member of the data steward council,
- Proposing, discussing, and voting on data policies and committee activities,
- Reporting to the data owner and other stakeholders within a data domain, and
- Working cross-functionally across lines of business to ensure their domain's data are managed and understood.

Challenges and Issues

Anyone who routinely consumes data has had a bad experience with third-party data. For a data exchange to be trusted, it must ensure it provides data integrity, quality, and consistency. Unless data consumers trust the data housed on a data exchange, they will not use it. Data exchange operator can do numerous things to ensure consumers trust the data, such as:

- Curation—While curation is challenging to scale, data exchange operators can certify and curate supplier data sets to ensure quality and consistency.
- Profiling—A data exchange should contain rich metadata about each data asset, including the number of rows, names and types of fields, cardinality, and other statistics, as well as the lineage of the data set, including source system collection methods and age.

- **Sampling**—Ideally, data exchanges allow consumers to query or sample data sets to validate data quality.
- **Ratings**—Much in the way e-commerce platforms provide ratings of products, data exchanges can allow consumers to give feedback on the quality and utility of data sets.

Data exchanges must reach a critical mass of data suppliers and consumers to gain traction and become viable, long-term operations. However, this is a chicken-and-egg situation: data consumers will not use a site that lacks sufficient breadth and depth of data assets, while data suppliers will not populate a site that lacks potential users. Unless a significant cloud platform runs a data exchange, it must target a specific market with built-in and demonstrated third-party data needs. Government agencies, industry consortiums, and membership organizations have a distinct advantage because they have a built-in audience and already serve as liaisons and intermediaries. However, even public-sector or cooperative data exchanges such as those referenced earlier need to embrace good data governance and cybersecurity principles to protect the data and potential users.

Although the primary function of a data exchange is to facilitate the transfer of data between suppliers and consumers, that does not add business value for data consumers. Data exchanges need to integrate third-party data with their existing data management systems so business users can access and query the data and all other internal data. To facilitate integration, data exchanges need to validate supplier data. Exchanges also need to ensure the data are clean, consistent, and delivered as described. That means the schema does not alter mid-stream, a column field does not change its name or type, and the number of records remains consistent. Exchanges also need to ensure that the data supplier has documented all the requisite metadata so data consumers can make informed purchasing decisions.

Issues surrounding the liability associated with data collection, aggregation, use, storage, and exchange are also important items to address. Data could be shared with the wrong people, exposing industry-protected information. Personal information could be captured and used to identify particular people or trips. The potential for tort liability claims related to having awareness of an infrastructure concern but not yet repairing it might increase. These situations and many more like them in the overall area of liability are complex and evolving.

Industry Interviews

The Data, Connectivity, Cybersecurity, and Privacy Subcommittee of the Texas Connected and Autonomous Vehicles (CAV) Task Force conducted interviews with companies in the CAV industry to better understand their view of data exchange. A wide range of topics were covered, including the meaning of the phrase *data exchange*, potential hurdles for adoption, privacy concerns, security, management, specification, and use cases. The following is the interview guide used while conducting the interviews. Not all questions were asked in each interview. Instead, these questions were used as a guide to make sure interviewers captured information on vital topics.

Interview Guide

Data Exchange

1. What does a data exchange mean to your organization?
2. What types of data does your organization not have but would like to have?

3. What types of data are most important to your organization?
4. What types of data would your organization be able to contribute?
5. Are there any use cases that would benefit your organization that you are missing data for? (Specifically, what are you missing in this use case?)
6. What would be the biggest hurdle your organization faces in participating in a data exchange?
7. How would your organization benefit most from a data exchange?
8. What do you think would be the ideal uses cases for a data exchange (public and private)?

Privacy

1. Would there be privacy issues with your organization participating in a data exchange for these use cases?
    a. Nondisclosure agreements
    b. Personally identifiable information
2. What kind of data would you be willing to share in these use cases?
3. What mechanisms would your agency need to have in place to ensure that proprietary data stay non-public?

Security and Management

1. Please provide your opinions on who should own/manage/operate a data exchange.
2. What types of safeguards would your organization require to participate in a data exchange?
    a. Access controls
    b. Use notifications/approvals
    c. Cloud configuration
    d. Encryption
    e. Disaster recovery response
3. Please provide your opinion on any data exchange best practices you have encountered.
    a. Safeguards
    b. Management
    c. Standardized formats
    d. Repeatable processes

Specifications

1. Does your organization adhere to any published data standards, or do you use your own specifications?
2. Would your organization require data it receives through a data exchange to comply with a data standard?

Panasonic

Panasonic's smart mobility division's goal is to develop intelligent mobility solutions. The division uses a combination of software, hardware, and advanced analytics to provide value to customers in every part of the mobility ecosystem.

Panasonic views a data exchange as a way to collaborate with other entities. The company is currently working on a vehicle-to-everything (V2X) platform called Cirrus. This platform provides an open development space for sharing, collaborating, defining, and standardizing smart mobility data.

In addition, Panasonic is working with departments of transportation and commercial transit vehicle fleets to scale connected vehicle deployment from concept to implementation, operations, and maintenance across states and municipalities. The company currently has several hundred participating vehicles but is looking to scale up soon.

Panasonic is most interested in how to achieve scale, techniques to merge multiple data sets, and development of standard ways for managing data. Panasonic's ideal use cases are intersections, safety, mobility operations, and maintenance. The company is currently developing connected vehicle data specifications and hoping to work with the data community to evolve them. The work the U.S. Department of Transportation has done with the WZDx specification is an example that it is taking lessons learned. In addition, Panasonic is currently developing a data exchange best practice document. It is still a work in progress, but one of the most significant issues encountered is the need for data dictionaries. Without a data dictionary, there is a high chance that people will assume the wrong unit of measurement.

Locomation

Locomation is working on autonomous relay convoys (i.e., platooning). Although the company's end goal is full automation, it has been working on lateral and longitudinal acceleration.

Locomation's long-term goal for data exchanges is to integrate them with its system to be a data source for convoys. A data exchange is not part of the company's initial product, but future generations could include data exchanges. The data exchange must be able to confirm data availability and data freshness. Being able to rely on the data is paramount for integration. Resources and penetration rate are other hurdles to potentially participating in data exchanges. While current resources are focused on building out a product line, future plans include a focus on data exchanges and data-exchange-related activities. With respect to penetration rate, Locomation wants to find out how much coverage there is nationwide. The company does not want to rely on work zone data without full coverage. Work zone data are viewed as its highest-priority need. This is echoed by other automated vehicle developers. Original equipment manufacturers (OEMs) want improved work zone data so their level 2+ and level 3 systems are not disengaged when their vehicles approach work zones. Other needs include weather, weigh station bypass, enforcement actions, and rail grade-crossing data. Locomation would be interested in sharing its data in a data exchange but would need more information on requirements and use cases.

Locomation's privacy concerns depend on the data. For example, it does not think sending an alert for a possible pothole on a roadway would be problematic, but anything that could be used to identify a truck or route would need to be scrubbed. When it comes to data specifications and standards for its V2X data, Locomation adheres to Society of Automotive Engineers standards although they are not a perfect fit for all data.

Case Studies

Work Zone Data Exchange

The WZDx enables IOOs to make standardized work zone data available for third-party use. The objective is to make travel on public roads safer and more efficient through universal access to data on work zone activity. Specifically, the project aims to get data on work zones into vehicles to help

automated driving systems and human drivers navigate more safely (*3*). The WZDx continues to mature, allowing for more and varied contributors, users, and types of information.

Via the WZDx, the Federal Highway Administration is leading efforts to develop a nationwide standard approach for collecting, organizing, and sharing data on the when, where, and how of work zone deployment. As the nation develops an increasing reliance on technology and next-generation transportation management, ensuring the availability of consistent, reliable data describing work zone events is critical to enabling agency management of highway operations.

This national initiative aims to create and accelerate the adoption of a consistent language for communicating work zone activity data across jurisdictional and organizational boundaries. Adopting this common language will result in enhanced work zone management practices, leading to improved mobility and safety in and around work zones for workers and the traveling public. Numerous parts of the Texas Department of Transportation (TxDOT) and other jurisdictions across Texas are actively using or investigating WZDx as a methodology for sharing critical work zone activity data. It is believed that not only will these data help both CAVs but could also help improve work zone safety overall and provide a source for advanced traveler information for non-connected vehicles (*3*).

Common challenges experienced when implementing a WZDx data feed include the following (*4*):

- The project seems intimidating.
- There are budgetary restrictions.
- The benefit to consumers is unclear.
- The benefit to the agency is unclear.
- There is uncertainty about whether JavaScript Object Notation (JSON) elements within the data feed are correct.
- There is uncertainty about how to provide data for recurring work zones.
- Work zone examples within the WZDx GitHub repository do not align perfectly.

Best practices include the following:

- Start small. Focus on specific and clear work zone parameters already available, potentially from another project.
- Start with manual data entry. Then, only attempt to immediately automate some aspects of the agency's work zone information. Get a few elements working successfully first.
- Use contract resources when available.
- Solicit inputs from data consumers. Ask automakers, mapping companies, etc., why the data are important. Ask other state agencies how their data are being used.
- Take advantage of connected field devices deployed at active work sites.
- Leverage JSON validation tools available through the WZDx GitHub repository.
- Request technical assistance from the WZDx Help Desk or GitHub.

Florida Department of Transportation Data Exchange

A Florida Department of Transportation (FDOT) project is the first in the United States to develop a V2X data exchange, with the aim of capturing data from thousands of devices across CAV and

infrastructure networks. A key aim of the project is to standardize the collection, analysis, and sharing of data from several proprietary systems, which have different coding and encryption methodologies, and to unify privacy and security mechanisms across different sources and users (5). Figure 2 illustrates the ecosystem of the FDOT platform.

The exchange will capture anonymous data both from standardized onboard units communicating directly with FDOT-owned roadside units and from the proprietary data feeds of various car manufacturers. These connected vehicle data will be fused with a range of other data from FDOT-owned infrastructure and third-party data feeds. This enriched data stream will be the basis for real-time and historical analysis, leveraging machine learning and traditional algorithms *(***Error! Bookmark not defined.***)*. Figure 2 illustrates how the data exchange ecosystem works.

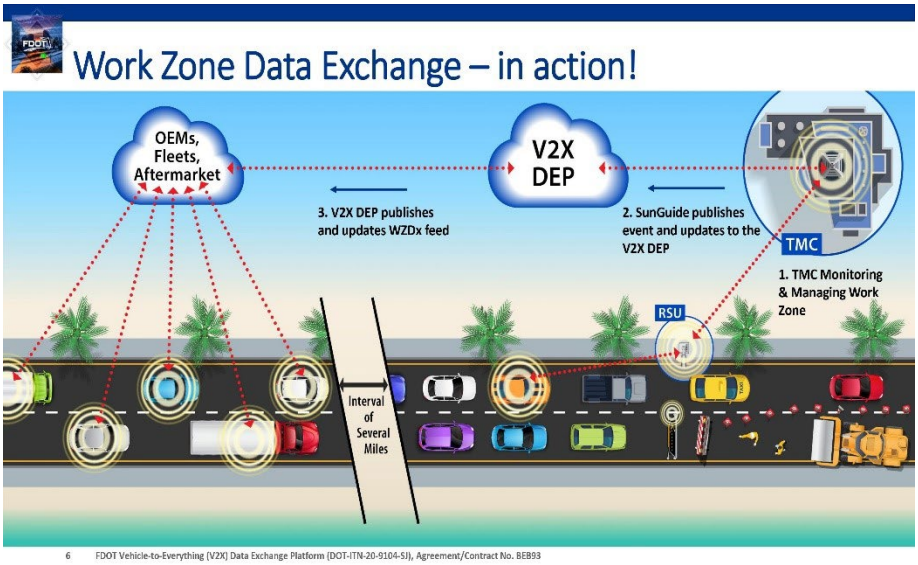

Figure 2: Work Zone Data Exchange Ecosystem (5)

FDOT program participants include Ford Mobility, which will supply V2X data from its connected vehicle platform; Iteris, a smart mobility company based in California; Florida International University; Amazon Web Services; Google; several OEMs; and logistics and fleet companies.

Owned by FDOT, the resulting data exchange will be available for use by other public agencies throughout the United States. In addition, this enables platform standardization across agencies and users and provides mutually beneficial cost sharing to develop new and improved functionality over time. Figure 3 shows the partners and participants at each level of the exchange.

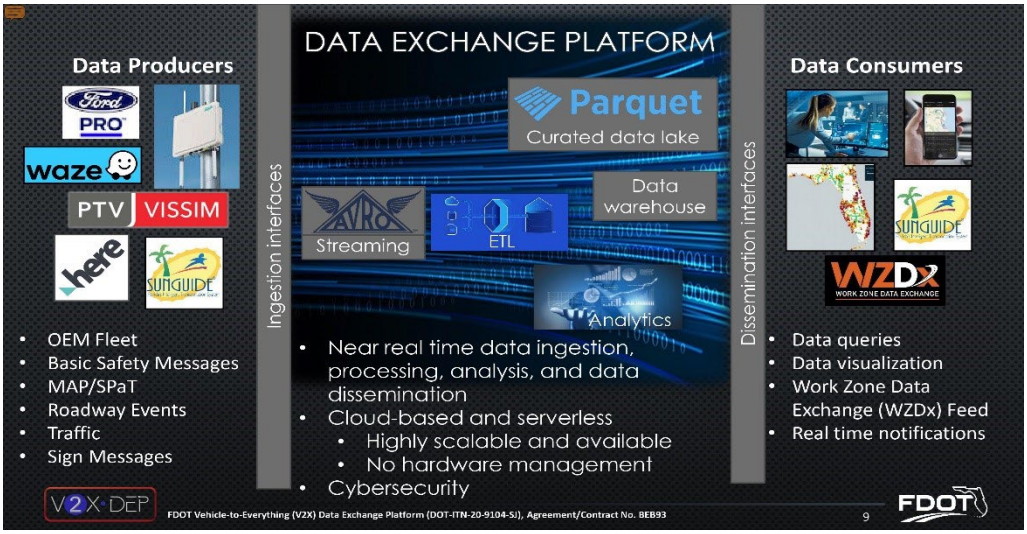

Figure 3: FDOT Data Exchange Platform Participants (5)

Summary and Conclusion

This white paper discusses, at a high level, the definition of a data exchange, the various components, data governance, security and compliance issues, industry leader thoughts, and case studies of operational data exchanges for transportation. Creating data exchanges for CAV data presents a unique opportunity to be able to expand research, product offerings, traveler information, and potentially roadway safety through cooperation and sharing. As vehicles become increasingly connected with each other and their surroundings, the data that are produced will increase exponentially. No single agency or company will have the ability to collect, clean, store, and analyze all of the data. This represents an opportunity to identify and elaborate on mutually beneficial methods of sharing these data, particular among public agencies and the users of the roads these agencies own, operate, and maintain. Developing resources such as the WZDx and the FDOT data exchange demonstrates an appetite for a platform that allows this type of sharing.

Because CAV data are a new genre of data, there are still many questions about many aspects of sharing, including privacy, security, and formatting. The needs of various users may be different depending on the use case. For example, the privacy considerations related to pothole locations are substantially different than those for vehicle journey data that may identify discrete points of a trip for individual vehicles. Currently, in the United States, there is no comprehensive approach to data privacy regulations. The Federal Communication Commission's nonbinding fair information practices guide data privacy protections, but federal law does not require companies to have a privacy policy or notify consumers of their privacy practices. States such as California, Nevada, and Maine have data privacy laws, but only California's pertain to non-online business practices.

Data standards and specifications are a critical area that also needs to be addressed. Like privacy, standards and specifications could vary based on use case. Numerous companies are attempting to follow and/or develop data standards, but the industry has not yet defined what those are in all cases.

Solving these issues can help provide a path for data exchanges moving forward, particularly for high-priority use cases such as work zones, real-time traffic and road conditions, and roadway inventories.

Opportunities

To effectively move forward with data exchanges to support the increasing levels of CAV activity in the state, Texas should consider taking an ownership role in participating in and/or developing data exchanges. Specifically, Texas should consider:

- Developing a comprehensive list of data exchanges that are pertinent to the development and deployment of CAVs and that also will improve operations and safety for human-driven vehicles. This would include an inventory of what private-sector companies would participate in data exchanges for any given use case.
- Identifying the most useful data exchange CAV and safety use cases for the state and its jurisdictions by collaborating with current and future users to identify needs.
- Developing an action plan for using or creating a data exchange for a particular use case that enjoys strong support from both public- and private-sector participants.
- Identifying potential failure points of data exchange collaboration and mechanisms to mitigate the concerns that could impact acceptance and usage.
- Encouraging TxDOT, with the help of metropolitan planning organizations, private contractors, and cities, to make a push for improved WZDx reporting statewide.
- Continuing the procurement of third-party data sources because these data platform-sharing initiatives promote standardization, cooperation, and data fluency at all levels of roadway operations.

References

1 Olavsrud, Thor. "What Is Data Governance? Best Practices for Managing Data Assets." *CIO*, March 24, 2023, https://www.cio.com/article/202183/what-is-data-governance-a-best-practices-framework-for-managing-data-assets.html.

2 U.S Department of Transportation. "Work Zone Data Exchange (WZDx)." March 22, 2023, https://www.transportation.gov/av/data/wzdx.

3 U.S. Department of Transportation. *Work Zone Data Early Adopters Guide*. March 2021.

4 Crowe, Robert. "SwRI Developing Connected Vehicle Data Exchange Platform for Florida Department of Transportation." Southwest Research Institute, November 9, 2021, https://www.swri.org/press-release/swri-developing-connected-vehicle-data-exchange-platform-fdot.

5 Ponnaluri, Raj. *Florida's Investments in the CAV Program*. FAV Summit, https://favsummit.com/wp-content/uploads/ppts/2019/3-Florida-Connects_FAV-Summit-Raj-Ponnaluri_111919.pdf.