COOPERATIVE & AUTOMATED TRANSPORTATION PROGRAM

TEXAS DEPARTMENT OF TRANSPORTATION

Can TxDOT Keep Up with CAT Data?

CAT Brief

Strategic Initiatives and Innovation Division

November 2024

Table of Contents

List of Tables

List of Figures

Acronyms and Abbreviations

AASHTO	American Association of State Highway and Transportation Officials
ADAS	Advanced Driver Assistance System
AI	Artificial Intelligence
API	Application Programming Interface
AV	Automated Vehicle
CAT	Cooperative and Automated Transportation
CAV	Connected and Automated Vehicle
CDOT	Colorado Department of Transportation
CEN	European Committee for Standardization
CRIS	Crash Records Information System
CTI	Connected Transportation Interoperability
CV	Connected Vehicle
C-V2X	Cellular Vehicle-to-Everything
DOT	Department of Transportation
DSRC	Dedicated Short-Range Communications
ERTICO	Intelligent Transportation System Europe
ETC	Eastern Transportation Coalition
FHWA	Federal Highway Administration
FOIA	Freedom of Information Act
GDPR	General Data Privacy Regulation
GM	General Motors
GPS	Global Positioning System
IOO	Infrastructure Owner-Operator
ITE	Institute of Transportation Engineers
ITS	Intelligent Transportation System
JPO	Joint Program Office
MAG	Maricopa Association of Governments
ML	Machine Learning
NAP	National Access Point
NGTMDD	Next Generation Traffic Management Data Dictionary
ODOT	Ohio Department of Transportation
OEM	Original Equipment Manufacturer
PII	Personally Identifiable Information

RDI	Roadway Digital Infrastructure
RSU	Roadside Unit
SAE	Society of Automotive Engineers
SENSORIS	Sensor Interface Specification
TMDD	Traffic Management Data Dictionary
TTI	Texas A&M Transportation Institute
TxDOT	Texas Department of Transportation
USDOT	U.S. Department of Transportation
VDH	Vehicle Data Hub
V2I	Vehicle-to-Infrastructure
V2V	Vehicle-to-Vehicle
V2X	Vehicle-to-Everything
WZDx	Work Zone Data Exchange
5G	Fifth Generation Mobile Broadband

1. Introduction

The Texas Department of Transportation's (TxDOT's) transportation network is the largest in the nation, with more than 80,900 miles of roadways, 271 general aviation airports, and 78 million tons of cargo moving through its seaports. For TxDOT, continuing to investigate innovative and economical methods for creating a safe and reliable transportation system for all users is critical. The integration of connected vehicles (CVs), automated vehicles (AVs), and other emerging transportation technologies into the state's multimodal transportation system offers numerous potential benefits to the traveling public. To support the efficient integration of these technologies, TxDOT launched the Cooperative and Automated Transportation (CAT) Program based on the CAT Strategic Plan and Program Plan.

The statewide CAT program supports the agency's mission of Connecting You With Texas. This program explores emerging technologies to address some of the state's greatest challenges in the areas of safety, mobility, environment, and funding, while executing strategies to accommodate disruptive changes and emerging technology trends. TxDOT seeks to proactively integrate CAT initiatives into transportation projects, from planning, design, and construction to operations and maintenance, rather than respond reactively to its proliferation in the multimodal transportation system.

A key initiative is to fully understand the current trends, legislative issues, and industry perspectives on challenges and opportunities presented by the exponential growth of CAT data, particularly from CVs. Though CAT data includes a large landscape of data options depending on the desired CAT use case, CV data is the focus of this brief because it provides new data points at a high frequency of transmission. Additionally, the CV market is inconsistent and still working through many unknowns regarding standards and legislative policies.

1.1 Project Background

The rapid growth of CAT infrastructure[1] and related CV technologies is generating an exponential increase in data volume, velocity, and variety. To effectively manage and leverage CAT data, TxDOT must develop a robust digital infrastructure to handle the vast amounts of data produced, which presents significant data management challenges.

TxDOT must upgrade existing digital infrastructure to handle the expected volume from emerging technology sources, especially CVs, which could revolutionize transportation management. CV data comes from vehicles with sensing and communication technologies, enabling vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication.

In addition to updating digital infrastructure, TxDOT and national transportation agencies must work together to build uniform national and local legislation around these emerging data sets. It is important to be transparent about the data collected and how it is used to benefit the traveling public.

Figure 1 shows the range of data generated by passenger vehicles. Transportation agencies are particularly interested in location, speed, driving behavior patterns, operational, and vehicle-to-everything (V2X) data. To manage this influx of data, new

[1] CAT infrastructure refers to the physical and digital assets required to support communication, data exchange, and coordinated actions within the CAT ecosystem, directly or indirectly influencing transportation system users and operational capabilities. Source: *CAT Brief: What is CAT Infrastructure?*

methods are needed for data consolidation and filtering, along with updated internal and external systems and architectures.

This CAT Brief, "Can TxDOT Keep Up With CAT Data?" aims to document current trends, legislative issues, and industry perspectives on standards and best practices in the CV data management landscape to help position TxDOT as a leader in this evolving field. This brief also summarizes TxDOT's current practices in utilizing CAT data compared to national practices.

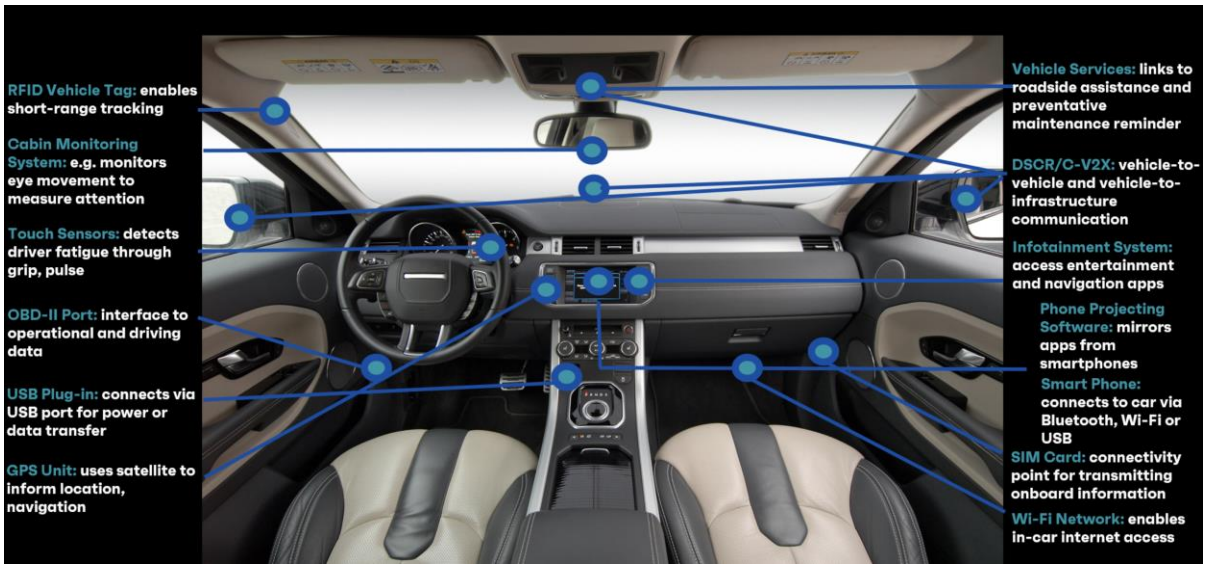


Figure 1: Illustration of Vehicle-Generated Data[2]

This CAT Brief is also part of a series of initiatives related to CAT data integration within TxDOT. **Figure 2** shows the relationship between the different CAT data-related initiatives.

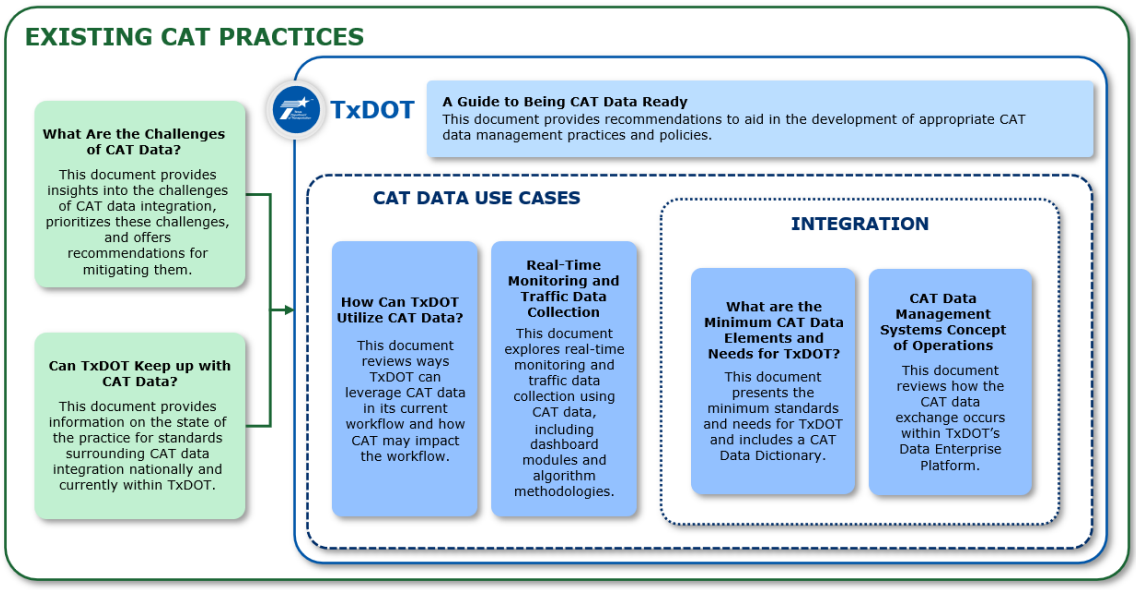


Figure 2: CAT Program Data-Related Initiatives

[2] Alliance for Automotive Innovation. https://www.autosinnovate.org/initiatives/innovation/connected-vehicles.

Figure 2 outlines a framework for TxDOT to effectively utilize and manage CAT data. It starts with establishing a foundation by assessing current practices ("Can TxDOT Keep Up with CAT Data?") and identifying challenges ("What are the Challenges of CAT Data?"). This informs the development of a guide ("A Guide to Being CAT Data Ready") for implementing CAT data use cases ("How can TxDOT Utilize CAT Data?"), such as real-time monitoring ("Real-Time Monitoring and Traffic Data Collection"). Finally, the framework addresses data integration ("What are the Minimum CAT Data Elements and Needs for TxDOT?" and "CAT Data Management Systems Concept of Operations") to ensure TxDOT is prepared to manage CAT data.

1.2 Methodology

This CAT Brief is informed by a comprehensive literature review examining existing knowledge on CV data in both the private and public sectors, as summarized in **Appendix A: References**.

A series of interviews were conducted with:

- **Industry leaders:** Including public agency officials, data aggregators, and original equipment manufacturers (OEMs) offering vehicle-based data and products, providing insights into current CV data market trends.
- **State Department of Transportation (DOT) representatives:** Identified through the American Association of State Highway and Transportation Officials (AASHTO), offering valuable perspectives on how peer agencies are addressing CV data challenges.

The list of interviewees can be found in **Appendix B: List of Interviewees and State DOT Representatives**. This combined research approach informed the development of a comparative matrix, which analyzes national guidance and state best practices in relation to TxDOT's current practices, considering policy issues and emerging data sources.

1.3 Organization of this CAT Brief

This CAT Brief is organized into the following sections:

- **Section 1: Introduction** – Provides an overview of the project background and summarizes the information presented in this document.
- **Section 2: Current Industry Trends of CV Data** – Reviews industry trends, including perspectives of data providers and OEM interests.
- **Section 3: Policy Issues: Privacy, Security, and Legislation** – Explores privacy, security, and legislative issues with a focus on CV data initiatives.
- **Section 4: National and State Activities: Standards and Best Practices** – Presents federal activities and prospects for national standards in the CV data space, and highlights state DOT organizational perspectives gathered from peer agencies.
- **Section 5: TxDOT Considerations for Keeping Up With CAT Data** – Summarizes findings and offers considerations for using, maintaining, and managing CV data.

2. Industry Trends for CV Data

2.1 Status and Outlook of the CV Data Provider Market

CVs generate vast amounts of data through sensors and on-board systems. A 2020 study forecasts the CV data market will reach $332 billion by 2030.[3] Non-automotive companies such as insurance, diagnostics, repair and maintenance, and fleet management are projected to represent 46% of the global connected services market by 2030.

CV mobility data products are offered in four forms: raw, concealed raw, aggregated, and synthesized. These are delivered via cloud services or web platforms from either the raw data provider (e.g., vehicle OEMs), third-party providers (e.g., Wejo [defunct] or StreetLight), or service provider platforms (e.g., INRIX/GM SafetyView).[4] As of late 2023, OEMs and data providers are reluctant to share raw CV data with public agencies.

To assess the current CV data market, interviews with several data providers were conducted. **Appendix C: Highlights of CV Data Providers and Products** highlights the information gathered during those interviews.

Given the growing volume of CV data, end users must balance in-house capabilities with external vendors. It is crucial to understand the tradeoffs between CV data types and operational needs, including technical capabilities, infrastructure, and privacy protections.

Table 1 illustrates these tradeoffs. As data end users, transportation agencies must assess benefits and costs of each data product for their use cases, considering impacts on data infrastructure, management, and analysis.[5]

Table 1: Relative Tradeoffs Between CV Data Products

Categories of CV Data Products	Flexibility in Application	Availability	End User Requirements		
			Technical Capabilities	Data Infrastructure	Privacy and Security Protections
Raw data(a)	High	Very Low	High	High	High
Concealed Raw data(b)	High	Very Low	High	High	Medium
Aggregated data	Medium	High	Medium	Medium	Medium
Synthesized data	Low	High	Low	Low	Low

(a) "Raw" CV data is defined here as individual way points having clear start and end points, available for all geographic locations, and applied in any use cases (i.e., no restricted applications).
(b) Concealed Raw data are individual way points with obfuscation at the start and end of journey to protect privacy

Vehicle Manufacturer Perspectives on CV Data

A 2023 survey of auto executives identified vehicles with connectivity and digital features as a top trend shaping future mobility, second only to electrification.[6] OEMs recognize the value of vehicle-generated data for internal operations and dealer networks. However, monetizing these data through third-party sales remains less understood and undervalued by OEMs.

[3] Ptolemus Consulting Group. (2020). Vehicle Data Market Global Study - Executive Summary.
[4] Singh et al. (2022). State of Emerging Mobility Big Data Sources and its Applications.
[5] Ibid.
[6] McKinsey & Company. (2023). Corporate Business Building to Unlock Value in Automotive Connectivity.

The survey highlighted several challenges OEMs face in data sharing:[7]

- Risk aversion and technical immaturity
- Slow business transformation and ecosystem development
- Data privacy, security, and regulatory compliance concerns
- Lack of standardization and interoperability
- Limited understanding of data value and new market opportunities

OEMs have shown reluctance to directly engage with transportation agencies for data sharing, preferring to work through data brokers. While open to sharing data for long-term infrastructure planning, OEMs face challenges in maintaining data quality and standardization across diverse vehicle models and formats.

A 2020 study, validated by industry experts, illustrates the automotive industry's maturity in working with data brokers or vehicle data hubs (VDHs). The study indicated varying levels of partnership willingness among OEMs, as shown in **Figure 3**.[8] Note that this figure was developed based on 10 years of expertise from 150 consulting assignments in the mobility and CV markets and 49 interviews of key stakeholders including OEMs.

Industry feedback suggests that OEMs are more inclined to support research or demonstrations rather than mainstream use of their data in DOT operations. For broader adoption, OEMs require a clear value proposition that addresses liability concerns and aligns with their revenue expectations.[9]

[7] Ibid.
[8] Interview with Hilary Cain, Alliance for Automotive Innovation. December 18, 2023.
[9] Interview with Stan Young, Chief Data Officer, and Sheryl Bradley, TSMO Director, Eastern Transportation Coalition. November 21, 2023

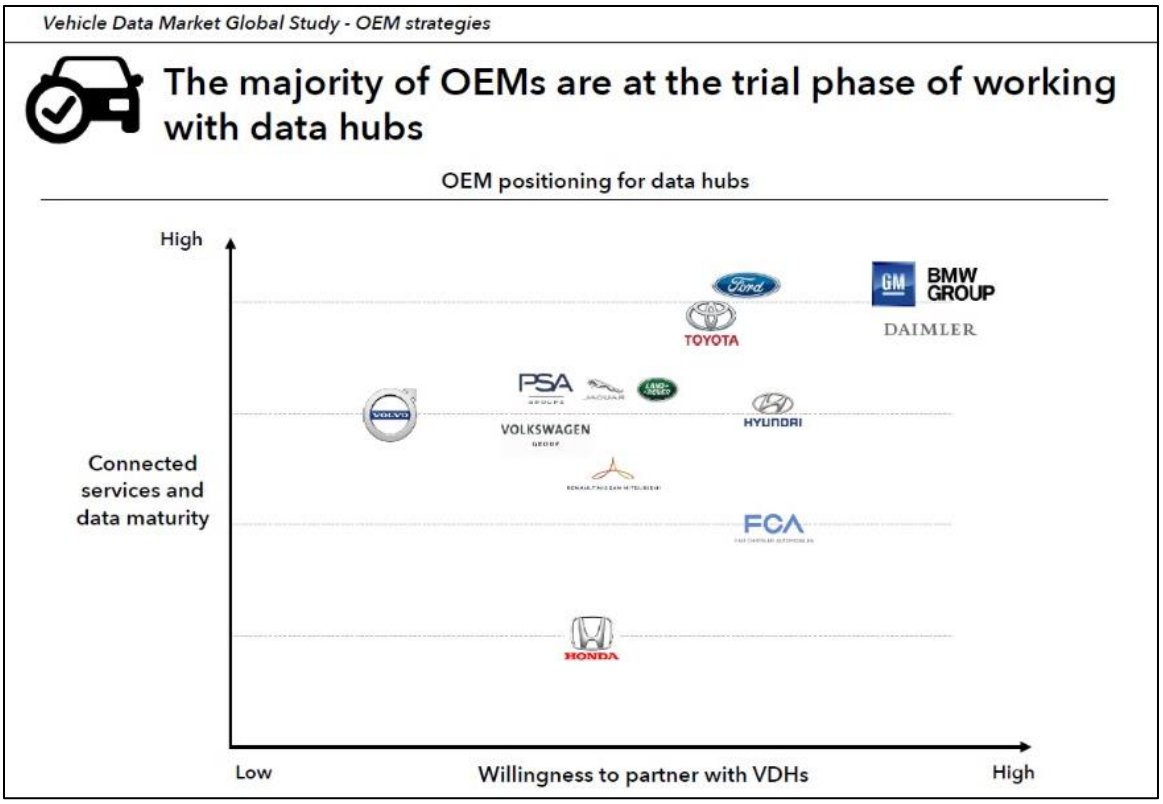


Figure 3: OEM Positioning for Data Hubs[10]

2.2 Future Sources of CAT Data

2.2.1 Automated Vehicles

AV sensor-generated data is currently unavailable to public agencies or third-party data vendors. However, research and testing have been conducted using supported data for various use cases.[11] As vehicle automation advances, both CVs and AVs are expected to consume significant amounts of data for safe operation and generate an increased volume and variety of data during operation.

Data generated by vehicles falls into two categories:

- **User-Generated**: Includes passenger identity, app usage, service information (e.g., tolls, parking), and communications. These data come from mobile device pairing or direct user interaction.
- **Vehicle-Generated**: Encompasses vehicle measures, safety data, environmental probes, diagnostics, and biometrics. These vehicle-generated data are produced from advanced sensors, processors, enhanced driver interfaces, and other on-board units that can record and deliver data internal and external to the vehicles.[12]

[10] Ptolemus Consulting Group. 2020. Vehicle Data Market Global Study - Executive Summary.
[11] Center for Transportation Research. (2024). Working with Autonomous Trucks to Improve Routine Maintenance Operations. TxDOT Research Project 0-7129.
[12] Texas CAV Task Force. (2021). Connected and Automated Vehicle Data Issues and Opportunities.

2.2.2 5G Automotive Data

5G technology is revolutionizing telecommunications with lower latency and greater bandwidth. This is largely due to 5G technology's:

- Enhanced data granularity and real-time capabilities
- Improved location precision through 5G microcells
- Expanded data sources, including vehicles, smartphones, and traffic control devices

5G enables advanced cellular vehicle-to-everything (C-V2X) functionalities with higher data rates, ultra-low latency, and massive connectivity. It supports virtual and physical roadside units (RSUs), as well as V2I. Meanwhile, dedicated short-range communications (DSRC), a legacy V2I technology, faces market decline.[13]

Outdoor Wi-Fi 6, the latest Wi-Fi technology for outdoor use, complements 5G in areas with low coverage. The prevailing technology will depend on development stage, infrastructure support, regulations, and market adoption.

5G-enabled vehicles will transform transportation business models. However, the increased data from 5G raises privacy concerns, necessitating improved data protection measures. While offering new opportunities, this data surge also amplifies the tradeoffs between a transportation agency's in-house capabilities and contracted services to manage CV data.

2.2.3 Growth in Data from Emerging Technologies

While the future of AVs and 5G technology in the automotive industry holds immense potential, the timeline for widespread adoption remains uncertain. TxDOT is actively engaged with the AV community via the Texas Connected and Automated Vehicle (CAV) Task Force, focusing on data sharing.[14] However, OEMs have been hesitant to adopt 5G due to limited business cases.[15] This hesitation, along with ongoing concerns surrounding data privacy, could impact the value of digital data for transportation planning and operations.

Future datasets are expected to expand significantly with the integration of data from sources such as satellites, LiDAR, and video. Furthermore, incorporating emerging technologies such as artificial intelligence (AI) and machine learning (ML) into transportation analytics presents both opportunities and challenges for decision-making and traffic optimization.

3. Policy Issues: Privacy, Security, and Legislation

3.1 Privacy and Security Issues

TxDOT has implemented data management policies to protect personal information and mitigate security risks. The unique nature of CV data raises specific concerns regarding the protection of personally identifiable information (PII). Key privacy and security issues include:

[13] 5G Americas. (2021). Vehicular Connectivity: C-V2X & 5G.
[14] Texas CAV Task Force. Texas CAV Task Force Subcommittees.
[15] Interview with Steve Schwinke, Sibros. November 27, 2023.

- **Aggregation and Anonymization:** Although data providers utilize various techniques to mask PII, anonymization may not be entirely reliable, and risks of data inference exist.[16]
- **Permissible Uses and Access Control:** Defining permissible data uses, implementing data minimization procedures, and weighing the cost of data retention against repurchasing are crucial for mitigating risks.
- **Private Companies vs. Public Agencies:** Differing data protection standards, especially regarding open records laws, create inconsistencies and concerns for companies sharing data with U.S. public agencies. This makes private companies hesitant to share sensitive information.

3.2 Legislative Issues

Data privacy and security in Texas are governed by laws and policies designed to protect the public interest. These regulations limit access to data containing PII.

Key Legislation:

- **SB 475 (2021):**[17] Requires state agencies to implement robust data management practices and obtain consent for any data identifying individuals or their locations. For CV data, this means providers must ensure they collect data with car owner consent.
- **HB 4 (2023):**[18] Establishes data privacy regulations for private companies, but excludes state agencies, raising concerns about enforcement, potential misuse of data by agencies, and legal risks associated with data retention.

Key Concerns:

- **Enforcement:** State agencies may lack resources to ensure data providers comply with privacy laws.
- **Misuse:** Policymakers worry about potential misuse of individual trip data held by state agencies, especially by law enforcement.
- **Data Retention:** Legal risks and obligations for data retention are a concern, especially regarding what TxDOT can redact from mobility data under open records requests.

4. National and State Activities: Standards and Best Practices

A "data standard" is a technical specification for consistent data storage and exchange across systems.[19] The literature review and interviews confirm that there are no national standards for CV data, only proprietary standards and definitions developed by individual companies.

4.1 Federal Activities

Several federal initiatives are underway to develop standards for CV data. These include:

[16] TxDOT Transportation Technology Task Force, Panel on Data Privacy and Cybersecurity. September 19, 2023.
[17] Texas 87th Legislature. (2021). SB 475.
[18] Texas 88th Legislature. (2023). HB 4.
[19] Federal Enterprise Data Resources. Standards Overview.

U.S. Department of Transportation (USDOT) Activities:

- **V2X Roadmap:** USDOT is finalizing a 10-year V2X national deployment roadmap through stakeholder engagement, but developing standards for OEM-supplied vehicle data is not a priority yet.
- **V2X Reference Architecture:** USDOT is updating the V2X reference architecture and collaborating with the Institute of Transportation Engineers (ITE) and the Society of Automotive Engineers (SAE) Connected Transportation Interoperability (CTI) Committee on standards development, focusing on connected intersections and vulnerable road users.

Federal Highway Administration (FHWA) Initiatives:

- **Work Zone Data Exchange (WZDx):** WZDx serves as a model for standardizing data sharing with the private sector. USDOT is currently working with ITE on connected work zone standards and expects an update in the next year.[20]
- **Next Generation Traffic Management Data Dictionary (NGTMDD):** ITE's NGTMDD aims to address emerging operational needs, including the role of private entities in real-time data management. The Traffic Management Data Dictionary (TMDD) in its current form provides the dialogs, message sets, data frames, and data elements to manage the interoperability of Intelligent Transportation System (ITS) devices.
- **Roadway Digital Infrastructure (RDI) Strategy:** FHWA is developing a National Roadway Network Digital Infrastructure Strategy centered on RDI to improve transportation system outcomes. This strategy focuses on transcontinental prototype corridors and the integration of data and assets to support end-user applications.

While these initiatives show progress, a cohesive national strategy for CV data standards is still needed. Furthermore, greater collaboration with OEMs is crucial to ensure that standards development aligns with industry practices and facilitates seamless data integration.

4.2 National and International Standards for CV Data

While there are no national standards for CV data in the United States, European countries have been actively developing standards for vehicle data and transportation systems for over two decades, focusing on creating a unified digital infrastructure. European standards for the vehicle ecosystem are illustrated in **Appendix D: European Standards for Vehicle and Infrastructure Ecosystem.**

Key Initiatives:

- **Sensor Interface Specification (SENSORIS):** Led by Intelligent Transportation System Europe (ERTICO), this initiative aims to standardize data exchange between in-vehicle sensors and cloud platforms, involving key players from the automotive, mapping, and telecommunications industries.

[20] Conversation with Bob Sheehan, USDOT JPO, and Govind Vakapat, Turner-Fairbanks Highway Research Center, January 12, 2024.

- **National Access Points (NAPs):** Each European country is establishing NAPs to facilitate the exchange and reuse of transportation data, supporting EU-wide interoperable travel and traffic services.

Established Standards:

- **DATEX II (CEN/TS 16157):**[21] Serves as the standard for exchanging road traffic and travel information in Europe, supporting seamless travel and expanding its focus to include connected and automated mobility.
- **Transmodel (CEN/TS 12896):**[22] Provides a comprehensive model for public transportation information, ensuring interoperability among different systems and operators.

4.3 State Practices

Two state DOTs and one multi-state coalition were interviewed about their CV data usage approaches. Key findings are summarized below.

- **Colorado DOT (CDOT):** CDOT is testing CV applications, focusing on emergency alerts, incident management, snowplow preemption, and curve speed warnings. CDOT prefers raw data procurement and seeks USDOT leadership in developing interoperability standards.[23]
- **Ohio DOT (ODOT):** ODOT pursues targeted data sets for specific use cases. They are exploring accident mitigation concepts with Honda and have developed an open-source data sharing model for work zone applications. ODOT is also investigating emergency vehicle alert solutions.[24]
- **Eastern Transportation Coalition (ETC):** ETC employs a multi-state model for vendor evaluation and data validation. While effective, the model has limitations, including restricted entry for new vendors and underutilization concerns.[25]
- **Maricopa Association of Governments (MAG):** In addition to the interviews, the literature review revealed a case study of MAG, which illustrates a balanced approach to public and private resourcing, emphasizing in-house expertise alongside external platforms. They continue exploring data applications to support informed decision-making.[26]

5. TxDOT Considerations for Keeping Up with CAT Data

The CV data market is rapidly evolving and experiencing volatility as it progresses toward mature business models. The anticipated growth of AVs and 5G networks will further increase data complexity and volume.

Table 2 provides a comparison of national and state best practices with TxDOT's current practices across various categories related to CV data. To effectively navigate this dynamic landscape, TxDOT should continue to proactively engage with industry stakeholders. This

[21] DATEX II (CEN/TS 16157) https://datex2.eu/
[22] Transmodel (CEN/TS 12896) https://www.transmodel-cen.eu/
[23] Interview with Heather Pickering-Hilgers, Colorado DOT. January 24, 2024.
[24] Interview with Nick Hegemier, Ohio DOT. February 9, 2024.
[25] Interview with Josh Gilman, Peter Van der Werf, TomTom. December 11, 2023, and TomTom trusted supplier proposal to TxDOT, November 3, 2023.
[26] Texas Technology Task Force. Powerful Data: Applications, Knowledge Sharing, and Platform Integration. White Paper by CTR for the Texas Technology Task Force. July 2023.

engagement will ensure that the agency's needs are considered as new data products and services are developed. Potential external actions for TxDOT to take to keep up with CAT data are described in **Table 3**. Potential internal actions TxDOT can take to keep up with CAT data are described in **Table 4**.

Table 2: Matrix of Findings[27]

Category	Description of Baseline Understanding	National and State Practices	TxDOT Practices	Gaps and Future Considerations
Data Sources	New data available: Private data markets (OEMs and brokers) are in a state of flux and expansion. CAV data and probe sources data such as cameras, detectors, and sensors.	• No national initiatives related to CV data. • Leading states obtain data from different sources and independently pursue data sources.	• TxDOT is a leading state in CV data procurement, but it has been disrupted by the departure of the largest vendor from the market. • Integration of CV data with TxDOT open data portal (roadway assets, boundary dataset, planning datasets, traffic-related data, safety-related data, and highway performance data) is desired.	• Define priority use cases to identify required data sets. • Remain engaged with CV data industry, potentially leveraging existing CAV Task Force activities. • Consider benefits of working collectively with other infrastructure owner-operators (IOOs) to influence the data marketplace. • AVs, 5G connectivity, and emerging data sources will expand future data sources. • Provide open Application Programming Interfaces (APIs) for data access.
Standards	Proprietary standards for internal uses; industry-led mapping consortia.	No industry, state, or federal standards for CV data.	No TxDOT CV data standards.	Establish or join a coalition with industry and other state DOTs/IOOs to develop CV data standards, including metadata standards, data quality, and security.
Legislative Outlook	Legislation mostly focuses on data privacy and security to protect public interest.	No national legislation on CV data; laws vary at state level.	TxDOT is operating under SB 475 for public agencies; new legislation HB 4 applies to private companies.	• Potential state laws related to privacy protections by public agencies. • Potential federal laws introduced compelling data sharing and strengthening privacy requirements, similar to the General Data Privacy Regulation (GDPR).

[27] National Science and Technology Council and U.S. Department of Transportation. Ensuring American Leadership in Automated Vehicle Technologies: Automated Vehicles 4.0. January 2020.
U.S. Department of Transportation. Preparing for the Future of Transportation Automated Vehicles 3.0. 2018
Texas State Senate. SB 2205: Relating to Automated Motor Vehicles. Legislative Session: 85(R). September 2017.
UC Davis. Federal, State, and Local Governance of Automated Vehicles. December 2018
U.S. Department of Transportation. Data for Automated Vehicle Integration (DAVI). https://www.transportation.gov/av/data
The TxDOT Data Classification policy
The TxDOT Information Resources and Security Requirements

Category		Description of Baseline Understanding	National and State Practices	TxDOT Practices	Gaps and Future Considerations
Organizational Perspectives	**Data Management Infrastructure**	Need to manage massive amounts of CAT data and integrate with current TxDOT infrastructure data. Requires efficient and legal process to receive, process, and share data.	No national guidance; varies by state.	Current TxDOT Data Pyramid was developed to facilitate decision making at the Commission, Administration, and District/Division levels. Enterprise Data Committee for data governance and management; Enterprise Data Platform includes Open Data Portal, TxDOT-Connect, Lonestar™, Crash Records Information System (CRIS), Site Manager, and Maintenance Management System.	Advance Storage Systems: Choose storage systems (third-party) and develop cloud-based platforms that can handle large volumes of data efficiently.
	Budgeting	Budgeting preparation and policy change to support CAT data development.	No national guidance; varies by state.	No specific TxDOT CAT data budgeting.	• Budget at appropriate scale for data processing, analytics tools, and software that handles real-time and batch processing of data. • Set aside funds for the regular maintenance and potential upgrades. • Budget for scalable solutions for increasing volume of data over time. • Assess the economic benefits of CAT deployment.
	Analytical Capabilities and Staff Training	Training and documentation to help staff and contractors keep up with CAT data.	No national guidance; varies by state.	Current data training includes Transearch, INRIX, and StreetLight. TxDOT Digital Roadway Data user call held monthly to provide updates to both internal employees and contractors.	Create regular training programs to enhance the skills of personnel involved in managing and analyzing CAT data.
Privacy Concerns		CV data may include personal information; limitations to accessing the data are through state laws or policies, or technical barriers that limit how data can be queried.	No national standards on data privacy for CV data; state-by-state policies vary.	TxDOT employs data privacy and security experts and has enacted policies and practices to protect PII.	• Stricter privacy requirements: Data Ethical Code for CAT data. • Transparent CAV data management plan to increase public awareness. • Secure CAV data transmission with encryption protocols.

Table 3: External Actions to Keep Up With CAT Data

Action	Description
Develop a Joint Value Proposition with OEMs	Work with OEMs and data brokers to demonstrate the societal benefits of sharing anonymized, aggregated CV data, such as improving safety and reducing travel time.
Establish or Join a Coalition with Other State DOTs and IOOs to Develop Standards	Data providers value IOOs collaborating on common data formats. A coalition could create a larger market than individual state efforts. While, European standards may offer a starting point, existing efforts, such as CAV task forces, could be leveraged to engage with industry representatives.

Table 4: Internal Actions to Keep Up With CAT Data

Action	Description
Define Priority Use Cases	Identify priority use cases for CV data to guide resource allocation and vendor selection. Developing in-house skills should be complemented by leveraging existing platforms for specific analyses. As CV data availability may be limited, industry engagement becomes critical.
Recognize the Tradeoffs in Privacy and Security with Different Forms of CV Data	Resourcing decisions should consider risks of purchasing and retaining CV data for internal analysis. TxDOT's in-house privacy and security experts should be involved in industry engagement and capability building. TxDOT's partnership with Texas A&M Transportation Institute (TTI) for CV data infrastructure provides additional capabilities and mitigates privacy-related risks, including direct privacy-related criticism by the public or policy makers. As a state agency, TTI follows the same laws as TxDOT, including potential Freedom of Information Act (FOIA) disclosure concerns.

5.1 Summary and Next Steps

Transportation data systems need upgrading to handle increasing volumes, particularly from connected vehicles. These data can improve operations, maintenance, and planning. Future technologies will only increase the frequency and volume of data exchange of these valuable datasets. **Figure 4** describes the evolution of CAT data practices for TxDOT to be prepared for a mature CAT data ecosystem. This progression aims to integrate emerging data sources for enhanced transportation efficiency and safety.

Figure 4: Evolution of CAT Data Practices For Keeping Up with CAT

- The existing system primarily relies on various independent data storage.
- There is limited storage availability and inadequate lifecycle management.
- Similar data from different sources are collected and stored in uniform formats.
- There is a lack of data transparency and limited resources from sources.
- There is a lack of skilled resources in data management and analytics.
- Cybersecurity practices and policies for receiving and sharing data from innovative sources are insufficient.
- The infrastructure inadequately supports real-time data collection, analysis, and dissemination.
- Systems require human intervention for data analysis and utilization.

- Independent data storage systems are shifting into a single platform.
- Requirements for data transparency and resources are now included during the procurement phase of data.
- Training and understanding of CAT data are beginning.
- Understanding of CAT data is guiding policy development to manage these data securely.
- Hiring needs are being evaluated for data management and analytics.
- A standard structure for different data sources is being developed to properly clean and merge data.
- Procedures are being established for determining data latency and life-cycle needs.
- A cybersecurity audit and updates are being conducted.

- **Safer**: Implement data policies and cybersecurity best practices for confident data sharing and reception.
- **Efficiency**: Utilize a single platform for effective data storage and streamlined utilization, with consistent structuring across sources.
- **Greener**: Apply best practices for data lifecycle management to minimize unnecessary data storage.
- **Accessible**: Ensure equal, affordable, and convenient data access for all travelers to foster economic opportunity.
- **Reliability**: Develop a reliable system for data ingestion, processing, and storage to enhance data utilization.
- **Output**: Manage a digital ecosystem to deliver the right information to the right people at the right time, enabling informed decision-making.

Appendix A: References

5GAA Automotive Association. (2024). *Road Traffic Operations in a Digital Age: A Holistic Cross-Stakeholder Approach*. https://5gaa.org/content/uploads/2024/01/5gaa-white-paper-road-traffic-operation-in-a-digital-age.pdf

Federal Highway Administration Joint Program Office. (2020). *Infrastructure and V2X Mapping Needs Assessment and Development Support: Final Project Report*. https://rosap.ntl.bts.gov/view/dot/57545

National Academies of Sciences, Engineering, and Medicine. (2020). *Guidebook for Managing Data from Emerging Technologies for Transportation*. Washington, DC: The National Academies Press. https://doi.org/10.17226/25844.

Office of Management and Budget, General Services Administration, and Office of Information Services. (n.d.) *Federal Enterprise Data Resources*. https://resources.data.gov/standards/

Ptolemus Consulting Group. (2020). *Vehicle Data Market Global Study - Executive Summary*. https://www.ptolemus.com/research/the-vehicle-data-market-global-study/

Singh, G., Sivaraman, V., & Hard, E. (2022). *State of Emerging Mobility Big Data Sources and its Applications*. https://static.tti.tamu.edu/tti.tamu.edu/documents/TTI-2022-8.pdf

Texas CAV Task Force. (2023). *Connected and Automated Vehicle Data Issues and Opportunities*. Subcommittee on Data, Connectivity, Cybersecurity and Privacy. https://www.txdot.gov/content/dam/project-sites/cav-task-force/docs/2023/08/Final_Texas_CAVTF-WhitePaper_Data_08162023_Final.pdf

Texas Technology Task Force. (2023). Powerful Data: Applications, Knowledge Sharing, and Platform Integration. White Paper by CTR for the Texas Technology Task Force.

Zmud, J., Tooley, M., & Miller, M. (2016). *Data Ownership Issues in a Connected Car Environment: Implications for State and Local Agencies* (No. TTI/SRP/16/165604-1). https://static.tti.tamu.edu/tti.tamu.edu/documents/165604-1.pdf

Appendix B: List of Interviewees and State DOT Representatives

Name and Title	Organization
Private Organizations	
Hilary Cain, Senior VP of Policy	Alliance for Automotive Innovation
Steve Schwinke, VP Customer Engagement	Sibros
Laura Schewel, CEO	StreetLight/Jacobs
Josh Gilman, Sr. Sales Representative Peter Van de Werf, Manager of Engineering Stephanie Leonard, Head of Government and Regulatory Affairs	TomTom
Terri Johnson, Director, Public Sector Services	INRIX
Daniel Abugov, Senior VP, Sales Lauri Keller, Public Sector/Texas Account Manager Alun Yung, Solution Architect	HERE
Nathan Wade, VP Sales	Flow Labs
Niti Anand, Head of Strategy and Business Development, Smart Cities Sam Johnson, Client Solutions, GM Future Roads	GM
Rafay Khan, CEO Justin Collins, CTO	Moove.AI
Public Agencies	
Greg Jones, Transportation Specialist	FHWA Resource Center
Bob Sheehan, Program Manager, Multimodal ITS Govind Vadakpat, Program Manager	USDOT – Joint Program Office (JPO) and Turner-Fairbanks
John Corbin, Automated Vehicle Program Manager John Harding, Team Lead, Connected and Automated Vehicles and Emerging Technologies	FHWA - Operations
Stan Young, Chief Data Officer Sheryl Bradley, TSMO Program Director	Eastern Transportation Coalition (ETC)
Heather Pickering-Hilgers, Assistant Director of Mobility Technology	Colorado DOT
Nick Hegemier, Managing Director of Infrastructure	Ohio DOT/Drive Ohio

Appendix C: Highlights of CV Data Providers and Products

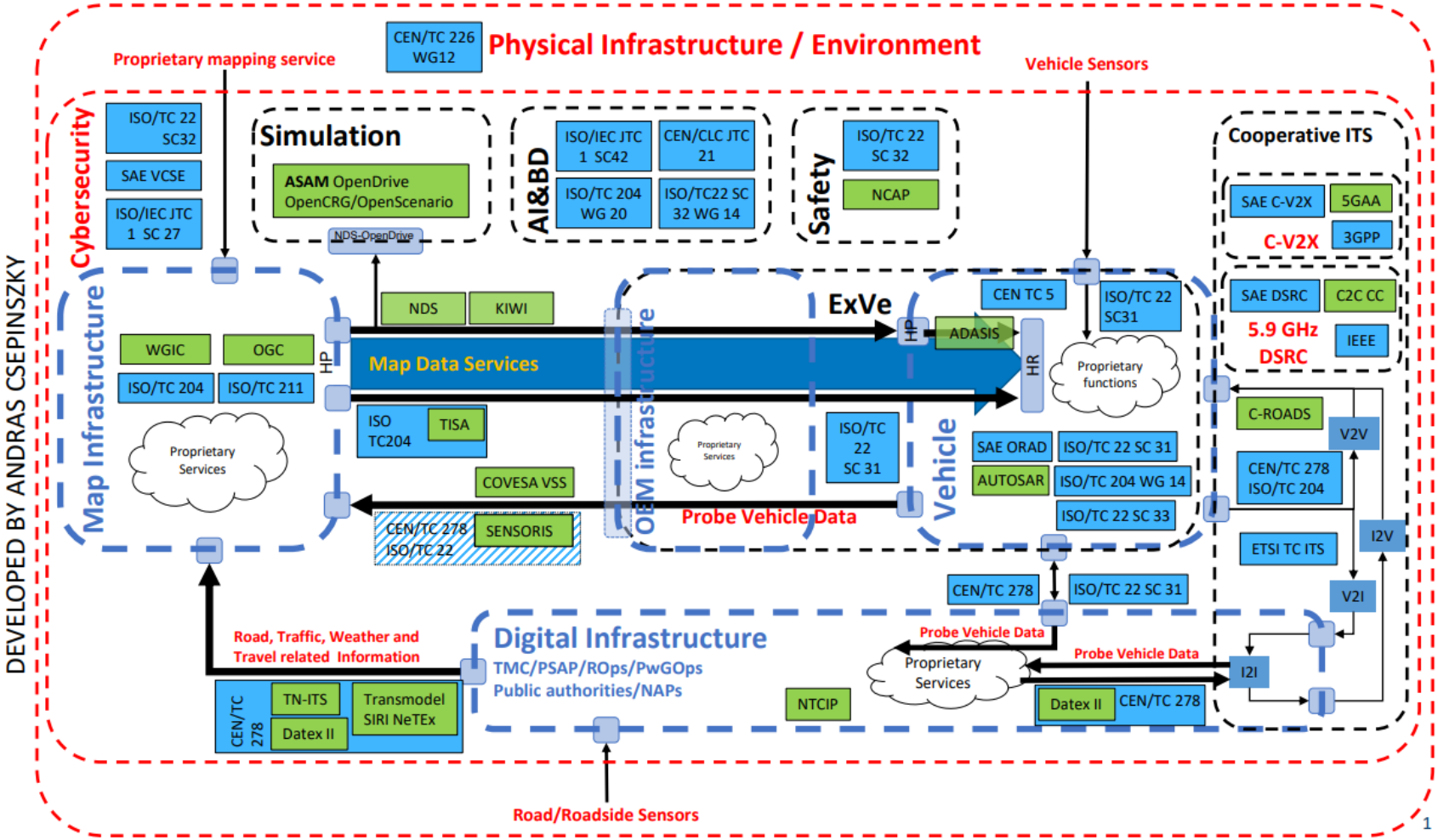
Data Provider	Product Offering Highlights
Jacobs/StreetLight[28]	• New product available in 2024 with individual trip data, like Wejo but with stricter privacy and cybersecurity controls, with trip ends "fuzzed" to ½ kilometer. • Asserts a 27% penetration on Texas roads. • StreetLight has smart phone data for walking and biking metrics and will have fuel consumption data from vehicles. • Data comes in sub-5 seconds, daily batch, and longer time horizons.
TomTom[29]	• Largest aggregator of vehicle data, using a variety of sources, capturing 1 in 4 cars on the road. • TomTom cannot reveal specific sources and must follow the GDPR as a European-based company; they provide tools to use anonymized data but do not provide raw data since it is aggregated from multiple sources. • Frequency: reported every minute, calculated every 30 seconds. • Penetration rate of data is higher than competitors and sets them apart.
INRIX[30]	• INRIX is primarily interested in the Global Positioning System (GPS) data coming from vehicles. • SafetyView is the GM/INRIX safety dashboard product for CV data using near-miss, hard-braking data. Working with GM on seatbelt usage in SafetyView.
HERE[31]	• Offerings include real-time synthesized traffic products with vehicle data as one of the inputs along with telemetry data, incident data, and other phone data. • "Road Alerts" product uses data from CVs regarding hard braking, windshield wipers, slippery roads, and other data combined. • State DOTs and automotive industry are HERE's largest clients. • Working closely with OEMs to capture changing roadside safety assets. • Will have a 2024 announcement about cloud-based platform with all products and services.
Mobilisights[32]	• New independent business unit established in January 2023 for exclusive access to Stellantis vehicle data. • Covers 13 million vehicles and is expected to grow to 34 million vehicles by 2030 across Europe and the Americas. • Data products cover full range of embedded data, including Advanced Driver Assistance System (ADAS). Customer sectors are currently insurance and fleet management.

[28] Interview with Laura Schewel, StreetLight/Jacobs. December 5, 2023, and StreetLight Transportation Solutions, Texas briefing deck, December 2023.
[29] Interview with Josh Gilman, Peter Van der Werf, TomTom. December 11, 2023, and TomTom trusted supplier proposal to TxDOT, November 3, 2023.
[30] Interview with Terri Johnson, INRIX. December 15, 2023.
[31] Interview with Daniel Abugov, Lauri Keller, Alan Yung. HERE, December 20, 2023.
[32] Mobilisights website, https://www.mobilisights.com

Data Provider	Product Offering Highlights
Flow Labs[33]	• Views their business model as a synthesized data provider, aggregator of multiple data sources, and solutions provider for transportation agencies. • Primary partnership with TomTom for CV data; claim 20-35% penetration, including rural. Also receives data from Airsage, Michelin, and NIRA dynamics. • Works closely with agencies on advanced applications and use cases, especially the efficacy of virtual sensors.
Moove.ai[34]	• Provides synthesized data with contextualized insights, classifying and normalizing CV data sets without "black box;" works with customers to build customized analytics. • Real-time or historical. • Uses multiple data sources; for CV, Wejo/StreetLight, Paccar and Volvo for truck data, and is in conversations with Stellantis (Mobilisights) and Ford about using their data.
GM[35]	• Data from 15 million vehicles with monthly upload; will not offer real-time insights until they see the revenue rationale. • Require customer to "opt-in" on data uses. • Offering its own platform – SafetyView with INRIX – providing insights and aggregated, anonymized data. • Certain datasets will not be offered on the market, including data from ADAS-equipped vehicles (400,000+ vehicles), e.g., hard braking, risky maneuvers; safety data will be available only through GM and its analytics products.

[33] Interview with Nathan Wade, Flow Labs, January 2, 2024.
[34] Interview with Rafay Khan, Justin Collins, and Pete Constello, Moove.ai, January 29, 2024.
[35] Interview with Harnit Anand and Sam Robinson, GM Future Roads, January 3, 2023.

Appendix D: European Standards for Vehicle and Infrastructure Ecosystem[36]



[36] SENSORIS website, https://sensoris.org/