165604-1

Strategic Research

PROGRAM

Data Ownership Issues in a Connected Car Environment: Implications for State and Local Agencies

November 2016

Texas A&M Transportation Institute

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
TTI/SRP/16/165604-1		

4. Title and Subtitle	5. Report Date
DATA OWNERSHIP ISSUES IN A CONNECTED CAR ENVIRONMENT: IMPLICATIONS FOR STATE AND LOCAL AGENCIES	November 2016
	6. Performing Organization Code

7. Author(s)	8. Performing Organization Report
Johanna Zmud, Melissa Tooley, and Matthew Miller	165604-1

9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
Texas A&M Transportation Institute College Station, Texas 77843-3135	
	11. Contract or Grant No. 10727

12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
Texas A&M Transportation Institute College Station, Texas 77843-3135	Technical Report: September 2015–August 2016
	14. Sponsoring Agency Code

15. Supplementary Notes

Supported by the State of Texas.
Project Title: Data Ownership and Use Issues with Connected Vehicle (CV) Applications

16. Abstract

This study examined the question of who owns the data emanating from connected cars, through the perspectives of three different stakeholder groups: automobile original equipment manufacturers, infrastructure facility owner-operators, and data aggregators. Connected cars have access to the Internet and a variety of sensors, so they are able to send and receive signals, sense the physical environment around them, and interact with other vehicles or entities. "Connected car data" is an umbrella term that refers to data generated by a car itself or in communication with other vehicles or infrastructure. It includes car data, infrastructure data, system performance data, and car occupant data. Stakeholders' perceptions of ownership are influenced by their data concerns, opportunities for monetization, and missions; therefore, in this study, stakeholder interests crossed data types. Six cross-cutting themes emerged from the research: where data are recorded matters, monetization for all, monetization impacts sharing, different roads to privacy, more is not better, and build a common lexicon. Connected car data represent an emerging data source with immense value for state and local transportation agencies, so these agencies should be proactive in determining the ways in which they can access those data, share them, and use them responsibly.

17. Key Words	18. Distribution Statement
Data Ownership, Data Privacy, Transportation Policy, Transportation Planning	No restrictions. This document is available to the public through NTIS: National Technical Information Service Alexandria, Virginia http://www.ntis.gov

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	50	

DATA OWNERSHIP ISSUES IN A CONNECTED CAR ENVIRONMENT: IMPLICATIONS FOR STATE AND LOCAL AGENCIES

by

Johanna Zmud
Senior Research Scientist
Texas A&M Transportation Institute

Melissa Tooley
Senior Research Engineer
Texas A&M Transportation Institute

Matthew Miller
Assistant Research Scientist
Texas A&M Transportation Institute

Report TTI/SRP/16/165604-1
Project 165604
Project Title: Data Ownership and Use Issues with Connected Vehicle (CV) Applications

November 2016

TEXAS A&M TRANSPORTATION INSTITUTE
College Station, Texas 77843-3135

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

ACKNOWLEDGMENTS

The authors recognize that the support for this research was provided by the State of Texas. The research team would like to thank Kevin Balke, senior research engineer at the Texas A&M Transportation Institute, for his review and comments on the report draft. We also acknowledge the contributions of the following individuals through their willingness to be interviewed as part of this research:

- David Agnew, Director of Advanced Engineering, Hyundai Mobis.
- Elizabeth Birriel, Intelligent Transportation Systems Program Manager, Florida Department of Transportation.
- Allison Cohen, Managing Counsel, Toyota Motor Sales.
- Dr. Mohammed Hadi, Associate Professor, Florida International University.
- Eric Hemphill, Director of System and Incident Management, North Texas Tollway Authority.
- Sarath Joshua, Senior Program Manager, ITS and Safety, Maricopa Association of Governments.
- William (Bill) King, Sr., Business Development Executive, Air Sage.
- Mark Kopko, Manager of Traveler Information and Advanced Vehicle Technology, Pennsylvania Department of Transportation.
- Melissa Lance, Operations Systems Manager, Operations Division, Virginia Department of Transportation.
- Greg Larson, Chief, Office of Traffic Operations and Research, California Department of Transportation.
- Bill Legg, State ITS Operations Engineer, Washington State Department of Transportation.
- Tim Reilly, Toll Operations, Central Texas Regional Mobility Authority.
- Dr. Anuj Sharma, Research Scientist, Iowa State University.
- Jeffrey Stefan, Public Policy, General Motors.
- Ted Trepanier, Executive Director, Public Sector, INRIX.

CONTENTS

EXECUTIVE SUMMARY

Connected car data are an emerging data source that could have immense value for state and local transportation agencies. Connected cars have access to the Internet and a variety of sensors, so they are able to send and receive signals, sense the physical environment around them, and interact with other vehicles or entities (PricewaterhouseCoopers [PwC], 2016). As used in this report, connected cars include but are not limited to the vehicle-to-vehicle (V2V) or vehicle-to-infrastructure (V2I) applications that define connected vehicles (CVs).

Connected car data are "generated by a car and its occupants either when the car is moving or stationary, by itself or in communication with other vehicles (V2V) or infrastructure (V2I)" (McKinsey, 2016). These data include car operations and diagnostics, telematics use, driver behavior, and car location and destinations. In a connected car environment, each car acts as a sender, receiver, and router to broadcast information. While the authors acknowledge that data from other types of vehicles, such as commercial fleets, could have value for agencies, this study examined issues surrounding car data.

Car data are also big data. Like all big data, car data can create information for the public good. Public agencies are considering how to use big data as a public utility—to use the information derived to make improvements in service provision and more effectively utilize public resources. Big data also have great monetary value. Some people consider big data, and the control and manipulation of those data, to be the new oil. At the World Economic Forum in 2009, a European Consumer Commissioner stated that "data is the new oil of the Internet and the new currency of the digital world" (Pentland, 2009). From a private-sector perspective, some of the largest Internet companies are built on the economics of monetizing big data. Big data also represent a new way for private individuals to look at their own value in the digital world—with data being an owned asset just like a house or a car.

Data have also been called the glue that binds numerous automotive innovations, from advanced driver assistance system (ADAS) technology to self-driving cars, since cars have become a crucial node in the incipient Internet of Things. As recognition of the volume and commercial value of data that cars can generate grows, an important question has become the following: *Who owns the data emanating from connected cars?*

Data Ownership

Ownership is straightforward when applied to a car or house since there is a formal transaction with written acknowledgment that makes ownership clear. However, when applied to data, ownership becomes complicated. For this reason, the data ownership question regarding connected car data has not been resolved to date. In a data context, there are many roles with which the notion of owner could be associated, from the data creator, to the data packager, to the data subject. Data ownership is also confounded by the recognition of its monetary value. Ownership of data defines who can control it and who can profit from it.

In addition, ownership implies a responsibility for information privacy. Information privacy is defined as the capability of individuals to "determine for themselves when, how and to what extent information about them is communicated to others" (Westin, 1967). Ownership

determines who can collect, process, use, and disseminate personal information. If the owner is not the individual, then privacy issues are raised. In the United States, there is no single comprehensive legislative framework for data privacy protection. There is also no single regulatory authority. Most states have enacted some form of privacy legislation. However, there is no regulatory framework that specifically addresses connected car data. There are guidelines developed by industry, such as those created by the Alliance of Automobile Manufacturers and Association of Global Automakers (2014), that are not legally binding but are part of self-regulatory efforts.

Ownership also implies a responsibility for data security. Data security refers to practices and processes that are in place to ensure data are not being used or accessed by unauthorized individuals or parties (Anderson, 2013). The proliferation of security breaches in recent years has led to expansive legal regulation at the state level. Virtually every state requires reasonable security procedures to protect data, a destruction policy, and notice of a security breach. There are no state laws specifically addressing car data security. There is also no current federal law regarding car data security, although the topic is receiving a great deal of attention by the legislative and executive branches of government.

Data ownership refers to both the possession of and responsibility for data emanating from connected cars. Ownership implies control as well as value creation. The control of data includes the ability to access, create, modify, package, derive benefit from, or sell the data. However, ownership also implies a broader responsibility for considering the consequences of how the data are used, particularly for how a particular use might impact data privacy or data security.

Data of Interest to Stakeholders

For this study, a typology of connected car data derived from various sources was used for information gathering among stakeholders (Hong et al., 2014; Walker et al, 2015; Telematics Task Force, 2014; Future of Privacy Forum, 2014). These data types were car-related data, infrastructure data, system performance data, and car occupant data. Stakeholders for this study were automobile original equipment manufacturers (OEMs), data aggregators, and infrastructure facility owner-operators. The interviews revealed that stakeholders' interests in data are not easily placed into neat categories (i.e., OEMS interested in only car-related data) but that their interests range broadly across the different data types. OEMs, data aggregators, and owner-operators all have legitimate business interests in connected car data. The answer to the question of who owns the data is influenced by data concerns, opportunities for monetization, and missions. Therefore, perceptions of data ownership also cross data types.

Cross-Cutting Themes across Stakeholder Perspectives

Interviews with stakeholders covered the following topics: data ownership, data functions and use, data privacy, data security, data retention policies, and data sharing. Six broad themes emerged from these interviews that provide insight into the different perspectives on data ownership: (a) where data are recorded matters; (b) monetization for all; (c) monetization impacts sharing; (d) different roads to data privacy; (e) more is not better; and (f) build a common lexicon.

Where Data Are Recorded Matters

Where data are recorded—inside a car or outside of a car—influences perceptions of data ownership. OEMs consider the owners of cars to be the owners of data generated by the cars, and they consider themselves to be stewards of that data with full access privileges once the owner has opted in to (or not opted out of) user agreements. Owner-operators typically use data that are recorded by devices outside of the cars (i.e., by roadside sensors). They perceive that they own the data from these sensors because the data are broadcast and therefore are public information. Data aggregators feel that they own the information derived from connected car data that they gather from various sources (including state departments of transportation [DOTs]), process and package, and then sell to buyers (including state DOTs). By processing the data, the data aggregators are creating a new information product and therefore are owners of that new product.

Monetization for All

All three stakeholder groups want to monetize connected car data. Data aggregators are openly in the business of data monetization because of how they generate revenue. OEMs do not consider themselves the owners of car data, but as the data stewards, they have access to use the data. In the past, the data were used primarily for internal purposes like improving safety, verifying vehicle quality, predicting customer behavior, and analyzing driver behavior. However, as cars become more connected, opportunities have arisen for car data monetization. Roadway owner-operators, who for the most part are public agencies, typically are mission-focused rather than revenue-focused. In the past, they have freely shared data coming from their sensors to fulfill this mission; however, in times of budget constraints, they are also beginning to consider the opportunities for monetizing the data.

Monetization Impacts Sharing

Perspectives on the issue of data sharing are influenced by the desire to monetize the data. OEMs and data aggregators as revenue-focused entities are less open to sharing data than are owner-operators. OEMs share data but only with entities that are involved in their internal processes. Data aggregators do not share data—they sell information derived from the data. To fulfill their missions, owner-operators typically share data on request. The benefits for the owner-operators are the opportunities to get useful information in return and to enable services or products that make them look more responsive, transparent, and effective in serving their constituencies.

Different Roads to Data Privacy

All three stakeholder groups are aware of the need to address privacy in data collection and use but take different approaches for doing so. The data aggregators are the most stringent in this regard, perhaps because they are the most embedded in the data business. Owner-operators acknowledged that maintaining privacy is an agency responsibility. However, there were less specifics provided on how data were to be de-identified. OEMs spoke less about data privacy issues, perhaps considering the industry's consumer privacy principles to be a complete solution to privacy protection.

More Is Not Better

More data are not better data because they take greater resources to use; information derived from data is best. While some data aggregators struggle with getting an adequate sample of data points along certain types of roads, generally the raw data that data aggregators provide to buyers are too much. State DOTs often struggle to format and use the data due to the volume and depth of the data and the data stream. They value the analysis that the aggregators perform to transform the data into useful information.

Build a Common Lexicon

Communication and collaboration among OEMs, data aggregators, and owner-operators might be improved if these entities had a common lexicon for discussing connected car data. A common language of key concepts would provide a basis for shared meaning and understanding and may facilitate resolving data ownership, data sharing, or other relevant data issues. Different labels or jargon were used by the three types of stakeholders (and even within a stakeholder group) to describe the different categories (or buckets) of connected car data of interest. The use of different labels for the relevant categories of data makes it challenging to derive consistent privacy protection, data sharing, or other important practices.

Conclusions and Recommendations

This study examined connected car ownership issues, and researchers drew conclusions regarding three key research questions.

Who Owns the Data?

Data ownership is complicated and nuanced. OEMs acknowledge that the owner or lessee of the car is the owner of the connected car data; however, they can access and control the data through user agreements. Data aggregators consider themselves to be the owners of the information that they sell, which is derived from the data. Owner-operators consider themselves to be the owners of the data collected by their sensors.

Under What Conditions Are Owners Willing to Share Data?

Desires to monetize data influence willingness to share connected car data. OEMs and data aggregators are revenue-focused entities and are only willing to share car-related data under certain terms and conditions. Owner-operators, on the other hand, typically share data on request; data sharing facilitates their missions, which for most involve moving people and goods safely and efficiently. Another benefit to the owner-operator of sharing data is to allow the data aggregator to process and reduce the data for them.

How Should Transportation Agencies Responsibly Use Connected Car Data?

Responsibly using connected car data relates to the concept of data stewardship. Data stewardship implies a fiduciary or trust relationship with individuals whose data are stored and managed by the steward. In this perspective, data ownership is less important. Stewardship relates to ensuring that personal information about individuals is protected and secure. In this

context, transportation agencies can take their lead from the OEMs in their enactment of the consumer privacy principles presented in this report. Principles geared toward state and local agencies can serve as a harmonized set of best practices for privacy and security risk mitigation.

Connected car data represent an emerging data source with immense value and the potential to vastly improve transportation planning, traffic management and operations, and safety. Faced with this opportunity, state and local agencies should be proactive in determining the ways in which they can have access to the data. They should begin to participate in, and even be leaders of, national, state, and local discussions and collaboration activities regarding how to responsibly collect, use, share, and disseminate connected car data.

CHAPTER 1. INTRODUCTION

This report presents findings from an exploration of the following question: Who owns the data emanating from connected cars? Connected cars have access to the Internet and a variety of sensors, so they are able to send and receive signals, sense the physical environment around them, and interact with other vehicles or entities (PwC, 2016). As used in this report, connected cars include but are not limited to the V2V or V2I applications that define CVs. While data from other types of vehicles, such as commercial fleets, could have value for agencies, this study examined issues surrounding car data.

Car data are an emerging data source that could have immense value for state and local transportation agencies for whom data collection for policy and planning purposes has strained limited budgets. Car data have been defined by McKinsey (2016) as "data generated by a car and its occupants either when the car is moving or stationary, by itself or in communication with other vehicles (V2V) or infrastructure (V2I)." These data include data about car operations and diagnostics, driver behavior, car location and destinations, and drivers themselves (e.g., their identities, contacts, schedules, destinations, and content choices; BC FIPA, 2015). In a connected car environment, each car acts as a sender, receiver, and router to broadcast information. Properly accessed, managed, and analyzed, such data could vastly improve transportation planning, traffic management and operations, and safety in cities and regions.

Connected car data are also big data. "Big data" is the term for a collection of large, quickly growing, and complex data sets that generally have some or all of the following features: digitally generated, passively produced, automatically collected, geographically or temporally trackable, and continuously analyzed (Hong et al., 2014). Car data represent massive amounts of information on people's movements, infrastructure conditions and performance, and environmental impacts. While applications of big data are somewhat more prevalent in the private sector, the potential benefits of integrating big data into transportation policy, planning, and operations are widely recognized. Such data can be combined with other data sources (e.g., demographics, transactions, or social networks) to create new knowledge about transportation activity and flows.

Objectives

For the reasons presented above, transportation agencies are interested in ownership-, sharing-, and use-related questions pertaining to connected car data. The research questions guiding this study were:

- *Who owns the data?* Data ownership principles are new and fluid. Technological advances are enabling new levels of ease for data creation, proliferation, and access; policy and regulation have not been able to keep up.
- *Under what conditions are owners willing to share car data with transportation agencies for societal benefit?* An important issue is how to divvy up the costs and benefits of sharing information; different players (i.e., individuals, private-sector firms, public agencies) do not always have the same goals.

- *How should transportation agencies responsibly use these data?* Agencies have a significant role to play in ensuring that data are properly collected and used while protecting confidentiality and privacy. Agencies also have a responsibility to collect accurate and timely data to meet the expectations of the public. Because of uncertainties in data ownership, among other issues, best practices have been slow to develop.

This study examined these questions, and this report provides recommendations for state and local transportation agencies in planning for the future.

Data of Interest

This study focused on the ownership and sharing of connected car data. Most likely, these data are owned by the private sector. However, some data are owned by public agencies. Moreover, there is a burgeoning social and policy movement to assign ownership of such data to the individual subjects, known as the New Deal on Data (Pentland, 2009). This new deal entails an exchange of data for monetary or service rewards and would require approval from the general public to use data collected from their digital interactions in the future.
To inform ownership and sharing questions, a typology for connected car data is presented in the following paragraphs, along with specific data elements comprising the general type. It is important to note that the typology has been inspired by several other categorization schemes, and that it does not solely represent CV data as defined by the United States Department of Transportation's (USDOT's) ITS Joint Program Office (Hong et al., 2014; Walker et al., 2015; Telematics Task Force, 2014; Future of Privacy Forum, 2014). The data presented here refer to a much broader set of data elements emanating from connected cars as defined in this report.

Car-Related Data

Moving and stationary cars are increasingly sources of data as connectivity and automation within cars increase. Advanced sensors, processors, enhanced driver interfaces, and other onboard or diagnostic units can record and deliver the data to centralized data centers through wireless networks. The data include basic car measures, car safety data, environmental probe data, diagnostics data, and inspections and emissions data. Specific data elements that are available, publicly or not, include:

- Car type and characteristics (length, width, bumper height).
- Time stamp.
- Speed and heading.
- Car acceleration and yaw rate.
- Turn signal status.
- Brake status.
- Stability control status.
- Driving wheel angle.
- Car steering.
- Tire pressure.
- Traction control state.
- Wiper status and run rate.

- Exterior lights.
- Global positioning system (GPS) status and vehicle position (longitude, latitude, elevation).
- Obstacle direction.
- Obstacle distance.
- Road friction.
- Current and average fuel consumption.
- Emissions data—measured emissions of specific cars comprised of exhaust pollutants including hydrocarbons, carbon monoxide, and nitrogen oxides.
- Air temperature and pressure.
- Weather information such as rainfall rate and solar radiation data.
- Electronic stability control.

Infrastructure Data

Data resulting from cars having connection to the physical infrastructure are an important product of a CV environment. Data are created via communication with dedicated short-range communication (DSRC)-capable roadside equipment (RSE) and more traditional intelligent transportation system (ITS) equipment distributed on and along the roadways, such as traffic detectors, environmental sensors, traffic signals, highway advisory radios, dynamic message signs, closed-circuit television (CCTV) cameras and video image processing systems, grade crossing warning systems, and freeway ramp metering systems. Such connections provide and exchange data elements related to roadway characteristics, road conditions, intersection status, and field equipment status. Data elements include:

- Roadway characteristics.
    - Friction coefficient.
    - Road geometry and markings.
- Road conditions.
    - Surface temperature.
    - Subsurface temperature.
    - Moisture.
    - Icing.
    - Treatment status.
- Road surface weather condition.
    - Air temperature.
    - Wind speed.
    - Precipitation.
    - Visibility.
- Intersection status.
    - Current operating status of the traffic signal equipment.
    - Signal phase and timing.
    - Intersection geometry.
- Advisory information based on current conditions.
    - Dynamic message signs.
    - Variable speed limit signs.

- Dynamic lane signs or control devices.
- Parking facility locations.
- Parking spaces available.

System Performance Data

Connected cars can report current position, speed, heading, and snapshots of recent events including route information, starts and stops, speed changes, and other information that can be used to estimate traffic conditions and support transportation planning and asset management. The information can be derived from roadside sensors or cameras and delivered wirelessly to traffic management centers (TMCs) or other ITS-related facilities for storage, integration, analysis, and reporting. System performance measures derived from connected cars include:

- Traffic speed.
- Travel times.
- Volumes.
- Occupancy.
- Density.
- Origin and destination data.
- Incident status.
- Video images.

Car Occupant Data

In a connected environment, travelers continuously interact with cars (e.g., via in-car route optimization and traffic information apps) to get real-time routing information and avoid congestion, or with the more advanced apps, to specify transportation parameters unique to their individual needs. Travelers can receive route planning information at fixed locations such as in their homes, at their place of work, and at mobile locations using personal portable devices and car-based devices. Travel information generated through these interactions includes:

- Trip origin and destination, and timing.
- Travelers' personal data such as address, trip records, and profile data.
- Service information (toll payment, parking reservations and fees, ridesharing options).
- Occupancy.
- Vehicle miles traveled.

The data elements across all types are wide ranging and the categories are overlapping, thus creating complex data ownership questions. The laws and regulations pertaining to data ownership are new, evolving, and have not kept pace with technology developments. Because of this, this study took a broad view of significant policy issues to provide context before delving into specifics on car data ownership and responsible sharing and usage practices.

Report Organization

After this introduction, the report is organized into the following sections:

- Chapter 2: Data Ownership and Data Privacy.
- Chapter 3: Stakeholder Perspectives on Data Ownership Issues.
- Chapter 4: Cross-Cutting Themes among Stakeholder Perspectives.
- Chapter 5: Conclusions and Recommendations.

The appendix includes a glossary of terms used in this report.

CHAPTER 2. DATA OWNERSHIP AND DATA PRIVACY

What Is Data Ownership?

Ownership pertains to having legal title or the right to possess something. When applied to someone's house or car, the concept is straightforward. However, when applied to data, the ownership question becomes messy. Ownership of data is tantamount to control; determining ownership defines who can collect, process, use, and disseminate data. Ownership also implies who can profit from what is owned. Yet, resolving the question of who owns a particular type of data is complex. The complexity of data ownership issues is made clearer by Loshin's (2002) listing of roles that have been used to claim data ownership:

- Compiler: selects and compiles information from different information sources.
- Consumer: uses the data.
- Creator: generates data.
- Decoder: unlocks data locked inside particular encoded formats.
- Funder: commissions the data creation.
- Packager: collects data and adds value for a particular use.
- Purchaser/licenser: buys or licenses data.
- Subject: is the subject of the data.

Data ownership is also complicated by the recognition of its monetary value as well as by awareness of privacy and security issues. The essence of ownership lies in the control of information as an asset. That control includes not just the ability to access, create, modify, package, sell, or remove data but also the right to assign these privileges to others. Since ownership is connected to economic value, an inappropriately low estimation of value can be as problematic as an overestimation of value. For example, having a low cost of obtaining data or low value estimation may encourage a lax approach to security by holders of the data, whereas an overly high value estimation can lead to data hoarding (Rosner, 2014).

The idea of owning information is relatively new. The earliest copyright law, which granted to the creator of a work exclusive rights to its duplication and distribution, first appeared in the early 18th century (Ng, 2011). It would still be hundreds of years, however, before the concept of data as currently understood even began to develop. The term "data," defined by Dictionary.com as being "transmittable and storable computer information," was first recorded in 1946. However, a literature search of the term "data ownership" indicated the earliest use was not until 1963, addressing ownership of patents in defense procurement (Leathem, 1963). More mainstream references go back three decades to the fields of health information management and research ethics (Grandison and Mohammed, 2013; Seashore, 1978).

Within the health information realm, concerns about potential risks to individual privacy eventually led to federal legal requirements on the use or disclosure of health information for research and other purposes (i.e., the Health Insurance Portability and Accountability Act of 1996 [HIPAA]). Prior to the 2000s, health and medical data were generally stored as physical records, and sovereign responsibility could be assigned. The classic rule was that the physical

record belonged to the health provider and the intangible data belonged to the patient (Demster, 2012).

In terms of research ethics, the pertinent issues were related to citing or sourcing, minimizing selective reporting, and mitigating the disclosure of confidential information obtained for research. Data were considered difficult and costly to collect, so data hoarding had become a significant issue. The scientific community was concerned that data hoarding led to several negative consequences including preventing other researchers from checking the accuracy of any results the original researchers might have published or from using research, clinical trial, or experimentation results to accelerate their own research. Because historically the collector of the information was assumed to be its owner and the entity responsible for it, the owner could hoard the data. For this reason, starting in about the 1990s, the concept of ownership was sometimes replaced by stewardship because it implied broader responsibility than just collection, including access, sharing, and reporting (Chisholm, 2011).

Data Stewardship

Since about 2000, the concept of data stewardship has received much attention in the transportation data world, especially among state DOTs in preparing data management plans. The concept of data stewardship is rooted in the practice of data governance and is meant to reflect the values of fair information practice as defined by the Federal Trade Commission (FTC; Diamond et al., 2009). The following fair information practices were developed to address online privacy (FTC, 1998):

- Notice: Provide clear and conspicuous notice of what information is collected, how it is collected, how it is used, and whether it is shared with other entities.
- Choice: Offer choices as to how personal identifying information is used beyond the use for which the information was provided (e.g., to consummate a transaction).
- Access: Provide reasonable access to the information that is collected about an individual, including a reasonable opportunity to review information and to correct inaccuracies or delete information.
- Security: Take reasonable steps to protect the security of the information collected from consumers.

Data stewardship denotes an approach to the management of data, particularly data that can identify individuals. Data stewardship refers to the design and application of data management principles covering collection, storage, retention, aggregation, de-identification, and procedures for data access, sharing, and use. The concept of a data steward is intended to convey a fiduciary (or trust) relationship with the individuals whose data are stored and managed by the steward.

However, these easily defined borders of ownership and stewardship are changing due to technological advances like advanced computing, the Internet, and mobile devices that have enabled new markets for data creation, use, compilation, packaging, etc. For example, how does one provide notice to individuals whose data have been collected via roadside Bluetooth® sensors?

Data have their own intrinsic value as well as added value as a byproduct of information processing. How much are data worth? According to the financial valuations of companies that were built on data, like Facebook, Uber, or Twitter, there is a gap between the actual market valuation and the perceived market valuation. Uber is now worth about $18.2 billion. United Continental Holdings with tangible physical assets, on the other hand, is worth $17.9 billion (Huet, 2014). The implications for data ownership in this new environment are immense. The value chain of who owns the data gets more complicated when speaking of big data, which involves aggregating data from many sources, analyzing it, storing it, repurposing it, and reusing it. As data are aggregated from ever greater numbers of sources, the issues of data provenance, ownership, and stewardship become murkier. This is certainly true of data emanating from personal vehicles.

Ownership of Event Data Recorder Data

The connected-car-related data ownership debate began to appear in the 1990s pertaining to electronic data recorders (EDRs), which are electronic sensors installed in motor vehicles (Koch, 2006). EDRs, also known as black boxes or sensing and diagnostic modules, store information produced immediately before and during an accident, such as date, time, vehicle and engine speed, steering angle, throttle position, braking status, force of impact, seatbelt status, and air bag deployment. EDRs have evolved from analog signal processing and recording in the 1970s to the electronic sensors of today that also provide diagnostic information. These data have been used to improve vehicle safety as well as investigate the causes of crashes.

EDRs track specific data elements as prescribed by National Highway Traffic Safety Administration (NHTSA) regulations. None of these data elements are personal information, but when combined with other technologies, such as onboard navigation systems or mapping apps, EDR data could be used to personally identify an individual (Canis and Peterman, 2014). Because of this, many individuals and advocacy groups have raised privacy concerns about law enforcement and other third-party access to the data.

The privacy aspects of EDRs and the ownership of the data they generate have been a subject of Congressional legislation since at least 2004 (Canis and Peterman, 2014). Since 2006, NHTSA has required that consumers be informed when an automaker has installed an EDR in a vehicle, and most car manufacturers currently install these devices in new cars. It was not until 2015, however, with the Driver Privacy Act, that the subject of ownership was specifically addressed. This legislation provides that all car manufacturers must have an EDR and must collect specific information. It also stipulates that the EDR information belongs to the owner or, in the case of a leased vehicle, the lessee of the vehicle in which the event data recorder is installed. EDR data are accessed via specialized software and are shared only with the consent of the vehicle owner or lessee. However, there are exceptions. These limited exceptions include (a) as authorized by a court or judicial or administrative authority, subject to the standards for admission into evidence; (b) to carry out investigations or inspections authorized by federal law; (c) to facilitate medical care in response to a car accident; and (d) for traffic safety research, as long as the personal information of the owner/lessee is not disclosed. Seventeen states have also enacted statutes relating to EDRs and privacy with a similar provision about who owns the EDR data (National Conference of State Legislatures, 2015). Various state statutes refer to EDR data as property, with the same ownership rights as tangible property (Pomerantz and Aisen, 2015).

Data Privacy

Data privacy is related to data ownership because determining ownership defines who can collect, process, use, and disseminate *personal* data. The EDR debate addressed data ownership and, in doing so, addressed data privacy as well. In the case of the EDR, data belong to the car owner or lessee, who controls their disposition. However, reflective of the current legal environment, the regulation is specific to the EDR and not to other types of car data. Thus, the legal environment surrounding the issue of data privacy and data ownership pertaining to other categories of car data is still very fluid.

Information privacy is defined as the capability of individuals to "determine for themselves when, how, and to what extent information about them is communicated to others" (Westin, 1967). This is particularly relevant to privacy of personally identifiable information (PII). PII is any data that could potentially identify a specific individual, including any information that could be used to distinguish one person from another or that could be used for de-anonymizing anonymous data. There is no one list of what constitutes PII. Some examples of information that could identify an individual include name, address, date and place of birth, and biometric data. A single piece of data can be PII, such as a social security number. Likewise, multiple pieces of data when merged can be PII, even when the individual pieces would not be. As an example, a license plate number does not identify a specific person; rather, it identifies a vehicle. However, the license plate number may be linked or associated with an identifiable person through a linkage with other information about the individual, such as date of birth, gender, and zip code. As a result, while license plate numbers are not inherently PII, their common affiliations and linkages with individuals constitute an increased risk to privacy.

Once information is associated with a specific individual, it becomes PII. Such associations can happen through consumers' smartphones, their use of in-car telematics systems, or some V2V or V2I applications. Radical transformation of computing, mobile, sensor, global positioning, and database technologies have weakened traditional means of protecting individuals' privacy, leading to increasing risks associated with misuse of PII. Treatment of PII is distinct from other types of data because it needs to be not only protected but also collected, maintained, and disseminated in accordance with the fair information practices. As such, data privacy has the potential to become a contentious issue for transportation agencies, unless harmonized policies and practices for data privacy risk mitigation are identified, documented, and applied. Balancing agencies' needs for using such data with individuals' concerns about their data privacy is a complicated challenge.

For example, according to a 2015 survey by the Pew Research Center, a majority of Americans believe it is important—often "very important"—that they be able to maintain privacy and confidentiality in commonplace activities of their lives (Madden and Rainie, 2015). Most strikingly, these views are especially pronounced when it comes to knowing what information about them is being collected and who is collecting it. These feelings also extend to a desire to maintain privacy when moving around in public. Survey results from early 2015 show that 63 percent felt it was important to be able to "go around in public without always being identified." All adults, regardless of age or gender, express comparable views.

Data Privacy Regulation

Data privacy regulation is typically associated with data privacy protection. In the United States, there is no single comprehensive legislative framework for data privacy protection. There is also no single regulatory authority. This contrasts with the European context. Privacy in the European Union (EU) is protected as a fundamental right under the European Union's Charter of Fundamental Rights, essentially making data privacy akin to a constitutional right for EU citizens. In contrast, the United States has a patchwork of federal and state laws and regulations that overlap, dovetail, and may even contradict one another (Jolly, 2014).

At the federal level, different privacy requirements apply to different industry sectors and data processing activities. The laws are often narrowly tailored and address specific data uses and users, such as EDR data. For those entities not subject to industry-specific regulatory authority (i.e., Gramm-Leach-Bliley Act for financial services, or HIPAA for health information), the FTC is the primary federal privacy regulator (Sotto and Simpson, 2014). It uses Section 5 of the FTC Act, which is a general consumer protection law that prohibits "unfair or deceptive acts or practices in or affecting commerce" to bring privacy enforcement actions. Yet, in general, FTC enforcement has been mostly procedural, focusing on companies' notice and consent actions, such as ensuring that online companies have privacy policies, that the policies are not hidden in obscure places on company websites, etc.

There are many laws at the state level that regulate the collection and use of personal data, and the number grows each year. Some federal privacy laws preempt state privacy laws on the same topic (Jolly, 2014). For example, the federal law regulating commercial email and the sharing of email addresses preempts most state laws regulating the same activities. Conversely, there are many federal privacy laws that do not preempt state laws, which means that a company or agency can find itself in the position of trying to comply with federal and state privacy laws that regulate differently the same types of data or types of activity.

Most states have enacted some form of privacy legislation. However, California leads the way in the privacy arena, having enacted multiple privacy laws, some of which have far-reaching effects at a national level. Unlike many federal privacy laws in the United States, California's privacy laws resemble the European approach to privacy protection. However, even in California, there is no regulatory framework that specifically addresses connected car data. Instead, there are many guidelines developed by governmental agencies and industry groups that are not legally enforceable but are part of self-regulatory efforts that are considered best practices in the context of connected cars.

Self-Regulatory Efforts

In describing its vision of the V2V system, NHTSA emphasized a few key points related to CV systems and data privacy. In deploying V2V, NHTSA affirmed that the technology *would not* (a) collect or store any data on individuals or individual vehicles; (b) track specific individual vehicles through space and time; (c) collect information linked to individuals, including PII or personal communications; or (d) require users to provide PII about themselves or their vehicle when enrolling in a V2V program (Stanley and Wagner, 2015).

In a similar manner, the automotive industry developed privacy principles in 2014 for vehicle technologies and services largely in response to data privacy and security concerns raised by Senator Markey (D-MA) and others in Congress about the increasing connectivity and automation of automobile technology (Markey, 2015). The Markey (2015) report concluded that "there is a clear lack of appropriate security measures to protect drivers against hackers who may be able to take control of a vehicle or against those who may wish to collect and use personal driver information." It called on NHTSA, in collaboration with the FTC, to promulgate new standards to protect the security and privacy of drivers in connected vehicles. The auto industry privacy principles, effective for new vehicles manufactured no later than model year 2017, represent a unified response to such concerns (Alliance of Automobile Manufacturers and Association of Global Automakers, 2014). Overall, the privacy principles require clear and prominent notices about the collection of information, the purposes for which it is collected, and the types of entities with which the information is shared. As discussed later in this report, some view these principles as a preemptive move by the auto industry to address data ownership questions.

Data Security

As noted above, data ownership implies a responsibility not only for data privacy but also for data security. Data security refers to practices and processes that are in place to ensure data are not being used or accessed by unauthorized individuals or parties (Anderson, 2013). It differs from data privacy in that data security refers to systems and data privacy refers to individuals. The more data are gathered and stored in databases, cloud-based services, or otherwise, the greater the potential for harms to data subjects from security breaches. According to Wikipedia, a data breach is a security incident in which sensitive, protected, or confidential data are copied, transmitted, viewed, stolen, or used by an individual unauthorized to do so.

Globally, 1 billion data records were compromised in 2014 (Singleton, 2015). In a 2015 PricewaterhouseCoopers survey on cyber threats, 76 percent of corporate executives said they were more concerned about cyber threats than in the previous 12 months, and 79 percent had detected a security breach in the past year (PwC, 2015). Since vehicles and the transportation networks they operate on have become increasingly digital, with a wide data flowing across systems, the threat of harm is moving from data breaches to interrupting network flows and threatening critical infrastructure. Cybersecurity is becoming a transportation policy and planning issue as well as a technical one.

In this current climate, trust that the collector of the information will keep data secure becomes very important. However, according to Pew Research Center surveys, Americans have little confidence that their data will remain private and secure; just 6 percent of adults say they are very confident that government agencies can keep their records private and secure, while another 25 percent say they are somewhat confident (Madden and Rainie, 2015). This means that over 50 percent do not feel confident that government agencies can keep their records safe and secure.

The proliferation of security breaches in recent years has led to an expansion of the patchwork system of security laws, regulations, and guidelines that is becoming one of the fastest growing areas of legal regulation (Anderson, 2013). To date, data security has been primarily addressed at the state versus the federal level. Virtually every state requires persons or organizations

possessing personal information of their residents to notify the state if there is a breach of security. Most states also require reasonable security procedures and practices to protect such information and require either a destruction policy or secure means of disposal for personal information. There are no state laws specifically addressing connected car data security.

While there is no current federal law regarding car data security, the topic is receiving a great deal of attention by the legislative and executive branches of national government. A report issued by U.S. Senator Markey's (D-MA) office found that nearly all vehicles on the road are vulnerable to hacking via at least one, and often many, wireless entry points (i.e., ways that vehicle electronics can be accessed remotely). Common wireless access points include Bluetooth, keyless entry, remote start, navigation, Wi-Fi, cellular communications, radio, and anti-theft systems and features. When asked via survey, most automobile manufacturers did not describe effective means of protecting customer data held in the automaker's data centers or those of its service providers (Markey, 2015).

In response to Markey's report and high-profile breaches, Congress introduced the Consumer Privacy and Protection Act (S. 1158) to establish a federal security breach notification law and provide protection for many types of data, including social security numbers, financial account information, online usernames and passwords, unique biometric data (including fingerprints), information about a person's physical and mental health, information about a person's geolocation, and access to private digital photographs and videos. The bill would preempt weaker state laws while leaving stronger state privacy laws in place. As of this report, there has been no Congressional action on the measure.

In summary, data ownership refers to both the possession of and responsibility for data emanating from connected cars. Ownership implies control as well as value creation. The control of data includes the ability to access, create, modify, package, derive benefit from, or sell the data. However, ownership also implies a broader responsibility—data stewardship—where the user must consider the consequences of how the data are used, particularly how a particular use might impact data privacy or data security.

CHAPTER 3. STAKEHOLDER PERSPECTIVES ON DATA OWNERSHIP ISSUES

Connected cars can generate large volumes of data, including data on engine performance, location, and driver behavior. While many stakeholders can benefit from connecting cars, there are three stakeholder groups with key interests in data ownership issues: automobile original equipment manufacturers, data aggregators, and owner-operators of transportation facilities. Interviews with key informants, supplemented by literature searches, were used to gather and document their varying perspectives on data ownership, data use, data privacy and security, and data sharing.

Automotive Original Equipment Manufacturers

General Motors (GM) was a pioneer of in-car connectivity in the 1990s with OnStar. Now, all major automakers are transforming their fleets into connected cars. Such cars have various systems and components that are controlled by in-vehicle computers, which communicate with each other and record data. Connected car technology also enables cars to communicate with the automaker, third-party suppliers to the automaker, other vehicles, and roadway infrastructure via DSRC, cellular, or wireless connectivity. Connected car penetration is at 9 percent in 2016 and is expected to hit 28 percent in 2020 (Statista, 2015). There is a spectrum of sophistication in connected car products, with some OEMs more mature than others (DePorre, 2016).

The kinds of systems that can be connected are music/audio; smartphone apps; navigation; automotive system diagnostics; Bluetooth; roadside assistance; voice command, hands-free controls; contextual help/offers; parking apps; automobile diagnostics; 4G Wi-Fi hotspots; traffic/safety/collision warnings; and text alerts to friends, family, and colleagues (Mohammed, 2015). All of these potential connections produce a lot of data in which OEMS have a significant interest. Car-specific data include:

- Vehicle identification information: VIN number, make, model, year.
- Diagnostic codes: odometer reading, oil life remaining, tire pressure, fuel economy.
- Information on when cars are refueled or recharged.
- Information indicating whether the car has been broken into or stolen.
- Information about preinstalled apps.
- Information about when the car ignition has been turned on or off.
- Information about collisions involving the car, like the direction from which impact happened and which air bags deployed.

OEM interest in connected car data goes beyond car-specific data to include data on driving behavior and customers' use of in-car technologies and services. As one OEM representative said, "One really cannot understand what a transmission diagnostic code might mean unless one knows how the car is driven." Driving behavior data include:

- Direction of travel.
- Time of travel.
- Speed.

- Average RPM.
- Cruise control status.
- Seat belt status.

Use of services data include:

- Customer chosen points of interest for navigation.
- Search content.
- Date, time, and duration of in-car calls via telematics.

Data Ownership

In interviews, automakers made important distinctions between data ownership, data access, and data stewardship. Historically, automakers have considered the owner of the car to also be the owner of the data generated by the car. As one automaker said, "Customers have control. We get consent to collect the data. If they opt out, we don't use." For this reason, many OEMs have enacted privacy principles and policies to provide transparency to their data collection, use, and sharing practices in order to not discourage customers from opting in.

However, the question of data ownership is becoming more complex and nuanced as automakers are realizing that data not only have intrinsic value, such as for analyzing vehicle health data or improving product quality, but also have added value as a byproduct of information processing or data integration and mining. An example of a value-added feature might be the creation of a smart driver feature to encourage better driving that could be monetized in association with the insurance industry.

With connectivity, there is greater perceived value and therefore greater interest among OEMs in clarifying their data access and use rights. One OEM representative thought that U.S. OEMs might be positioning themselves to be the sole owner of the data because of the potential for monetization. He thought that if the OEM could produce its own automated vehicle (AV) or ADAS, it had a better chance of owning the data. Google was cited as an example of an OEM that wants to keep ownership by producing its own self-driving cars, of which it would be the fleet owner. However, there was also recognition that creating value requires application of advanced computing and big data analytics, which takes resource investment. Thus, the OEMs are carefully considering their options.

Automakers have direct access to much of the car data of interest because their systems are collecting those data. While some OEMs, particularly technology firms, might be positioning themselves to be owners of connected car data, most conventional automakers consider themselves to be the *stewards* of the data. Data stewards handle data governance responsibilities, such as crafting appropriate data use policies to protect privacy and security, tracking data-related legislation, and monitoring internal compliance with the data use policies.

In accordance with this role as data stewards, in 2014, 19 OEMs signed on to a set of consumer privacy protection principles for vehicle technologies and services under the auspices of the Alliance of Automobile Manufacturers and Association of Global Automakers. The principles refer to the fact that the automotive industry is developing innovative technologies and services

that are based on information obtained from a variety of vehicle systems and that involve the collection of information about a vehicle's location or a driver's use of a vehicle. Consumer trust in terms of opting in to this data collection is essential for the ongoing development. An OEM participant suggested that this might be a preemptive strike by OEMs to maneuver to be in control of monetization: "Once people click the user agreement, they have given up the data."

The principles are general and broad:

- Transparency: Provide ready access to clear, meaningful notices about the collection, use and sharing of information.
- Choice: Offer certain choices regarding the collection, use, and sharing of information.
- Respect for context: Use and share information in ways that are consistent with the context in which it was collected.
- Data minimization, de-identification, and retention: Collect information only as it is needed for legitimate business purposes.
- Data security: Implement reasonable measures to protect information against loss and unauthorized access or use.
- Integrity and access: Implement reasonable measures to maintain accuracy of information and give customers reasonable means to review and correct personal subscription information.
- Accountability: Take reasonable steps to ensure that they and other entities that receive information adhere to the principles.

This broad specification provides latitude and discretion to OEMs in their implementation. OEMs' privacy policies and user agreements are written from their business perspective. An ongoing collection and analysis of data enables them to improve their product and develop a new suite of customized product offerings. This helps to establish an ongoing customer relationship as well as incremental revenue streams over the life of the car. An OEM mentioned that as the steward of the data, the customers were entrusting their data to the OEM, and the OEM had the rights to use that data in many broad areas.

Data Functions and Use

Interviews with OEMs identified several purposes for which data are being used. These included verifying vehicle quality, developing new features, analyzing vehicle trends, improving safety, and preventing fraud or misuse of systems. With the advent of higher automation in vehicles, it was mentioned that connected car data will be used to "teach cars how to drive, leading to more effective and safer ADAS systems and autonomous vehicles." If one knows how the system is performing out in the field, that information can be used for updating the design.

In interviews, OEMs made a distinction between their right to use data that are tied to the individual and data that are de-identified. If data were anonymized or aggregated, then OEMs felt they could do with the data what they want—use it for any purpose or share it with any third party.

While automakers use the data for their own purposes, they are quick to point out that they also seek to develop services or features that customers will find valuable and improve the customer

experience, such as coaching on techniques that optimize fuel efficiency or highlighting routes that minimize time in traffic congestion.

Data Privacy, Security, and Retention Policies

Automakers' privacy and protection policies attempt to explain to customers the OEM's data collection, use, and handling practices specific to connected car data. These are typically housed on a privacy portal on the OEM website. In at least one interview, the OEM said that its portal was in direct response to signing on to the consumer privacy principles mentioned previously. Most policies cover what types of information are collected, how the data are used, what information is shared and with whom, and data retention and security practices. Three privacy statements were analyzed for similarities and differences. While the types of information collected were consistent, the retention practices differed.

- Toyota collects personal contact information, location data, data on real-time status of the car (e.g., status of powered features, oil life, and distance-to-empty), diagnostic trouble codes and related data from the car's onboard diagnostic system, and multimedia system screen operation log information. A customer needs to proactively opt out of the data collection. Most data are retained for seven years after expiration of subscription to connected services. Vehicle health and multimedia screen operation log information are retained not longer than 20 years to support ongoing research. Only the latest real-time status data are retained; the old data are purged. (See Toyota Connected Vehicle Services Privacy and Protection Notice at http://www.toyota.com/privacyvts/images/doc/privacy-portal.pdf.)
- GM OnStar collects account information (including contact and billing information and use of OnStar services and websites), car-related information, and driving information. OnStar states that it can use the information for any purpose. Customers opt out of data collection by declining to subscribe to OnStar. Data are retained for "if we need it." (See OnStar Privacy Statement at https://www.onstar.com/us/en/footer-links/privacy-policy.html?source=ct.)
- Hyundai collects subscription, billing, and registration information; information about the car (e.g., model, year, VIN, engine type, service data); eco-related driving performance data per trip; information about use of technologies and services (e.g., customer chosen points of interest for navigation, search content); information about use of the car (including direction and time of travel, odometer reading, refuel indication, EV battery status); and geolocation and driver behavior information (such as speed, average RPM, cruise control use, and seat belt status). Customers cannot opt out of collection of diagnostic information. Customers can decline to subscribe to telematics services, but if they subscribe, they cannot opt out of data collection. Data retention practices are not covered. (See Hyundai Vehicle Owner Privacy Policy at https://www.hyundaiusa.com/owner-privacy-policy.aspx.)

Data Sharing

The entities with which OEMs share data are typically listed in their privacy policies and are standard across the OEMs. OEMs do not share data with law enforcement unless required to by law or legal process. OEMs share data with:

- Emergency responders and roadside assistance providers.
- Internal credit corporations.
- Internal parent company.
- Their own third-party service providers.
- Their own dealers.

These are entities that are involved in the OEM's internal business processes. The OEMs have always been an industry in which internal information is proprietary for market advantage. That said, one OEM representative raised the possibility that NHTSA might have a role in regulating the sharing of data among automakers working on higher levels of automated vehicles for the public good. He advised that opening safety data among the automakers would help improve active safety systems: "The data sharing could save lives." He thought this would be a good use of government regulation.

More recently, NHTSA, in laying out guidance to companies developing the next generation of cars, recommended that companies commit to sharing data. The guidance stipulates that during testing and when fully autonomous cars begin driving, "Data generated from these activities should be shared in a way that allows government, industry and the public to increase their learning and understanding as technology evolves but protects legitimate privacy and competitive interests" (U.S. Department of Transportation, 2016). It is unclear how the auto industry will respond.

When asked if they would be willing to share data with transportation agencies, there were varied responses from the OEMs. There was the sense that an additional consent might be needed to share the data, but that the owner of the car might be willing to share the data with public agencies. Thus, this could be the focus of a study at a particular point.

The notion of continuous sharing of data was seen as more challenging because "data is very valuable." One OEM stated that it might be willing to share for a cost since it is interested in monetizing the data. However, it would be less interested in sharing information that has commercial value, such as driver behavior data. Such data have a bigger impact on OEMs' internal systems and have a higher value for monetization.

Additional Comments of Interest

One OEM started out the interview by stating, "Data ownership is a legal question. We don't think in terms of ownership." Having legitimized access to the data appeared to be more important to OEMs.

OEMs are looking to monetize data. A participant indicated that it is "a connected big data company. Our mission is to determine how to monetize the data." Another said that they are analyzing the data for themselves, for the greater good, as well as the customer. Another specifically mentioned that the OEM "wants to use the data for the greater good. We want to be part of the story that mitigates congestion." Another mentioned that for a customer, "Opting in to data collection could also be looked at as a contribution back to society by an individual if the ADAS systems become much better and traffic safety is significantly enhanced."

Data Aggregators

Car-based and system performance data are gathered and shared within a data ecosystem that includes companies gathering the data from all available sources about all types of vehicles, and then formatting and innovating new uses by various customers, including public agencies. These companies are known as data aggregators. Data aggregators are typically private companies that utilize a particular part or series of components of the wider vehicle-based data ecosystem to gather car-based and systems performance data. These data aggregator companies assemble data streams from a variety of sources, including:

- GPS devices.
- Bluetooth-enabled devices.
- Wi-Fi-enabled devices.
- Onboard unit transmitters from original automotive manufacturers.
- Smartphone applications.

Data aggregators then clean the data of personal identifiable information and format those data as per the needs of the customer. This section describes findings from outreach with universities, state DOTs, and data aggregators involved in the use of these aggregated data sources.

Data Ownership

Data aggregators advised that they own the connected car data that they gather from various sources. As one company representative indicated, "Much of the data used these days is not owned by the end user—it's licensed by the end user." This company and several others advised that as part of their business model, they sell access to the data to various customers via data licensing agreements with a variety of conditions on use.

Smartphone application-based data are obtained and effectively owned by the company responsible for developing the application. Examples of smartphone application-based data sources are companies like Waze, Metropia, Google Maps, and INRIX. The most common individual source user agreement indicates that the user agrees to share his or her data for use by the mobile application developer as it sees fit, as long as privacy and security are maintained.

Data aggregators may be from either a public or a private orientation, so their perceptions of data aggregation and ownership have subtle differences. Private data aggregators generally advised that they rely on a chain of legally binding agreements to access source data and to package and resell it to buyers for limited uses and functions. However, for one quasi-public toll authority, the perception is that connected car data may be collected, owned, and put to use by anyone capable of gathering the data, as long as the data are scrubbed of personally identifiable information. The representative indicated, "It's whoever can grab the data—we are tracking technological developments in vehicle-to-vehicle and vehicle-to-infrastructure technologies, and found that…there are sensors transmitting this information and to get data back from these sensors you have to have some network in place. So we have built a fiber backbone down the middle of our roadway and put a bunch of laterals off to infrastructure capable of collecting and transmitting this data."

This toll authority indicated that it collects the car-based data along its toll roads through several sources, including:

- Radio frequency identification (RFID) tags installed in cars for toll-use records captured at entry points to the toll road.
- Bluetooth sensors installed and connected to high-speed bandwidth infrastructure installed by the toll authority along its toll road.

The Bluetooth data are shared among a set of public agencies in the region where the toll authority operates. For now, these data are limited to tire pressure sensor data from cars and smartphone-based Bluetooth sensor data. The data that they can use mostly from this set are for speed monitoring functions. The car tire data are not used by the authority. The toll authority indicated that it plans to eventually deploy more sensors capable of collecting additional connected car data in order to gather more detail such as car occupant information and incident status.

Different sources of data are presently entering a gray legal jurisdiction based on the product type. For example, according to a data aggregator, "In some cases the original data source owner considers themselves as the data owner for data emitted by the vehicle…the automotive manufacturer is an example of an original source owner where some of the data from a car is owned by the automaker. Recently some courts have ruled that the data produced by the vehicle, such as tire sensor data, is owned by the car owner." This not only contrasts with the automaker but also with the toll authority, who felt that this telemetry data (tire sensors being one) was its right to collect if it built the network and sensors capable of receiving the data. There are also legal requirements for cellular companies in their license agreements with individual customers that allows them to opt out of providing their data for use by data aggregators. On the reverse side, court cases are also increasingly finding for the prosecution, allowing them to admit data from vehicle controller area networks, which are the internal circuitry responsible for gathering data through wireless means, as evidence in criminal cases.

Data Functions and Use

Car-based data and system performance data are obtained from data aggregators for a variety of purposes. Some data aggregators provide connected car data for various services that include:

- OEM product design.
- Transportation planning and operations.
- Site-based business development customer analyses.

Formats vary by the need for geolocating connected car data along certain road networks or transportation modes and users, and for various arrangements in time, whether real-time streams or in historic bins of 1, 5, and 15 minutes.

Presently, public agencies are in the process of verifying the quality of car-based and systems performance data. For example, one DOT purchased data in both a batch purchase (which is largely for planning activities) as well as a real-time data stream in order to measure the quality of data and determine if the data could be used in traffic operations. One researcher examining

different uses of probe data in comparison to sensor-based data explained how he/she "found that it was mainly on freeways where [the aggregator] had enough data to perform planning and traffic operations functions (using the data to respond to incidents). However, the lower volume state roads did not have enough data saturation [from car-based probe data] to enable application of both functions." Some data aggregators struggle with getting a data sample size reflective of actual traffic conditions on non-freeway roads and arterials, which then limits their applicability in adequately serving regional traffic management functions.

For other data aggregators, the opposite is the case, and too much information is being provided. One data aggregator developed a data sharing agreement with a state DOT providing it with the raw data stream from a smartphone application with a high rate of users that saturate the arterials, freeways, and local roads. However, these data are provided in a raw, real-time format that must be filtered and cleaned so that the state DOT can put it to use in traffic management operations such as incident detection on arterials. The state DOT official in charge of the data sharing agreement data advised, "They [the data aggregator] don't scrub the data; they provide it to us in a real-time feed. A lot of the data is not of value to us. For example, the location of police officers noted by drivers as they proceed along their route is not of any value to us. That's not something we would put thru the state 511 traveler information system, and so we have to filter it. They also give us pothole and roadway, but we strip it from our data and then archive it for a certain amount of time. The way the data is gathered is very difficult to manipulate."

These same data also have many other data points that could potentially be of use in the future, but the state DOT struggles to format it due to sheer size and depth of the data and data stream. This is also a problem in that the state DOT is prevented from sourcing out work to experts in the matter of data formatting and archiving due to a stringent data licensing and exchange agreement. The university representative at this meeting with the state DOT advised, "The data aggregator with the data exchange agreement has many constraints on sharing data beyond the state DOT, even with academia. It would be critical to figure out a way to revise the data sharing agreement."

All data aggregators indicated that license agreements and associated terms and conditions are applied to access car data. The use of these data is then limited to terms and conditions spelled out in the user agreement. For example, one data aggregator places requirements into the licensing agreement preventing the combination of its data with other data sources if they will be used to generate reports and information for public consumption. As further explained, "Data retention policies from our requirements are very loose. A while back, someone was combining our data with others and making really bad data, and we ended up having our name on an inferior product because it was combined with inferior data." Such terms and conditions protect the reputation of the data aggregator's data from being compromised by erroneous combinations with other data sets of poorer quality or different uses. Otherwise, the data aggregator was fine with the data being combined with other sources of data internally for the purposes of research and data quality improvements.

One data aggregator indicated that in its user agreements, it provides the option to make the data "open data," which enables it to charge 25 percent more to its clients. This means that the buyer may share it freely with other organizations, consultants, and universities. Several DOTs and local metropolitan planning organizations (MPOs) have purchased their data set under this open

data user agreement. This same data aggregator also advised that it will provide its data oriented toward a time period, for a given route and purpose of analysis (transit ridership, planning, etc.). It indicates in the terms and conditions of the user agreement that the quality of analysis will decrease if it goes beyond the prepared uses of the data set since the data must be prepared in a certain way for the specific conditions of the analysis. One of the reasons the aggregator is okay with data sharing is that "data becomes obsolete and has a shelf-life. So many companies will allow sharing knowing this data will expire, and are comfortable in that most state DOTs will come back to them to renew it with a new rate."

Connected car trajectory data (also known as origin-destination data) can be obtained and used to map out paths taken by cars from origin and destination zones while in transit between zones. Data aggregators provide these data as a specific function to planning agencies and site-selection consultants. Data accuracy for this trajectory data, similar to systems performance data, is often a result of the size of the data set and the representative sample of available data points gathered over a transportation network. As a result, these trajectory data are based on averages over a given time period, rather than real-time data. In this way, travel time averages over a year can be used accurately for planning but not for real-time traffic management and operations.

Another data aggregator indicated that its largest client is the OEMs who use the data for onboard navigation, rerouting around incidents, and parking in real time. This same data aggregator is currently in talks with automotive manufacturers to establish a new source of data known as extended floating car data, which are based on SIM chip onboard units and could eventually relay a wealth of data, including number of occupants in the car.

Data Privacy, Security, and Retention Policies

By law, data source owners such as cellular networks and data aggregators must apply securitization protocols, which remove personally identifiable information. Data aggregators abide by this law through a variety of encryption and infrastructure placement approaches. One data aggregator advised that it takes an umbrella approach, where the data it receives from various source owners has already been anonymized based on the data aggregator's agreement and instructions with the source owners. This first round of data anonymization thus takes place prior to the data aggregator's receipt of the data, and it then further anonymizes the data by automating the rotation of vehicle IDs to add a second layer of data encryption. A second data aggregator that relies on cellular networks to provide data developed its server infrastructure to run at the actual cellular provider's facility, behind existing security and IT firewalls maintained by the cellular networks. In this case, the data aggregator receives data that has already been scrubbed of personally identifiable information, which removes the PII liability from the data aggregator.

Data aggregators collecting trajectory data remove personal identifiers by removing the first and last minutes of the trip. There is an increased risk in the retention of trajectory data in that, as one data aggregator indicated, it is possible to use other data sources in combination with trajectory data to potentially re-identify users. The reason is that even with the origin and destination removed, these trajectory data display the path taken by a car over a transportation network daily. As a result, various other data sets in combination could link these now-visible trips to certain individual patterns of travel. One data aggregator providing trajectory data ruminated on

re-identification of data from trajectory data and advised, "Regarding whether records could be reidentified, if in an area with high fidelity and with small bubbles, I could reverse ID someone. If I knew where you lived, and in which neighborhood, and where you worked, and some pattern you took, such as two days out of the week you go over to this restaurant, … there is a high probability that I could re-identify your record." This data aggregator advised that, as a result, it was initially subject to security audits by the cellular source owner as per its agreement to share trajectory data with the data aggregator.

One research institution estimated that as connected cars come online, the associated data will follow Society of Automotive Engineers (SAE) standards, which require that the car ID be temporary and change every five minutes. This will help to secure this source of data and possibly ready it for use in wider data sharing applications similar to the anonymization efforts applied to data sets held by data aggregators today.

A potential complication to this connected car security standard is the degree of latitude that public agencies develop in their efforts to collect data using their own network of sensors. One toll authority indicated that as far as it was concerned, if it can obtain car-based data from sensors it has installed, then it will put that information to use, especially in its sharing agreements with regional public agencies. However, it also indicated that it strictly separates its data collected from Bluetooth sensors from its RFID-based toll sensor data, which contains PII. Furthermore, data from financial transactions are not shared with other transportation agencies: "We do pay-by-mail and we have locations and times when cars went thru the tolls; but they can access their own information their own toll data by using the information, and that is the only personal data we have. We don't share any of that [PII] with other state or regional agencies." This is a result of federal and state laws requiring tolling authorities to protect this PII from being used in combination with other sources of data or applied for use in transportation operations and other similar functions.

Connected car data provided in real-time stream format may be retained and archived at the discretion of the buyer, as long as it does not violate the original terms and conditions set forth in the license agreement. One toll authority indicated that it retains transponder records from RFIDs installed on cars going back approximately 10 years, after which new toll data records begin to rewrite over these data. Presently, the 2005 data set is being written over.

Data Sharing

Data sharing for car-based and system performance data is oriented around a data ecosystem that involves gathering and sharing activities. Data aggregators gather their original source data from what are called source owners, which may be internal to the data aggregator or could include cellular network providers, GPS probes in freight fleets, smartphone applications, and original automotive manufacturers. Data aggregators establish formal data sharing agreements with source owners that establish how the data are to be used and how sharing and privacy issues will be resolved. Some data aggregators establish data sharing arrangements with public agencies that provide them with road-based sensor data (speeds, volumes) in exchange for a variety of vehicle-based data. One smartphone-based data aggregator providing these data placed very tight requirements to keep the data within the DOT and to not share it with other organizations such as consultants or university partners.

On the data sharing side, data aggregators typically set up terms and conditions with the customer, detailing who the purchased data can or cannot be shared with beyond the buyer on the contract and setting up limitations on types of use. Typical customers for some data aggregators include state DOTs with associated research institutions. There are no data licensing standards that they rely on in developing license agreements other than that they have provisions for anonymizing or protecting personally identifiable information within the data sets.

One toll authority indicated that it is planning to share the car-based data it collects through its data sensors installed along the roadway with other operations entities in the region for use in real-time routing and traveler information systems.

Additional Comments of Interest

A third area of data use from car-based data is land-use data. These data are based on origin and destination data and are typically purchased by site-selection consultants to determine level of competition for businesses considering a location. These car-based data can potentially be applied to land-use analyses performed by public agencies as well.

Infrastructure Facility Owner-Operators

Infrastructure is a critical component of the connected car environment. The infrastructure subsystem provides and exchanges data elements related to roadway characteristics, road conditions, intersection status, and field equipment status (Michigan DOT et al., 2014). Data resulting from this interchange of information are typically collected and provided by the agencies or authorities that own the infrastructure. The V2I Deployment Coalition, which has the mission "to work collaboratively with the industry, state and local governments, academia and USDOT to achieve the goal of deploying and operating a functioning V2I infrastructure," states on its website that achieving this goal will require "long-term cooperation, partnership, and interdependence between the infrastructure owners and operators (state, county, and local level transportation agencies); the automobile industry OEMs, and aftermarket manufacturers; and a variety of other stakeholders" (Vehicle to Infrastructure Deployment Coalition, 2016). The owner-operators who participated in the interviews represent the viewpoints of state DOTs, large MPOs, and regional authorities.

Data Ownership

In general, state DOTs advised that once car data are collected by a DOT's sensors, those data become the DOT's. As one DOT said, "Any broadcast data is public information, and if DOT collects it with their sensors, then the DOT owns it." One DOT did acknowledge that the OEMs and individual car owners may have differing opinions. The MPOs and authorities are not collecting their own connected car data yet, and if they need the information, they are getting it from the state DOT.

There are ongoing discussions between agencies and OEMs regarding data ownership, and as one state DOT representative stated, "Data ownership is a complicated question." A September 2016 meeting was organized between OEMs and state DOTs to discuss AV/CV data ownership. State DOT interviewees stated that in these discussions, the OEMs indicated that they see the data coming from cars as theirs. However, car owners may feel that the data belong to them. As

one interviewee stated, "Ownership of the data depends on where the data is in the stream. For example, when it is generated in the car, it is the property of the car owner. If it is collected by agency sensors, it is the property of the agency. If a third party compiles it, the data becomes the property of the third party."

Data Functions and Use

Data collected are typically used for:

- Planning purposes, such as running models and forecasting, performance monitoring, and project evaluations.
- Operations, including feeder information for advanced traffic management systems (for example, travel time information and slow-down reports) and ramp metering.

Locating bottlenecks and routing drivers accordingly is important because, as one DOT indicated, "An informed traveler is a safe and efficient traveler." Information is also routed to the media for drive-time reporting. Probe-type data are fed by one agency into the 511 system.

Several participating agencies stressed the need to recognize the difference between data and useful information and noted that the information derived from the data is what they need. As one state DOT stated, "Data is a single point of reference and information is aggregated from the data in a way that is useful. For example, data might be the speed of a single car and information would be the average travel speed." Interviewees stated that public agencies do not have the expertise for analysis and are dependent on third-party data aggregators to perform the analysis to transform the data into useful information. These third parties include companies like INRIX and, in some cases, academic/university partners. The data are freely shared with these third parties in exchange for useful information.

Some stated that future plans include using connected car data for issuing alerts to roadway users regarding wrong-way drivers and alerts or appropriate responses in adverse road conditions such as ice or heavy rain. Other future applications may include incident notification if an air bag deploys. One DOT stated that it will focus on signalized intersection applications, where drivers can recognize basic safe message (BSM) data and react in real time. The example given was for a driver at a timed intersection late at night; it would be advantageous for the signal to have the ability to let the driver through rather than having him or her wait at a light when there is no traffic. Applications for signalized intersections are currently under development.

All owner-operators interviewed shared that, as public agencies, they are bound to protect the privacy and security of the data and ensure its quality; otherwise, travelers will lose confidence in the data's value. There were concerns that "most public agencies do not have the resources to ensure high levels of accuracy. Data may need to be 'cleaned up' to some extent before it is useable." Data provided by public agencies must be accurate and up to date; for example, signal phase and timing (SPaT) and signal timing data must be current and up to date in real time.

Data Privacy, Security, and Retention Policies

The agencies interviewed shared that privacy of connected car data is an acknowledged concern, and that maintaining privacy is an agency responsibility. Some agencies are bound by legislation

to provide privacy in data collection, meaning that data cannot be tracked to individuals, but there were differing opinions on how to accomplish this. For example, some agencies interviewed routinely collect Bluetooth data and consider that to be private. Others are re-labeling Bluetooth data as Anonymous Re-Identification Device (ARID) data in response to privacy concerns regarding Bluetooth. One interviewee stated that "SPaT data broadcasts required for CV can be contentious. Some local agencies are contemplating sharing this data freely with vendors, but professional staff at many agencies is very concerned."

The aggregation necessary to make the data useful, for example in origin-destination studies, helps to protect the identification of specific data points. One DOT indicated that for current pilot studies, the user must opt in before any personally identifiable, non-aggregated data are collected. Privacy concerns are so strong in one state that cameras are not allowed to zoom in on license plates. In the case of several agencies, TMC cameras do not record in order to address privacy concerns; they are solely used for live traffic monitoring.

Security was of concern to all owner-operators participating in the interviews, although some local agencies only use archived data for planning purposes so have less vulnerability to malicious mischief. One agency requires a cryptocard (i.e., secure account system) for the use of safety-related data, which requires a signed agreement. Toll authorities must adhere to credit card standards because user accounts are linked to them, which imposes a high level of security. Agencies that use data collected by others stated that security is "the collecting agency's issue." For state DOTs, multiple firewalls protect data and routine quality assurance/control checks will detect anomalies. One DOT has a private network for AV/CV data, separate from its ITS systems, with separate fibers, etc. A concern was expressed that the "DSRC is not mature—we need to make sure it can't be compromised by hackers." The work in this area is still in the study stage, but this potential vulnerability must be addressed before the data can be used in practice. One DOT shared that while Bluetooth and wireless data are cheap and easy to collect, it has not addressed how to ensure security. As one state DOT interviewed stated, "Security—this is still a big question. The system could possibly be hacked into, although SCMS [Security Credential Management System] is used. Theoretically, it should be secure."

Data retention policies vary widely among the agencies surveyed. Some are currently keeping data indefinitely, have not yet considered how to address the volumes of AV/CV data on the horizon, and have no policy in place. One state DOT's policy is to archive data for up to one year, according to individual district policy, but not all districts archive. This DOT currently does not archive video but will change its policy this year to keep video for five days. If there is a freedom of information request or subpoena in that five days, it will keep the video for one year. The representative stated that "this video is good for training purposes, so why not keep for the good of the public for a limited time?" Other state DOTs contacted do not store video from CCTVs at all because of the potential liability issues, and because they do not want their video to be discoverable for litigation. In one state, data are incorporated into the Performance Management System and maintained in perpetuity. In another state, state law limits data retention to three years; however, the DOT was granted a waiver for AV/CV type data and will be keeping it in perpetuity.

Data Sharing

All of the state DOTs interviewed for this study stated that data will be shared upon request, although the conditions and mechanisms for doing so vary. In one state, if the requested data are part of an active police investigation, they cannot be shared. One DOT requires others requesting data to recognize the DOT as the source if they use it. Several have portals or web pages in place to make the data accessible, and they believe that because they are a public agency, it is appropriate that the information be available upon request. Some of the access mechanisms used are an Application Protocol Interface (API), the Commercial Wholesale Web Portal (CWWP), and a Traffic Video and Data Dissemination (TVDD) system. Of these three, the TVDD system does not currently contain AV/CV data but is planned to in the future.

There are some conditions for access to data sharing websites and portals at some DOTs. The API is available to anyone, but users must provide an email address. The TVDD requires users to go through a sign-up process. The CWWP is available through the state DOT website and is available to anyone; it is currently being used by data aggregators such as Waze and INRIX.

An MPO interviewed stated that its state DOT is willing to share data with the MPO for its use, but an agreement is required and ownership of the data is retained by the DOT. It cannot be passed along to any other entity from the MPO.

State law prohibits the tolling authority interviewed from sharing any information that is collected for the purposes of collecting money, which would include camera or video images. The law is fairly new, so its long-term effects are yet to be determined. The only way this information can be shared is by subpoena. The tolling authority's attorneys are currently working on whether the law would apply to AV/CV data because "anything generated by a phone could trigger the law, since there is an ability now to pay through phones," and this could apply to Bluetooth data.

Additional Comments of Interest

Owner-operators interviewed acknowledged that most public agencies do not have the resources required to process data into information—that will require partnerships between OEMs and third parties. There are too much data and too many resources required. As one owner-operator said, "Agencies should target the information that is absolutely required instead of trying to do everything. Then they could possibly manage some of it." Also, at some point after agency data have been processed by data aggregators such as INRIX or others, those entities become the data's owner. Should these aggregators have to pay the agencies for the data since they profit from those data?

One individual said it would be good to know if USDOT will be taking steps to include CV infrastructure projects that improve safety as eligible for federal safety funds. This could speed implementation considerably, but no one has indicated that this would be allowable with federal safety funds.

Owner-operators are interested in OEM data and acknowledge that at some point, the OEMs may recognize that the car data are worth money and will figure out a model to monetize the data. Owner-operators do not want to execute user agreements for every OEM and every car make and

model. They would prefer a third-party solution where the third party negotiates with agencies and OEMs for their data, converts those data into useful information, and sells them back, which one respondent said "could be a successful business model." One state DOT representative stated that "we look to the third-party solution because data management is not part of our core business, and we aren't very good at it. The sensors are hard to maintain, hard to use, etc. We do need the information derived from the data, however, although we do not have to collect it ourselves, and would prefer not to. We would be happy to purchase it." OEMs would like access to agency-collected data as well.

One owner-operator advised that there are two kinds of data: BSM 1 and BSM 2. BSM 1 is information such as location, heading, and speed—basically probe data (with perhaps 10–12 data points). BSM 2 is information from the car itself, such as whether the lights or wipers are on, the amount of traction, etc. BSM 2 data are much more voluminous (100–200 data points), and the OEMs have discretion over whether that information is transmitted. BSM data could go to a third party for collection and management, and then it would provide processed information from these data for a fee. Most applications require BSM 2 data, which are OEM data, and therefore access would have to be negotiated with OEMs. BSM 2 data and types of data are unique to each OEM.

Based on the national discussion between agencies and OEMs, there are currently six key applications that will use OEM data; however, there will be more applications in the future.

- Queue warnings.
- Data for traffic operations.
- Intelligent traffic signal systems.
- Reduced speed zone warnings/warnings about upcoming work zones.
- Curve warnings.
- Road weather motorist advisories and warnings.

One individual summed up by saying, "We will only know what is feasible after full AV/CV deployment, so we need to do things incrementally, and address problems as they occur. Trying to anticipate all the potential problems beforehand would not be feasible."

CHAPTER 4. CROSS-CUTTING THEMES AMONG STAKEHOLDER PERSPECTIVES

OEMs, data aggregators, and owner-operators all have legitimate business interests in connected car data. The answer to the question of who owns the data is influenced by their data concerns, opportunities for monetization, and missions. Six broad themes emerged from an inductive analysis of the interviews that provide insight to the different perspectives on data ownership: (a) where data are recorded matters, (b) monetization for all, (c) monetization impacts sharing, (d) different roads to data privacy, (e) more is not better, and (f) building a common lexicon.

Where Data Are Recorded Matters

Where data are recorded—inside a car or outside of a car—influences perceptions of data ownership. OEMs consider the owners of cars to be the owners of data generated by the cars and consider themselves to be stewards of those data with full access privileges once the owner has opted in (or not opted out). According to one OEM, "Once people click the user agreement, they have given up the data." These data (like real-time status of features, diagnostics, or location) are recorded by devices in the car. Some data remain in the car and some are transmitted to the OEM or a third-party provider wirelessly and directly.

Data aggregators and owner-operators typically use data that are recorded by devices outside of the car (e.g., by roadside sensors). State DOTs feel that they own the data collected by their sensors. From the owner-operator point of view, since the data are recorded outside of the car, the data are fair game: "Any broadcast data is public information, and if the DOT collects it with their sensors, then the DOT owns it."

Data aggregators feel that they own the connected car data that they gather from various sources (including state DOTs), process and package, and then sell to buyers (including state DOTs). Data sources include smartphone apps, roadside sensors, and GPS data from fleets. By processing the data, the data aggregators are creating a new information product and therefore are owners of that new product.

Data ownership is complicated and nuanced. As one interviewee stated, "Ownership of the data depends on where the data are in the stream." When the information is generated in the car, it is considered the property of the car owner and is protected and used by the OEM. If it is collected by agency sensors, it is considered the property of the agency. If a third party compiles it, the information becomes the property of the third party.

Monetization for All

All three stakeholder groups want to monetize connected car data. Data aggregators are openly in the business of data monetization because of how they generate revenue. They clean data of personally identifiable information, manufacture the information, and sell it. Unfortunately, they typically do not own the raw data they use and so must purchase the data or acquire them through usage agreements with the data owners—much like the raw materials for any product. However, they do own the processed and formatted information that they sell. As profit-generating entities, they are always trying to find new data products to sell, such as the one data

aggregator in talks with OEMs to establish a new source of data—extended floating car data. Such data would augment car speed and trajectory information they currently use with information about the number of people in the car.

OEMs do not consider themselves the owners of connected car data, but as the data stewards, they have access to use it. In the past, the data were used primarily for internal purposes like improving safety, verifying vehicle quality, predicting customer behavior, and analyzing driver behavior. For example, the OnStar service offered through GM has used connected car data to provide customers with vehicle health reports since the late 1990s. However, as cars become more connected, opportunities arise for car data monetization. One OEM said, "Our mission is to determine how to monetize the data." These opportunities include identifying additional ways in which to improve safety (e.g., advancing vehicle automation), generating revenues by selling car-related services/features to consumers (and the related consumer insights to third parties); and leveraging data to reduce manufacturing costs and/or safety risks. Thus, monetization takes two forms: creating new revenue and identifying ways to reduce business costs.

Owner-operators, who for the most part are public agencies, typically are mission-focused rather than revenue-focused. Most missions focus on moving people and goods safely and efficiently. In the past, they have freely shared data coming from their sensors to fulfill this mission. However, in times of budget constraints, they are also beginning to consider the opportunities for monetizing it. As one interviewee asked, "Should aggregators have to pay the agencies for access to the data since they profit from it?"

Monetization Impacts Sharing

Perspectives on the issue of data sharing are influenced by the desire to monetize the data. OEMS and data aggregators as revenue-focused entities are less open to sharing data than owner-operators. OEMs share data, but only with entities that are involved in their internal processes. Sharing with outside entities, such as transportation agencies, is challenging because "data is very valuable." They might be willing to share for a cost or for a one-off study, but they would be less interested in sharing information that has a commercial value for them, such as driver behavior data. Owner-operators would like to have OEM data like driver behavior data, but they do not anticipate that happening.

Data aggregators do not share data—they sell it. They typically set up terms and conditions with the buyer regarding with whom the data can and cannot be shared beyond the buyer on the contract. They also typically set up limitations on how the data can be used. Data aggregators themselves will have formal data sharing agreements with the source owners of the data that establish how the data are to be used and how sharing and privacy issues will be resolved. At least one data aggregator monetizes data even further by charging a premium when data come with the opportunity to be shared freely with other organizations.

To fulfill their missions, owner-operators typically share data on request. With the proliferation of apps and connected devices, data are moving from people to owner-operators through crowdsourcing or sensing devices and from owner-operators to people and data aggregators through open data initiatives. The benefits for the owner-operators are the opportunities to get useful information in return and to enable services or products that make them look more

responsive, transparent, and effective in serving their constituencies. Some agencies make the data easily accessible through portals or web pages with minimal conditions for access, such as providing an email address or having an account. In a few cases, the owner-operators have data sharing agreements in place for specific uses of the data.

Different Roads to Data Privacy

All three stakeholder groups are aware of the need to address privacy in data collection and use but take different approaches for addressing it. The data aggregators are the most stringent in this regard, perhaps because they are the most embedded in the data business. By law, data aggregators and the data source owners they work with (such as cellular networks) must apply securitization protocols to connected car data to remove PII. Some aggregators contractually rely on the source owner to remove PII, which removes liability from the aggregator. Others strictly separate databases that contain PII from those that do not in order to decrease opportunities for using a second database to re-identify PII. Others are looking to apply SAE standards for connected vehicles, which will require that the vehicle ID be temporary and changing every five minutes.

Owner-operators acknowledged that maintaining privacy is an agency responsibility. However, there were less specifics provided on how data are to be de-identified. For example, one owner-operator opined that Bluetooth data by nature are private because only a MAC address is logged—this is the Wi-Fi or Bluetooth identifier that is broadcast when the Wi-Fi or Bluetooth is turned on. However, a privacy debate has centered on the use of Bluetooth and hinges on whether a MAC address should be considered private information. There was also the thought that privacy is safe if the agency only uses anonymous aggregate data, but recent experiments have demonstrated the ease with which anonymous data can be reidentified, that is, combined in a manner that results in identifying individuals to a great degree of certainty. Another owner-operator mentioned that public agencies are contemplating sharing SPaT data broadcasts required for CVs freely with vendors. Professional staff at many agencies are concerned because of privacy issues.

OEMs spoke less about data privacy issues, perhaps because they consider the set of consumer privacy principles to be a complete privacy solution. However, the principles do not address data de-identification. It appears that once a customer has opted in to (or not opted out of) the user agreement, the agency has the right to use the data. At least one OEM stated that when data are tied to the individual, it needs to be careful with how data are treated. When data are de-identified, the OEM can do what it wants. Another said it "shared anonymized and aggregated vehicle data in analytic reporting to federal and state regulatory agencies and non-profit organizations for the purpose of education and research related to environmental and energy issues, advanced technologies, and usage analysis." This was done without regard for the fact that these data could perhaps be de-anonymized.

More Is Not Better

More data are not better data because they take greater resources to use; information derived from data is best. While some data aggregators struggle with getting an adequate sample of data points along certain types of roads, generally the raw data that data aggregators provide to buyers

are too much. State DOT users often struggle to format and use the data due to the volume and depth of the data and the data stream. One interviewee explained, "Data is a single point of reference and information is aggregated from the data in a way that is useful." As an example, data might be speed of a single car, whereas information would be the average travel time on a facility. Owner-operators value the analysis that the aggregators perform to transform the data into useful information, and so have freely shared their data with them. One interviewee noted, "We look to the third-party solution because data management is not part of our core business, and we aren't very good at it." Another explained, "If agencies could target the information that is absolutely required (instead of trying to use everything), they could possibly manage some of it."

Build a Common Lexicon

Communication among OEMs, data aggregators, and owner-operators might be improved if these entities had a common lexicon for discussing connected car data. A common language of key concepts would provide a basis for shared meaning and understanding and may facilitate resolving data ownership, data sharing, or other relevant data issues. Different labels or jargon were used by the three types of stakeholders (and even within a stakeholder group) to describe the different buckets of connected car data of interest.

Some of the owner-operators talked about BSM 1 and BSM 2 data, without specifying what specific data elements are implied. Aggregators expressed interest in connected car data, generally, and were more specific on the sources of data than on the data elements themselves. OEMs typically mentioned customer contact data, vehicle health data, diagnostic data, and driving information, but all used different labels for discussing the data elements.

A potential classification scheme might have a hierarchical listing of data categories and associated definitions. The classifications would be comprehensive and exclusive at the upper levels of the hierarchy, expandable at the lower levels, and simple, consistent, and scalable at all levels. The listing would not take the place of the connected vehicle reference implementation architecture, for instance, which is very detailed, but would be a less technical lexicon that would be focused on facilitating collaboration among different industries.

CHAPTER 5. CONCLUSIONS AND RECOMMENDATIONS

This study examined connected car ownership issues, and researchers drew conclusions regarding three key research questions. Recommendations for state and local transportation agencies are presented within the answers to the questions.

Who Owns the Data?

Data ownership is complicated and nuanced. Data ownership seems to be in the eye of the beholder. When viewed from three different stakeholder perspectives, the research team found that:

- OEMs acknowledge that the owner or lessee of the car is the owner of the connected car data; however, they are able to access and control the data through user agreements. Privacy principles are used to provide transparency to their data collection, use, and sharing practices so as not to discourage customers from opting in to the agreements. Customer trust in terms of opting in is essential for the OEMs' ongoing use of the data to improve their automotive products and develop new customized offerings.

- Data aggregators consider themselves to be the owners of the information that they sell that is derived from the data. These data have been gathered from many sources, processed, and formatted into new information products. While they may not be the owners of the source data, they are the owners of the new information products that they create.

- Owner-operators consider themselves to be the owners of the data collected by their sensors. Since the data are recorded by sensors outside of the car, they view the data as fair game. Broadcast data are viewed as public information.

Given these different views on data ownership, does ownership matter? It matters because ownership of data is tantamount to control, determining who can collect, process, use, and share the data. Ownership also implies who can profit from the data. However, just as important, ownership implies a broader responsibility—data stewardship—where the owner must consider the consequences of how the data are used, particularly for how a particular use might impact data privacy or data security. Data privacy and security are becoming more important to individuals, as indicated by Pew Research Center data. Two-thirds of respondents in a 2015 survey felt it was important to be able to "go around in public without always being identified." Half did not trust that government agencies would keep their personal information safe and secure (Madden and Rainie, 2015).

In this climate, private individuals may begin to overtly control access to their data, including data generated by cars that they own or lease. In much the same way that OEMs were preemptive in their enactment and communication of privacy principles with individual owners or lessees of cars, state and local transportation agencies should begin to *transparently* communicate with users of their facilities about collection, storage, retention, use, and sharing practices related to car-based data captured by roadside sensors.

Under What Conditions Are Owners Willing to Share Data?

Desires to monetize data influence willingness to share connected car data. OEMs and data aggregators are revenue-focused entities and are only willing to share connected car data under certain terms and conditions. OEMs regularly share data with entities that are involved in their internal business processes, such as their third-party service providers, parent companies, or credit corporations. OEMs would be willing to share data with transportation agencies in one-off situations in which the car owner agreed to the special data sharing, but not in any type of continuing collaboration. A caveat even to this is that they would be less interested in sharing data that might have commercial value for the OEM. Data aggregators sell information derived from data; they are not interested in sharing data.

Owner-operators, on the other hand, typically share data on request; data sharing facilitates their missions, which for most are moving people and goods safely and efficiently. Another benefit to the owner-operator of sharing data is to have the data aggregator process and reduce the data into useful information. Transportation agencies often do not have the human or monetary resources to transform streams of data into useful information.

However, much like some private individuals may be looking at their own value in a digital world—with data being an owned asset just like a house or a car—some owner-operators may be shifting in the value they place on their sensor data. The question of whether agencies should be paid for access to sensor data by entities that profit from it may become more commonly considered by state and local transportation agencies. Determining appropriate business models for this has yet to occur.

Data sharing entails collaboration among entities, and collaboration is facilitated by a common language of key concepts. This research revealed that different labels or jargon are used by the three stakeholder groups to describe the different types of connected car data of interest. The necessary common lexicon would be non-technical and easily cross-walked among the data interests of OEMs, data aggregators, and state and local transportation agencies.

How Should Transportation Agencies Responsibly Use Connected Car Data?

Responsibly using connected car data relates to the concept of data stewardship. Data stewardship implies a fiduciary or trust relationship with individuals whose data are stored and managed by the steward. In this perspective, data ownership is less important. Stewardship relates to ensuring PII is protected and secure. Because of increasing concerns about data privacy and security among the public and policy makers, these issues have the potential to become contentious for state and local transportation agencies.

In this context, transportation agencies can take their lead from the OEMs in their enactment of the consumer privacy principles presented in this report. Certainly, these principles reflect best practices from the business perspective of the OEM. Another source discussed in the report is the FTC's fair information principles. These relate primarily to Internet- or web-based practices.

Because of uncertainties in data ownership, among other issues, best practices for transportation agency use of connected car data have been slow to develop. A recent National Cooperative Highway Research Program (NCHRP) report consolidated best available guidance (i.e., White

House, National Institute of Standards and Technology, Organization for Economic Cooperation and Development, International Association of Chiefs of Police) into the following practices (Zmud et al., 2016):

- **Transparency and openness**: Agencies should notify or otherwise communicate the types of information they collect and how that information is used, disseminated, and shared to individuals within their jurisdictions.
- **Purpose specification**: Agencies should clearly communicate why they are collecting information and under what authority; a change in purpose requires an update of the communication.
- **Data minimization, retention, and use limitation**: Agencies should only collect information that is necessary to meet their specified purpose, retain it for only as long as needed, and restrict its use to only specified purposes.
- **Data quality and accuracy**: Agencies should ensure that data are accurate and of high quality, and—when relevant—enable individuals to review and correct any information.
- **Accountability**: Agencies should define explicit policies and procedures for complying with data protection principles.
- **Security**: Agencies should protect personal data with reasonable measures to prevent loss, unauthorized access, or disclosure.

These principles can serve as a harmonized set of practices for risk mitigation, with the recognition that balancing agencies' needs for using connected car data with individuals' concerns for data privacy and security is a complicated challenge.

In conclusion, transportation agencies tend to move forward deliberately and slowly in reaction to change. Connected car data represent an emerging data source with immense value and the potential to vastly improve transportation planning, traffic management and operations, and safety in cities and regions. Faced with this opportunity, state and local agencies should be proactive in determining the ways in which they can have access to the data. They should begin to participate in, and even be leaders of, national, state, and local discussions and collaboration activities regarding how to responsibly collect, use, share, and disseminate connected car data.

REFERENCES

Alliance of Automobile Manufacturers and Association of Global Automakers (2014). Consumer Privacy Protection Principles: Privacy Principles for Vehicle Technologies and Services. Retrieved July 15, 2016. Accessed from https://www.globalautomakers.org/media/papers-and-reports/privacy-principles-for-vehicle-technologies-and-services

Anderson, B. (2013) The Difference between Data Privacy and Data Security. Retrieved June 1, 2016. Accessed from http://blog.eiqnetworks.com/blog/bid/313892/The-Difference-Between-Data-Privacy-and-Data-Security.

BC Freedom of Information and Privacy Association. (2015) *The Connected Car: Who is in the driver's seat: A Study on Privacy and Onboard Vehicle Telematics Technology*. Vancouver: British Columbia Freedom of Information and Privacy Association.

Canis, B., and D. Peterman. (2014) *"Black Boxes" in Passenger Vehicles: Policy Issues.* Washington, DC: Congressional Research Service.

Chisholm, M. (2011). "What is Data Ownership." Retrieved August 10, 2016. Accessed from http://www.b-eye-network.com/view/15697

Demster, B. (2012) "Data Ownership Evolves with Technology." *J AHIMA*, V(83), 9.

DePorre, E. (2016). Statement made during panel session, "Balancing Security, Privacy, and Innovation in AV Data Use" at Legal Breakout Session, Day 2, Automated Vehicle Symposium 2016, San Francisco, CA.

Diamond, C.C., F, Mostashari, and C. Shirky (2009) "Collecting and sharing data for population health: a new paradigm." *Health Affairs*, V(28), 2, pp. 454–466.

Federal Trade Commission (FTC). (1998*) Privacy Online: A Report to Congress.* Washington, DC: FTC.

Future of Privacy Forum (2014). The Connected Car and Privacy: Navigating New Data Issues. Washington, DC: Future of Privacy Forum.

Grandison, T., and Mohammed, A. (2013) "Patient Data Ownership." Retreived April 8, 2016. Accessed from http://www.slideshare.net/tgrandison/patient-data-ownership.

Hong, Q., Wallace, R, Krueger, G. (2014) *Connected v. Automated Vehicles as Generators of Useful Data.* Michigan Department of Transportation, Center for Automotive Research, Leidos. Lansing: Michigan Department of Transportation.

Huet, E. (2014) "At $18.2 Billion, Uber Worth More Than Hertz, United Continental" Retrieved September 26, 2016. Accessed from http://www.forbes.com/sites/ellenhuet/2014/06/06/at-18-2-billion-uber-is-worth-more-than-hertz-united-airlines/#5898605639d6.

Jolly, L. (2014) "Data Protection in the United States: Overview." *Practical Law*. Multi-Jurisdictional Guide 2014/15. Association of Corporate Counsel. Retrieved August 5, 2016. Accessed from http://us.practicallaw.com/6-502-04671.

Koch, J. (2006) *Event Data Recorders and their Role in Automobile Accident Litigation.* Retrieved July 28, 2016. Accessed from http://www.jlolaw.com/wp-content/uploads/2015/07/AutoDiagnosticModules.pdf

Leathem, E. (1963) "Defense Procurement—A Complex of Conflicts and Tensions*." Boston College Industrial and Commercial Law Review*, V (V), 1, pp. 1–19.

Loshin, D. (2002) "Rule-based data quality". In *Proceedings of the 2002 ACM CIKM International Conference on Information and Knowledge Management* November 4-9, 2002, McLean, VA, USA. pp. 614-616.

Madden, M., and L. Rainie. Pew Research Center, 2015, "Americans' Attitudes About Privacy, Security and Surveillance." Retrieved August 10, 2016, Accessed from http://www.pewinternet.org/2015/05/20/americans-attitudes-about-privacy-security-and-surveillance/

Markey, E. (2015). *Tracking and Hacking: Security and Privacy Gaps Put American Drivers at Risk.* Report written by the staff of Senator Edward J. Markey (D-Massachusetts). Retrieved June 20, 2016. Accessed from https://www.markey.senate.gov/imo/media/doc/2015-02-06_MarkeyReport-Tracking_Hacking_CarSecurity%202.pdf.

Mckinsey & Company (2016). "Creating Value from Car Data." Retrieved August 12, 2016. Accessed from http://www.mckinsey.com/industries/automotive-and-assembly/our-insights/creating-value-from-car-data.

Mohammed, J. (2015) "How Connected Cars Have Established a New Ecosystem Powered by IoT." Retrieved September 8, 2016. Accessed from https://techcrunch.com/2015/01/31/how-connected-cars-have-established-a-new-ecosystem-powered-by-iot/

Ng, A. (2011) *Copyright Law and the Progress of Science and the Useful Arts*. Northampton, MA: Elgar.

National Conference of State Legislatures (2015) Privacy of Data from Event Data Recorders: State Statutes. Retrieved July 28, 2016. Accessed from http://www.ncsl.org/research/telecommunications-and-information-technology/privacy-of-data-from-event-data-recorders.aspx

Pentland, A. (2009). "Reality Mining of Mobile Communications: Toward a New Deal on Data." In *The Global Information Technology Report 2008-2009: Mobility in a Networked World*, edited by S. Dutta and I. Mia. Zurich: World Economic Forum.

Pomerantz, F. and A. Aisen (2015) "Auto Insurance Telematics Data Privacy and Ownership." In *MEALEY's Data Privacy Law Report.* V (1), 1.

PwC (2015). "Key findings from the 2015 US State of Cybercrime Survey." Retrieved June 20, 2016. Accessed from http://www.pwc.com/us/en/increasing-it-effectiveness/publications/us-cybercrime-survey-2015.html.

PwC (2016). "Connected car report 2016: Opportunities, risk, and turmoil on the road to autonomous vehicles." Retrieved October 12, 2016. Accessed from http://www.strategyand.pwc.com/reports/connected-car-2016-study.

Rosner, G. (2014). "Who Owns Your Data"? In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, Adjunct Publication, pp. 623–628. Seattle, WA: Ubicom.

Seashore, S. (1978) "Plagiarism, Credit Assignment, and Ownership of Data." *Professional Psychology*, pp. 719–722.

Singleton, M. (2015) "Nearly 1 billion records were compromised by data breaches in 2014: Data breaches were up 49 percent from 2013." Retrieved August 20, 2016. Accessed from http://www.theverge.com/2015/2/12/8028399/hackers-compromised-1-billion-data-records-2014

Sotto, L., and A. Simpson (2014) Data Protection & Privacy 2015, United States, *Getting the Deal Through*. Retrieved July 29, 2016. Accessed from https://www.huntonprivacyblog.com/wp-content/uploads/sites/18/2011/04/DDP2015_United_States.pdf

Stanley, K., and J. Wagner. (2015) *Data Privacy Considerations for Automated and Connected Vehicles.* College Station: Texas A&M Transportation Institute.

Statista (2015). "Connected Car." Retrieved September 13, 2016; Accessed from https://www.statista.com/outlook/320/109/connected-car/united-states#

Telematics Task Force. (2014) *White* Paper: Telematics Data Definition. Retrieved June 15, 2016. Accessed from http://aftermarkettelematics.org/wp-content/uploads/2014/12/Telematics-Data-Definition-White-Paper-12-10-copy.pdf.

U.S. Department of Transportation (2016). Fact Sheet: Federal Automated Vehicles Policy Overview. Retrieved September 20, 2016. Accessed from https://www.transportation.gov/AV-factsheet

Vehicle to Infrastructure Deployment Coalition website (2016). Retrieved September 14, 2016. Accessed from http://www.transportationops.org/V2I/V2I-overview

Walker, S., S. Shogun, L. Huntsinger, R. Wallace, Q. Hong. (2015). *Use of Data from Connected and Automated Vehicles for Travel Demand Modeling.* Michigan Department of Transportation, Center for Automotive Research, Parsons Brinckerhoff. Lansing: Michigan Department of Transportation.

Westin, A. 1967. *Privacy and Freedom*. New York: Atheneum.

Zmud, J., J. Wagner, and M. Moran. 2016. *License Plate Reader Technology: Transportation Uses and Privacy Risks*. Forthcoming. NCHRP Project 08-36, Task 136. Washington, DC: Transportation Research Board.

APPENDIX. GLOSSARY OF TERMS

Selected terms used in this report are defined below.

Aggregated Information: Information elements collated on a number of individuals, typically used for the purposes of making comparisons or identifying patterns.

Anonymized Information: Previously identifiable information that has been de-identified and for which a code or other association for re-identification no longer exists.

Confidentiality: Preserving authorized restrictions on information access and disclosure, including means for protecting personal privacy and proprietary information.

Connected Car: A vehicle that comes equipped with technologies and services that transmit and receive data wirelessly. This allows the car to share Internet access with other devices both inside and outside the vehicle.

Context of Use: The purpose for which PII is collected, stored, used, processed, disclosed, or disseminated.

Data Breach: The intentional or unintentional release of secure information to an untrusted environment.

Data Ownership: The act of having legal rights and complete control over a single piece or set of data elements. It defines and provides information about the rightful owner of data assets and the acquisition, use, and distribution policy implemented by the data owner.

Data Privacy: The relationship between collection and dissemination of data, technology, the public expectation of privacy, and the legal and political issues surrounding privacy. Also called information privacy.

Data Security: Protecting data, such as a database, from destructive forces and from the unwanted actions of unauthorized users

Data Steward: A person responsible for the management and fitness of data elements—both the content and metadata. Data stewards have a specialist role that incorporates processes, policies, guidelines, and responsibilities for administering organizations' entire data in compliance with policy and/or regulatory obligations.

De-identified Information: Records that have had enough PII removed or obscured such that the remaining information does not identify an individual and there is no reasonable basis to believe that the information can be used to identify an individual.

Distinguishable Information: Information that can be used to identify an individual.

Harm: Any adverse effects that would be experienced by an individual (i.e., that may be socially, physically, or financially damaging) or an organization if the confidentiality of PII were breached.

Linkable Information: Information about or related to an individual for which there is a possibility of logical association with other information about the individual.

Linked Information: Information about or related to an individual that is logically associated with other information about the individual.

Ownership: Legal title coupled with exclusive legal right to possession. Co-ownership, however, means that more than one person has a legal interest in the same thing.

Personally Identifiable Information (PII): Any information about an individual maintained by an agency, including (a) any information that can be used to distinguish or trace an individual's identity, such as name, social security number, date and place of birth, mother's maiden name, or biometric records; and (b) any other information that is linked or linkable to an individual, such as medical, educational, financial, and employment information.

Privacy: The ability of an individual or group to seclude themselves, or information about themselves, and thereby express themselves selectively.

Privacy Concerns: Wherever PII or other sensitive information is collected, stored, used, and finally destroyed or deleted—in digital form or otherwise. Improper or nonexistent disclosure control can be the root cause for privacy issues.