# Connected Vehicle Pilot Deployment Program Phase 3

## Data Management Plan – Tampa (THEA)

www.its.dot.gov/index.htm

**Final Report — March 2021**

**Publication Number: FHWA-JPO-17-462**

U.S. Department of Transportation

Produced by Tampa Hillsborough Expressway Authority (THEA)
U.S. Department of Transportation
Intelligent Transportation Systems (ITS) Joint Program Office

# Notice

**Technical Report Documentation Page**

| 1. Report No. | 2. Government Accession No. | 3. Recipient's Catalog No. |
|---|---|---|
| FHWA-JPO-17-462 | | |

| | |
|---|---|
| **4. Title and Subtitle**<br>Connected Vehicle Pilot Deployment Program Phase 3, Data Management Plan – Tampa (THEA) | **5. Report Date**<br>02/2021<br>**6. Performing Organization Code** |
| **7. Author(s)**<br>Sisinnio Concas, Steve Johnson, Steve Novosad, Dave Miller | **8. Performing Organization Report No.** |
| **9. Performing Organization Name And Address**<br>Tampa Hillsborough Expressway Authority<br>1104 East Twiggs Street, Suite 300<br>Tampa, Florida 33602 | **10. Work Unit No. (TRAIS)**<br>**11. Contract or Grant No.**<br>DTFH6116RA0007 |
| **12. Sponsoring Agency Name and Address**<br>ITS-Joint Program Office<br>1200 New Jersey Avenue, S.E.,<br>Washington, DC 20590 | **13. Type of Report and Period Covered**<br>Final Report<br>**14. Sponsoring Agency Code** |

**15. Supplementary Notes**
Govind Vadakpat, Tampa CV site Agreement Officer's Representative and Sarah Tarpgaard, Agreement Officer

**16. Abstract**
The Tampa Hillsborough Expressway Authority (THEA) Connected Vehicle (CV) Pilot Deployment Program is intended to develop a suite of applications that utilize vehicle-to-infrastructure (V2I) and vehicle-to-vehicle (V2V) communication technology to reduce traffic congestion, improve safety, and decrease emissions. These CV applications support a flexible range of services from advisories, roadside alerts, transit mobility enhancements, and pedestrian safety. The pilot is conducted in three phases. Phase 1 includes the planning for the CV pilot including the concept of operations development. Phase 2 is the design, development, and testing phase. Phase 3 includes a real-world demonstration of the applications developed as part of this pilot. This document represents the Data Management Plan. The Data Management Plan is intended to describe how data will be collected, managed, integrated, and disseminated before and during Phase 3.

| 17. Key Words | 18. Distribution Statement |
|---|---|
| Applications, Connected Vehicles, Data, Performance Measures, Design Deployment, Mobility, Pilot, Safety, Software | |

| 19. Security Classification. (of this report) | 20. Security Classification. (of this page) | 21. No. of Pages | 22. Price |
|---|---|---|---|
| Unclassified | Unclassified | 56 | |

**Form DOT F 1700.7 (8-72)** **Reproduction of completed page authorized**

# Contents


**1 Introduction** ..... **1**
1.1 Study Area ..... 1
1.2 Purpose of the Plan ..... 3
1.3 Organization of the Plan ..... 4
**2 Data Management Approach** ..... **5**
2.1 Data Sources ..... 5
2.1.1 CV Data ..... 5
2.1.2 Non-CV Data ..... 6
2.1.3 Participants Surveys ..... 6
2.2 Data Sharing ..... 7
2.2.1 ITS DataHub ..... 7
2.2.2 Data Sharing with Independent Evaluators ..... 7
2.2.3 THEA Master Server ..... 8
2.2.4 CUTR Server ..... 8
2.2.5 Other Data Sharing ..... 9
2.3 Data Privacy ..... 10
2.3.1 Key Privacy Terms ..... 10
2.3.2 Collected PII/SPII Data Categories ..... 12
2.3.3 Controls ..... 13
2.3.4 Intellectual Property Issues ..... 16
2.4 Data Quality Control ..... 16
2.5 Archiving and Preservation ..... 18
**3 Application Data Management Plan** ..... **21**
3.1 Application Overview ..... 21
3.1.1 End of Ramp Deceleration Warning (ERDW) ..... 21
3.1.2 Wrong Way Entry (WWE) ..... 23
3.1.3 Vehicle Turning Right in Front of Transit Vehicle (VTRFTV) ..... 26
3.1.4 Transit Signal Priority (TSP) ..... 27
3.1.5 Forward Collision Warning (FCW) ..... 29
3.1.6 Emergency Electronic Brake Light Warning (EEBL) ..... 30
3.1.7 Intersection Movement Assist (IMA) ..... 31
3.1.8 Intelligent Signal System (I-SIG) ..... 32
3.1.9 Pedestrian Collision Warning (PCW) ..... 33
**4 Data Collection Analysis** ..... **35**
4.1 BSM Data Generation Simulation ..... 35
4.1.1 Refinement to the Baseline Scenario ..... 36
4.1.2 Simulation Results ..... 38
4.2 Data Collection Approach ..... 39
4.2.1 Data Collection Within the Study Area ..... 39
4.2.2 Data Collection Outside the Study Area ..... 39
4.2.3 BSM Data Collection Bottleneck Approach ..... 40
4.2.4 Alert and OBU Data Log Collection ..... 40

**References ........................................................................................................................ 42**

# List of Figures

Figure 1. CV Pilot Deployment Overview................................................................................ 2
Figure 2. THEA CV Pilot Planned Data Flow.......................................................................... 3
Figure 3. Data Sharing Flow from CUTR Server to USDOT .................................................. 9
Figure 4. Exit Curve of the Reversible Express Lanes ......................................................... 22
Figure 5. WWE Functional View ............................................................................................ 24
Figure 6. WWE Physical Overview ........................................................................................ 25
Figure 7. Functional View of VTRFTV.................................................................................... 27
Figure 8. Functional View of TSP............................................................................................ 28
Figure 9. Functional View of FCW .......................................................................................... 30
Figure 10. Functional View of EEBL ...................................................................................... 31
Figure 11. Functional View of IMA.......................................................................................... 32
Figure 12. Functional View of PCW ........................................................................................ 34
Figure 13. Study Area Parking Lots ........................................................................................ 37
Figure 14. Estimated Data Storage at Master Server............................................................ 39

# List of Tables

Table 1 Controls for Protection of Data Classes Used in the CV Pilot ................................ 15
Table 2. CV Pilots SAE J2735 Message Sets......................................................................... 21
Table 4. Baseline BSM Data Estimation Scenario.................................................................. 35
Table 5. Travel Times to/from RSUs........................................................................................ 37
Table 6. Simulation Results ..................................................................................................... 38
Table 7. OBU Message and Alert Data Priority...................................................................... 41

# List of Acronyms

| ACRONYM | DEFINITION |
|---|---|
| ADP | Application Deployment Plan |
| AP | Authority and Purpose |
| API | Automated Program Interface |
| AR | Accountability, Audit, and Risk Management |
| AVL | Automated Vehicle Location |
| BSM | Basic Safety Message |

| ACRONYM | DEFINITION |
|---|---|
| CBD | Central Business District |
| CCB | Change Control Board |
| CIA Triad | Confidentiality, Integrity, and Availability |
| ConOps | Concept of Operations |
| CU | Controller Unit |
| CUTR | Center for Urban Transportation Research |
| CV | Connected Vehicle |
| CVRIA | Connected Vehicle Reference Implementation Architecture |
| Detector | Infrastructure device that senses moving objects |
| DI | Data Quality and Integrity |
| DM | Data Minimization and Detection |
| DMP | Data Management Plan |
| DPP | Data Privacy Plan |
| DSRC | Dedicated Short Range Communications |
| EEBL | Emergency Electronic Brake Light |
| ERDW | End of Ramp Deceleration Warning |
| FCW | Forward Collision Warning |
| GIS | Geographic Information System |
| GPS | Geographic Positioning System |
| HART | Hillsborough Area Regional Transit |
| HMI | Human Machine Interface |
| HUA | Human Use Approval |
| ICD | Interface Control Document |
| IE | Independent Evaluator |
| IMA | Intersection Movement Assist |
| InCD | Informed Consent Document |
| IP | Intellectual Property |
| IRB | Institutional Review Board |
| I-SIG | Intelligent Signal Systems |
| ISM | Infrastructure Safety Message |
| ITS | Intelligent Transportation System |
| MAP | Map Message |
| MPH | Miles Per Hour |
| NARA | National Archives and Records Administration |
| NIST | National Institute of Standards and Technology |
| NTCIP | National Transportation Communications for Intelligent Transportation System Protocol |
| OBU | On-Board Unit |
| OTA | Over-the-air |
| PCW | Pedestrian Collision Warning |

| ACRONYM | DEFINITION |
|---|---|
| PID | Personal Information Devices |
| PII | Personally Identifiable Information |
| PMWS | Pedestrian Mobility Warning System |
| Proxy | Software application that converts Detector output to BSM based on detection zone |
| PSM | Personal Safety Message |
| PTMW | Pedestrian Transit Movement Warning |
| REL | Reversible Express Lanes |
| RSA | Roadside Alert |
| RSU | Roadside Unit |
| RTVM | Requirements Traceability Verification Matrix |
| SAD | System Architecture Document |
| SCMS | Security Credential Management System |
| SDC | Secure Data Commons |
| SE | Security |
| SMOC | Security Management Operating Concept |
| SPaT | Signal Phase and Timing Message |
| SPII | Sensitive PII |
| SRM | Signal Request Message |
| SSM | Signal Status Message |
| SSN | Social Security Number |
| SyRS | System Requirements Specification |
| TB | Terabyte |
| THEA | Tampa Hillsborough Expressway Authority |
| TIM | Traveler Information Message |
| TMC | Traffic Management Center |
| TMDD | Traffic Management Data Dictionary |
| TSP | Transit Signal Priority |
| UL | Use Limitation |
| USDOT | United States Department of Transportation |
| V2I | Vehicle-To-Infrastructure |
| V2V | Vehicle-To-Vehicle |
| V2X | Vehicle-To-Everything |
| VIN | Vehicle Identification Number |
| VM | Virtual Machine |
| VTRFTV | Vehicle Turning Right in Front of a Transit Vehicle |
| WWE | Wrong Way Entry |

# 1 Introduction

The Tampa Hillsborough Expressway Authority (THEA) Connected Vehicle Pilot Deployment (CV Pilot) is one of the three CV projects selected by the Federal Highway Administration (FHWA) in September 2015 as part of a U.S. Department of Transportation (USDOT) funded program. The deployment consists of three phases: (1) Concept Development, (2) Design-Build-Test, and (3) Operations and Maintenance (O&M). The Pilot identified areas of traffic management in Tampa, Florida, that may be improved by the deployment of CV applications. The project team developed a system concept for deploying these CV applications and after approval by USDOT, designed, deployed, and currently operates the system.

The deployment includes several CV applications, deployed across highway, transit, and pedestrian modes of transportation on a variety of facility and vehicle types. This Pilot aims to create a connected urban environment to measure the effect and impact of CVs in Tampa's vibrant downtown.

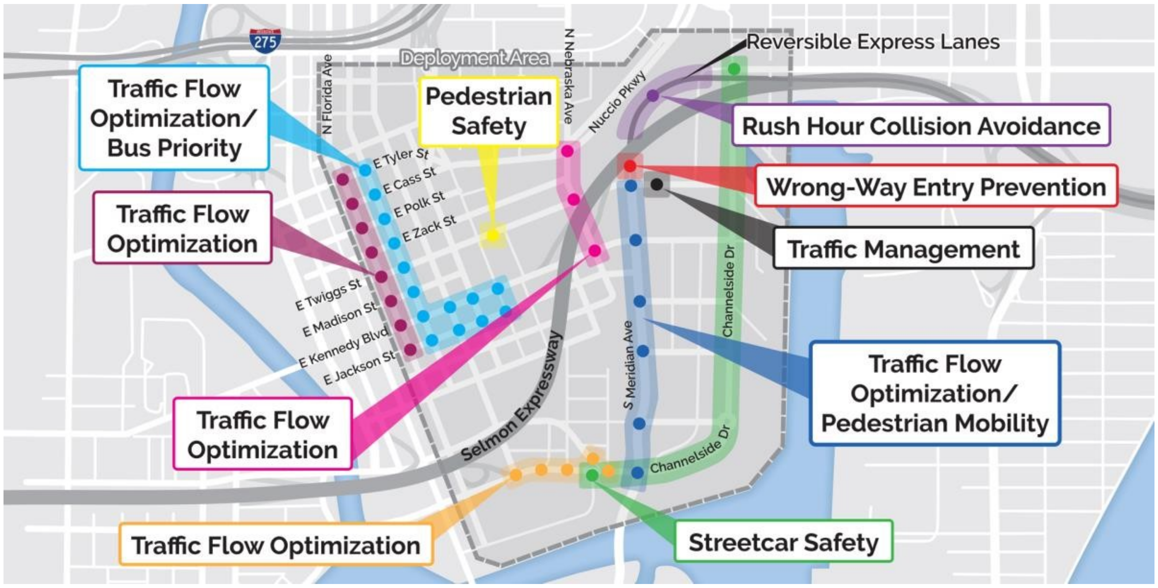
This documents details the THEA CV Pilot data management plan (DMP). The DMP is based on data management plan guidance for extramural researchers established by the United States Department of Transportation (USDOT), the Cooperative Agreement for the CV Pilot Deployment Program, guidance provided by USDOT for the Smart Cities Program, and data management plan guidance provided by USDOT. This update to the DMP reflect operational and deployment changes that occurred during Phase 3. Therefore, this update removes all data elements that were planned to be collected and shared during Phase 1 and Phase 2 as related to applications that were not either implemented or modified during the last phase of the deployment.

This section provides a short CV Pilot overview, discusses the purpose of the DMP, and details how the DMP is organized.

## 1.1 Study Area

The Pilot deploys CV applications in Tampa's Central Business District (CBD) and environs to create a more connected downtown. Downtown Tampa is bordered by Ybor Channel (Cruise Ship and Commercial Port Channel) to the east, Garrison Channel (local waterway) to the south, Florida Avenue to the west, and Scott Street to the north. A virtually flat topography near sea level helps to simplify the evaluation of traffic flow parameters.

Figure 1 depicts the focused pilot area and deployment of the thirteen THEA CV pilot applications. Additional information on the applications can be found in the THEA CV Application Deployment Plan.




*Source: HNTB, 2017*

**Figure 1. CV Pilot Deployment Overview**

**V2I Safety** – V2I safety applications wirelessly exchange critical safety and operational data between vehicles, roadway infrastructure, and personal information devices to help avoid motor vehicle crashes. V2I safety applications will complement V2V safety applications, enabling vehicles to have a 360-degree awareness and inform drivers through advisories and warnings of hazards and situations they cannot see. The THEA CV pilot team plans to deploy the following V2I applications:

- Pedestrian Collision Warning (PCW)
- Wrong Way Entry (WWE)

**V2V Safety** – V2V safety applications wirelessly exchange data among vehicles traveling in the same vicinity to offer significant safety improvements. Each equipped vehicle on the roadway – including automobiles, trucks, transit vehicles, and motorcycles – will be able to communicate with other vehicles. This rich set of data and communications will support a suite of active safety applications and systems. Vehicles will communicate with one another broadcasting basic safety messages (BSMs) that will inform drivers of hazards and situations they cannot see. These applications will only function when the involved vehicles are both equipped with V2V devices. The THEA CV Pilot team plans to deploy the following V2V safety applications:

- Forward Collision Warning (FCW)
- Emergency Electronic Brake Light (EEBL)
- Intersection Movement Assist (IMA)
- Vehicle Turning Right in Front of Transit Vehicle (VTRFTV)

**V2I Mobility** – V2I mobility applications communicate operational data between vehicles and infrastructure, intended primarily to increase mobility and enable additional safety, mobility, and environmental benefits. Applications may use real-time data to increase safety and operational efficiency while minimizing the impact on the environment and enabling travelers to make better-informed travel decisions. The THEA CV Pilot team plans to deploy the following V2I mobility applications:

- End of Ramp Deceleration Warning (ERDW)
- Intelligent Traffic Signal Systems (I-SIG)
- Transit Signal Priority (TSP)

**Agency Data** – Agency Data applications use probe data obtained from equipped vehicles along the corridor to support Traffic Management Center (TMC) operations. Vehicle data can be used to detect changes in vehicle speeds indicating congestion or a disruption of traffic flow as well as calculate travel times. When a TMC notices a slow down on a corridor, the TMC may decide to act by, for example, altering signal timing based on traffic flows.

All of the above applications transmit and receive data (i.e., message sets). Each vehicle Onboard Unit (OBU) will save data it generates and transmits (e.g., Traveler Information Message [TIM]), save events (e.g., FCW, EEBL) that occur, and save BSMs it receives. V2V and V2I applications store the data on OBU storage and download the data to the Roadside Units (RSU) over-the-air (OTA) as the vehicle passes an RSU. An RSU will save data it generates and transmits (e.g., Infrastructure Safety Messages [ISMs]), save data it receives (e.g., BSMs, PSMs), and OBU data downloads. As the OBU sends its data to the RSU, it deletes the data from its storage. The RSU will perform a similar process sending the data to the Master Server over the network between the RSU and Master Server. As the RSU sends its data, the data is deleted from its storage. The Master Server is the home for all data collected for the CV Pilot. The Master Server software archives and transmits the data to the Center for Urban Transportation Research (CUTR) performance measurement team so that data can be scrubbed and sent to the USDOT data sharing platforms. CUTR is responsible for making data available to the Secure Data Commons (SDC) and ITS DataHub. The flow of data generated throughout the pilot is illustrated in Figure 2.

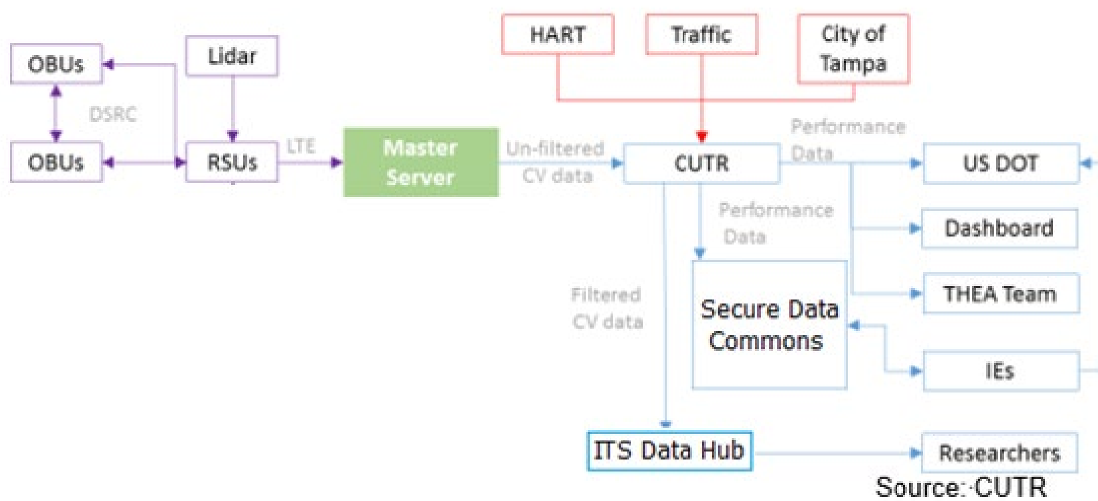



**Figure 2. THEA CV Pilot Planned Data Flow**

## 1.2 Purpose of the Plan

The purpose of the DMP is to document the types of CV Data that will be used and/or stored within the system and detail how the CV Data will be created, captured, transmitted, maintained, accessed, shared, secured and archived. These data include real-time and archived data that are used to control or are generated by systems that are managed by THEA or the City of Tampa.

The DMP serves as an operational guide for managing data collectively as a strategic asset of THEA. This plan details how and where data will be shared, subject to applicable privacy, security and other safeguards, and how data will be made available to others to enable performance measurement and support independent evaluation.

The DMP documents the flow of data from generation through its use to applications in the pilot deployment, including:

- Data sources and destinations
- Volume of data flow (currently under analysis)
- Contents of data flow

The DMP is intended to provide clear operational procedures consistent with data-related elements of multiple deliverables:

- Data Privacy Plan (DPP) [1]
- Security Management Operating Concept (SMOC), including a Privacy Operational Concept
- Performance Measurement and Evaluation Support Plan (PMESP) [2]
- System Requirements (SyRS) [3]
- Human Use Approval Summary (HUA), including feedback from a participant data collection and use-related Institutional Review Board (IRB)
- Application Deployment Plan (ADP) [4]
- System Architecture Document (SAD) [5]

The DMP provides an assessment of the variety, volume, and velocity of deployment-related data that can be accommodated to ensure the end-to-end delivery of data to all identified recipients/users. The DMP establishes CV data quality control procedures, and in cases where data includes Personally Identifiable Information (PII) or other restrictions, the DMP relies on the strict handling of PII detailed in the DPP [1].

## 1.3 Organization of the Plan

The DMP is organized into the following major sections:

- Chapter 1. Introduction
- Chapter 2: Data Management Approach
- Chapter 3: Application Data Management Plan
- Chapter 4: Data Collection Analysis
- References

The Data Management Approach section discusses the approach for managing data from the CV applications, taking into consideration Data Sharing, Data Privacy, Intellectual Property Issues, Data Quality Control, and Data Archiving.

The Application Data Management Plan section describes the data generated, received, and stored for each CV Pilot application. This data consists of not only CV data, but traditional Intelligent Transportation System (ITS) data as well (e.g., traditional detection devices). The section discusses TMC systems/software, data quality control, and data storage. The Data Collection Analysis section details a simulation of the amount data generated by the Pilot.

# 2 Data Management Approach

This section describes the team's general approach to data management and data-management plans that typically apply across all or most applications. The sub-sections include: Data Sharing, Data Privacy, Intellectual Property Issues, Data Quality Control, and Data Archiving and Preservation. A critical starting point for the management of data is the initial recognition and statement of ownership. All data have a specific or implied owner, and that owner must be respected in the collection, storage, and use or sharing of that data. It has already been determined that the data generated, transmitted, collected and stored is owned by the vehicle owner/participant. Thus, any sharing of the data must be pre-approved by the individual owner/participant in addition to the Salus IRB. Likewise, the purpose, use and sharing of data must not exceed the parameters under which permission was granted. This allowed purpose, use and sharing is described in the informed consent documents (InCDs), whereby the participant grants their pre-approved permission.

The overarching approach to data management for the THEA CV Pilot is to affect the best possible use of the data within the confines of the InCD parameters. The THEA Pilot Team will implement safeguards to ensure that data management remains within the allowed parameters of the InCDs and experimental protocol as approved by Salus IRB. The following sections describe the principles to be followed for data sharing, data privacy, data quality control, data archiving and preservation of data.

## 2.1 Data Sources

### 2.1.1 CV Data

The Tampa CV Pilot generates several datasets from the interaction between vehicles (via OBUs) and between vehicles and infrastructure (OBU/RSU interaction). Vehicles traveling or operating (i.e., public transportation and participant vehicles) generate data in the form of BSMs, which are collected by RSUs and transferred over the air to THEA's secured master server. At the same time RSUs broadcast relevant information to the OBUs. All the CV data are planned to be shared with USDOT following the sanitization protocols detailed in this document along metadata as detailed in Appendix C of the PMESP [2].

#### *2.1.1.1 RSU Data*

RSUs transmit and/or collect the following data:

- BSMs from the participant and public transit vehicles (up to 10Hz). These are also called "sniffed" BSMs or BSM collected by vehicle operating in range of an RSU.
- Signal Phase and Timing Messages (SPaT) from RSUs (10Hz).
- MAP Data Message (MAP) from RSUs (1Hz).
- Traveler Information Message (TIM) from RSUs at 1Hz .
- Signal Request Message (SRM) transmitted by OBUs within the range of Dedicated Short-Range Communication radio of an RSU.

- Signal Status Message (SSM) broadcasted by RSUs for conveying back to OBUs the status of its SRM.
- Pedestrian Safety Message (PSM) that triggered the collision alert.

#### *2.1.1.2 OBU Data*

Public transportation and participant vehicles record all received and transmitted data from the interaction with nearby vehicles and RSU in range via a OBU Data Log recording protocol. OBU Data Logs contain various data elements falling into one of the following categories:

- WAVE Short Message Protocol (WSMP) messages sent or received.
- Warnings issued to the driver.

Driver warning event records are created whenever one of the applications triggers a warning. The OBU creates a unique warning ID used to identify multiple Warning Event Data records belonging to the same warning event. As detailed in the Performance Measurement Evaluation Support Plan (PMESP), the OBU creates a set of Warning Event Data records per warning. Each record of the set represents a point in time before, during, and after the warning triggered [2].

### 2.1.2 Non-CV Data

The performance evaluation and measurement support team collected data from other sources for the purposes of data integration and to measure observable confounding factors. The following datasets were collected throughout the deployment:

- Bluetooth reader data
- Transit GTFS feeds data from Hart OneBusAway API (Application Programming Interface)
- Weather event data

All the above data are planned to be shared with the USDOT Independent Evaluators (IEs) following the metadata structure detailed in Appendix C of the PMESP [2].

### 2.1.3 Participants Surveys

In coordination with USDOT IEs, CUTR designed a series of surveys to be administered throughout the deployment. The first survey (initial survey) is administered at the installation appointment to collect information about the participants' travel preferences to the study area, their knowledge about CV technologies and to gauge the main reasons for joining the study. An exit survey is administered throughout Phase 3 to collect data as participants dropped from the study. The goal of this survey is to obtain information on the exit motivations and assess participant attrition.

Two separate surveys are administered to obtain information on the participant's experience with the CV technologies and equipment, and for those exposed to the warnings via the Human Machine Interface (HMI), to obtain information on how they perceive and respond to the warnings. One survey is administered at about midpoint of Phase 3, and one at towards the end of Phase 3. These two survey instruments contain the same questions on the CV application effectiveness to allow combining or comparing responses between waves.

## 2.2 Data Sharing

A major objective of the CV Pilot Program is to generate data that can be widely used for further research on connected vehicle applications and deployments, including other CV Pilot projects, and early CV deployers. This section describes the approach to data sharing, including archives and institutional relationships.

### 2.2.1 ITS DataHub

The United States Department of Transportation (USDOT) ITS DataHub is a publicly available data sharing site. Connected vehicle, mobile device, and infrastructure sensor data captured during deployment are expected to be broadly shared with the community. However, data sharing is subject to the protection of intellectual property rights and personal privacy and must be handled securely. Appropriately prepared system control, performance and evaluation data, stripped of PII, will be shared with the USDOT and posted to the ITS DataHub. Data are stored in the CUTR Server with a documented data structure. Data posted to the ITS DataHub will be documented, such that they are easily discoverable via a search function on the site. The CUTR Servers also handles any stripping of PII from data collected before they are uploaded to the ITS DataHub. The cleansed data is stored in a local directory on the CUTR Server.

Upload of the cleansed data to the ITS DataHub uses the established ITS DataHub data transfer methodology. Developing a transfer methodology is outside the scope of the project. The Data Schema and interface for transferring performance measurement information to the USDOT ITS DataHub was agreed during the Detailed Design phase.

THEA expects to work closely with the USDOT to ensure that data produced during the demonstration is shared efficiently and cost effectively, leveraging these and other shared resources as appropriate to increase the completeness and timeliness of data exchange. Data exchanges are anticipated to be manual exchanges requiring human intervention by manual steps in order to ensure and double check that the data is sufficiently cleansed.

### 2.2.2 Data Sharing with Independent Evaluators

Collected data will also be shared with the Independent Evaluators (IEs) via the USDOT Secure Data Commons (SDC) data archiving and sharing platform. The data uploaded to the SDC will include three broad types of information. Safety, mobility and survey result data will be provided at the frequency and in the format that was agreed upon in discussions with the Tampa team, USDOT and the IEs during the IE site visit. Safety and mobility measures and data are expected to be provided weekly with survey results provided as they are administered and completed.

No PII is sent to the SDC. Methods for trip identification, vehicle recognition and survey participant consistency within the Tampa CV Study Area have been proposed and are being developed to ensure that IEs receive useful and meaningful research information. The transfer of the data will be included in the Detailed Design. The data requirements and data formats are being developed by the USDOT to serve all three sites and the entities that will act as independent evaluators for the USDOT.

The IEs will use data collected before and during deployment to derive quantitative and qualitative measures of system impact.

### 2.2.3 THEA Master Server

THEA's Master Server receives the data sent from the CV deployed devices. The Master Server handles data sanitation and stores the following sanitized messages:

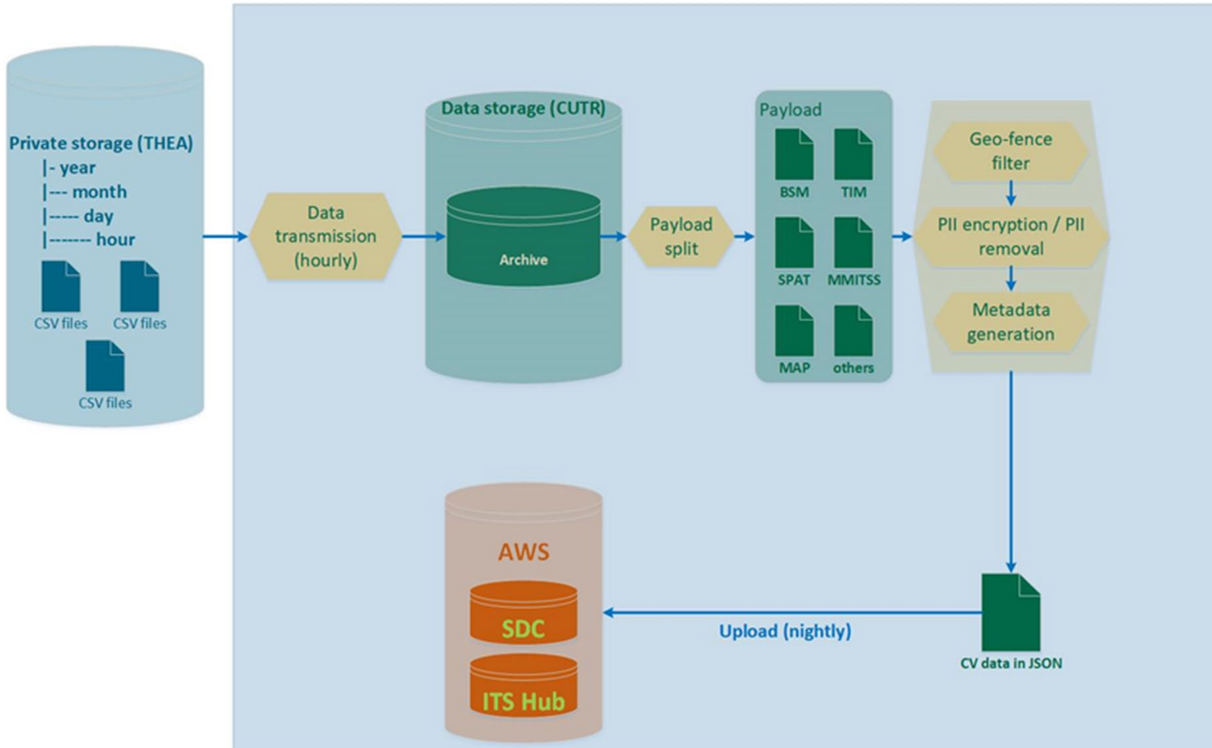
- Basic Safety Messages (BSMs)
- OBU Data Logs
- Map Message (MAP)
- Signal Phase and Timing Message (SPaT)
- Signal Request Message (SRM), Transit Signal Priority Request
- Signal Status Message (SSM), Transit Signal Priority Status
- Personal Safety Message (PSM)
- Traveler Information Message (TIM)

The data fields of the message were defined during Phase 2 and detailed in the PMESP [2]. The details of data cleansing are discussed in the THEA CV Pilot Data Privacy Plan.

### 2.2.4 CUTR Server

CUTR will maintain a server that will house data from the THEA Master Server and other sources for processing, analysis and dissemination to USDOT. In addition to data mentioned above, CUTR will be collecting and maintaining information on weather, incidents, Hillsborough Area Regional Transit (HART) schedule changes, transit incident reports, and survey results. Information will be provided to the IE via the SDC from the CUTR Server. In order to protect PII participant identifiers, such as "participant number," will be removed prior to dissemination.

The CUTR server has been deployed and configured to receive and archive all CV and non-CV data. CV Data are stored as highly compressed flat files in THEA master server protected storage. As shown in Figure 3, the CV data are first split by payload (e.g., BSMs, TIMS, SPaTs), then subject to PII removal, and finally repackaged in a file format suitable for upload to the SDC and ITS Data Hub.




*Source: CUTR*

**Figure 3. Data Sharing Flow from CUTR Server to USDOT**

The SDC developers have designed and deployed a set of data governance procedures to manage access data generated and uploaded by the USDOT CV Pilots. The SDC developed and implemented a "New Dataset Provider Form," which has been provided to the CV Pilot Team. The form allows the data provider specifying the information and access level for each data field. In coordination with the SDC and ITS Data Hub developers, data transmission is carried in nightly batches. Each night at 1:15 A.M. EST, the CUTR server processes and uploads CV data generated during the previous day. The upload schedule contains provisions to re-upload data that are not transmitted during the previous period due to technical issues encountered either by the CUTR server or the SDC and ITS Data Hub platforms.

The flow of data between servers follows security protocols detailed in the Data Privacy Section and conforming to data transmission security protocols established by the SDC and ITS Data Hub developers.

## 2.2.5 Other Data Sharing

No direct release of Pilot data shall be made outside of those platforms discussed above and no uncleansed (non-filtered) data will be released outside of the THEA CV Pilot Team. The THEA CV Pilot Team retains no control over, or responsibility for the accuracy or integrity of data once released to the ITS DataHub, SDC, and/or other archival systems.

## 2.3 Data Privacy

A key consideration in providing data from the CV Pilot program is to ensure that proprietary data or data containing PII is not released to the public. If any data containing PII is generated during the operation of the program, the provider must ensure that PII is removed before the data are shared with the public and internal team members, if not explicitly authorized by the DPP and approved Salus IRB associated documents. These documents include at a minimum the research protocol and informed consent documents.

The THEA CV Pilot DPP provides details and procedures for the handling of PII associated with the collection and analysis of data required for the CV Pilot. The Phase I Security Management Operating Concept (SMOC) provides principles for a holistic approach to security in general, whereas the DPP provides the physical, administrative and technical controls to be implemented to ensure the protection of PII (Data Privacy). The Crash Avoidance Metrics Partnership will operate the Security Certificate Management System (SCMS) for the CV Pilot. This system is also discussed in the DPP. Sections 1.1.9 – 1.1.12  provide an overview of the major areas of Data Privacy regarding the Data Management Plan. For detailed specifics on the procedures and processes to be followed in protecting privacy, please refer to the DPP.

Security of the data system that is used to centralize public documents may have minimal security requirements, whereas a system that stores personal data would require a much higher level a security. Please refer to SMOC and DPP for guidance on securing data and protecting data privacy.

### 2.3.1 Key Privacy Terms

**Data Owner:** by default, the owner is the subject of the PII Data. Informed consent of the owner must be acquired and documented prior to collection of PII. For the CV Pilot, the data owner shall be the vehicle owner who registers for participation in the Pilot. Owner/registrants receive training and sign informed consent documents to ensure they fully understand:

- The data to be collected
- The purpose for which the data will be used
- Their rights as the data owner
- Their protection from disclosure of Sensitive PII (SPII)/PII
- Any additional entities with whom the data may be shared during/after the Pilot

The owner/registrant is the participant and is responsible for informing other persons whom they may allow to operate their vehicle during the Pilot.

In the case of transit buses and streetcars, HART, as the owner of the vehicles, shall sign an InCD. An initial HART InCD was approved by the IRB (June 2016). While HART vehicles will be operated by HART employees, these employees, although not technically participants, will receive training and be extended the oversight for safety and safety equipment reviews that other participants will receive. If an incident occurs, standard HART procedures will be followed and the Pilot Safety Manager will be informed, who will then follow the standard operating procedure outlined in the Phase I Safety Management Plan [6]. No PII data will be collected on the HART drivers, except that which HART already possesses as the employer, so there will be no danger to their PII in the Pilot database. Raw CV data will not be available to HART to avoid the possibility of monitoring individual drivers. However, HART already has access to Geographic Positioning System (GPS) and Automated Vehicle Location

(AVL) equipment that is currently in use pursuant to HART's union contract (Hillsborough Area Regional Transit and Amalgamated Transit Union Local 1593, October 1, 2012). (See also Section 5.1.)

**Privacy:** control over the extent, timing, and circumstances of sharing oneself (physically, behaviorally, or intellectually) with others.

**Personal Identifiable Information (PII):** Information that can be used to distinguish or trace an individual's identity, such as their name, Social Security number, biometric records, etc., alone, or when combined with other personal or identifying information, which is linked or linkable to a specific individual, such as date and place of birth, Mother's maiden name. The definition of PII is not anchored to any single category of information or technology. Rather, it requires a case-by-case assessment of the specific risk that an individual can be identified by examining the context of use and combination of data elements. Non-PII can become PII whenever additional information is made publicly available. This applies to any medium and any source that, when combined with other available information, could be used to identify an individual.
**SPII:** is a subset of PII, which if lost, compromised or disclosed without authorization, could result in substantial harm, embarrassment, inconvenience, or unfairness to an individual. Sensitive PII requires stricter handling guidelines because of the increased risk to an individual if the data are compromised. The following PII is always (de facto) sensitive, with or without any associated personal information:

- Social Security number (SSN)
- Passport number
- Driver's license number
- Vehicle Identification Number (VIN)
- Biometrics, such as finger or iris print
- Financial account number such as credit card or bank account number
- The combination of an individual identifier and date of birth, or mother's maiden name, or last four digits of an individual's SSN

In addition to Sensitive PII, some non-sensitive data may be deemed sensitive based on context, as discussed next.

**Non-sensitive Data as PII:** some information may be non-sensitive or anonymous by itself, but when coupled with other available or discoverable data, can become PII and even SPII.

**CV Data:** for the purposes of this document, CV Data shall mean data collected from the system pertaining to functional performance of the various components, devices and communications; and necessary for the CV Applications to successfully operate as intended. This also includes BSMs, TIMs, MAPs and SPaT data, which will generally NOT contain ANY "direct" PII or SPII. However, as non- sensitive data can be utilized to extrapolate PII, a "scrubbing" process will be applied to filter against this possibility before CV Data is shared on the RDE, the CV PEP or other portals.

**Published Data:** industry guidelines establish that data which is "published" is de facto "public data" and, therefore, has no expectation of privacy or protection from exposure/exploit. The (Official (ISC) Guide to the CISSP CBK, Fourth Edition, 2015) includes "broadcast communications" in the definition of published data [7]. CV Pilot data that is "Broadcast" over DSRC channels is protected by the SCMS system and not publicly available from the site operator. As such, broadcast live CV Data is not to be construed as "published" or "public" within that context.

**The CIA Triad:** is the term for the "Big 3" tenets of information security – Confidentiality, Integrity and Availability as defined below:

- Confidentiality: Prevention of intentional or unintentional unauthorized disclosure of data
- Integrity:
  - Prevention of modification by unauthorized persons or processes
  - Prevention of unauthorized modification by authorized personnel or processes
  - Ensuring that data is internally and externally consistent
- Availability: Ensuring the timely and reliable access to data by appropriate personnel

Each category/class of data outlined in the DPP will be assessed following the principles of the CIA triad[7]. This will include, at a minimum, considering the vulnerability of each class/category of data regarding confidentiality, integrity and availability. This assessment has been completed and the results are documented in the DPP and repeated herein. Table 1 (Section 2.3.3.3) displays the results of the assessment and shows the control(s) to be applied for each class/category of data. The assessment complies with NIST Publication FIPS SP800-53 [8].

**Access Control Terms:** Identification: the means by which users claim their identities to a system. Identity is a required precursor to authentication and authorization.

- Authentication: the testing or reconciliation of evidence of a user's identity. It establishes and verifies that a user is who they say they are.
- Authorization: the rights and privileges granted to a person or process.
- Accountability: the processes and procedures by which a system obtains its ability to determine the actions and behavior of a single individual or process within the system and to identify that individual person or process. Audit trails and logs are examples of tools supporting accountability.

### 2.3.2 Collected PII/SPII Data Categories

Data to be collected by the CV Pilot may include, based on motorist participants, many of the following forms of personal information about individual participants and their motor vehicle and motor vehicle use. The following data represent the minimum amount of data required for the research to be effective and statistically relevant. The occupation/affiliation-type data was requested by the IE for socio-demographic analysis, and also supports the provision of anonymized data. This type of data, along with all data obtained from surveys and interviews will be distributed only in combined socio-demographic-type reporting and will not be individually specific.

Participant Background Information (All Participants)

- Individual Identifiers;
- Full Name (First, Middle, Last);
- Socio-demographic information, including age, gender, marital status, and income;
- Driver's license number, issuing state, and qualifiers.

Vehicle Identifiers (Driver Participants Only)

- Personal VIN and registration information;

- VIN of government issued vehicles; and,
- Identifiers for equipment installed by Pilot in personal or participant vehicle.

Contact Information (All Participants)

- Mailing/Residential Address;
- Phone number(s);
- Email address(es);
- Institutional or organizational affiliation;
- Work/Business related contact information; and,
- Occupation and work schedule.

Eligibility Information (Driver Participants Only)

- Driver history and habits;
- Proof of insurance;
- Proof of Florida vehicle registration; and,
- Completion of Pilot participant training.

Project Information (Driver Participants Only)

- Vehicle sensor information;
- Dynamic information about a vehicle, including location, heading, velocity, proximity to and interaction with other vehicles and infrastructure; and,
- Data collected from drivers by means of surveys, focus groups, or interviews.

Some data categories are collected for validating the statistical sample and are not specifically used for performance measurement directly. This type of data obtained from surveys and interviews will be distributed only in combined socio-demographic-type reporting and will not be individually specific. Examples are:

- Data related to occupation, age, driving history and habits
- Data collected from drivers by means of surveys, focus groups, or interviews

## 2.3.3 Controls

Security methodologies used to protect sensitive data are referred to as "Controls".

#### *2.3.3.1 Types of Controls*

Security controls can be classified by three types and three means. The three types of controls are:

- Preventive: Are put in place to "inhibit" harmful events.
- Detective: Are put in place to "discover" harmful events
- Corrective: Are put in place to restore systems after a harmful event.

These security controls follow a progression from blind optimism (believing that prevention will eliminate ALL negative events) to the sky is falling (we cannot stop them, better prepare to pick up the pieces). The best security plans utilize a balance of the available controls to accomplish the best solution based on multiple factors including:

- Risk tolerance of data owner
- Value of data at risk
- Damage expected from loss or exposure

- Likelihood of loss or exposure
- Cost of various safeguard options compared to the level of assurance they bring and the above factors.

The Pilot will identify and manage Security Controls following the steps recommended by NIST in its FIPS SP800-53 Document, and the requirements traceability verification matrix (RTVM) will be constructed around these steps:

- Categorize the information system based on a FIPS Publication 199 impact assessment; (partially completed by USDOT - pre-award, preliminary re-assessment based on current state of design at point of DPP creation, and another re-assessment to follow final design)
- Select the applicable security control baseline based on the results of the security categorization and apply tailoring guidance;
- Implement the security controls and document the design, development, and implementation details for the controls;
- Assess the security controls to determine the extent to which the controls are implemented correctly, operating as intended, and producing the desired outcome with respect to meeting the security requirements for the system;
- Authorize information system operation based on a determination of risk to organizational operations and assets, individuals, other organizations, and the Nation resulting from the operation and use of the information system and the decision that this risk is acceptable; and,
- Monitor the security controls in the information system and environment of operation on an ongoing basis to determine control effectiveness, changes to the system/environment, and compliance to legislation, Executive Orders, directives, policies, regulations, and standards.

#### *2.3.3.2 Means of Control*

There are also three means for implementing the first two types of controls:

- Administrative: Includes policies and procedures, security awareness training, background checks, and levels of supervision.
- Logical or Technical: Targets the restriction of access and includes encryption, smart cards, access control lists, and biometrics.
- Physical: Incorporates security guards, alarm systems, locks etc.

#### *2.3.3.3 Selected Controls by Data Class*

Table 1 details the Controls selected for protection of the data classes comprising the CV Pilot. The checkmark indicates that the control shown in the row is to be applied to the data class shown in the column. The checkmark has been removed for scrubbed data. “Need to know” is part of the authorization process to grant access. An individual does not gain access solely on the basis of clearance level, but also a valid role based on a need to know. Both components are required to be granted access. This a very basic principle of security, which is described in detail in the DPP. SP 800- 53 Table J-1 defines the legend for this column. The National Institute of Standards and Technology (NIST) publication provides a correlation between the ISC2 CIA assessment and the NIST assessment [8].

**Table 1 Controls for Protection of Data Classes Used in the CV Pilot**

| Data Type / Description Control Used (Type/Means) | NIST SP 800-53 Table J-1 Controls | Live CV Data (real-time data accessed in the field on OBUs, RSUs, or sniffers) | Stored CV Data, Raw | Stored CV Data, Scrubbed | CV Data of any type when in transit | PII/ SPII in any state | Hard Copy Participant Data | Electronic Participant Data |
|---|---|---|---|---|---|---|---|---|
| **SCMS Certificates/CRL (Preventive/ Technical)** | AR, DI, SE, DM | Ö | | | | | | |
| **Anonymity (Preventive/Technical)** | AR, DI | Ö | Ö | | | | | |
| **Encryption (Preventive/Technical)** | AR, DI, SE | Ö | Ö | | Ö | Ö | | Ö |
| **Access Control- Cabinet locks etc. (Preventive/Physical)** | AR, DI, SE, DM | Ö | Ö | Ö | Ö | Ö | Ö | Ö |
| **Access Control- remote to devices and via system (Preventive/Technical)** | AR, DI, SE, DM | Ö | Ö | Ö | Ö | Ö | | Ö |
| **Authorization - ID Based (Preventive/Technical)** | AR, DI,SE, UL, DM | | Ö | Ö | Ö | Ö | Ö | Ö |
| **Authorization - Role Based (Preventive/Administrative)** | AP, AR,DI, SE, UL, DM | | Ö | Ö | Ö | Ö | Ö | Ö |
| **Penetration Testing (Preventive/ Technical)** | DI, SE | Ö | Ö | Ö | Ö | Ö | | Ö |
| **System Monitoring(Detective/ Technical)[1]** | AR, DI, SE, UL, DM | Ö | Ö | Ö | Ö | Ö | | Ö |
| **Anti-Virus (Detective// Technical)[1]** | SE, DM | | Ö | Ö | Ö | Ö | | Ö |
| **Filtering/ Scrubbing (Preventive/Technical)** | AR, DI, SE, UL | | | Ö | | | | |
| **Need to Know (Preventive/ Administrative)** | AR, DI,SE, UL, DM | Ö | Ö | | Ö | Ö | | Ö |
| **Compartmentalization (Preventive/Administrative)** | AR, DI, SE, UL | Ö | Ö | | Ö | Ö | | Ö |
| **Internal Audits Detective/ Administrative)** | AR, DI,SE, UL, DM | Ö | Ö | | Ö | Ö | | Ö |
| **Independent Audits (IRB) (Detective/Administrative** | AR, SE, UL | Ö | Ö | | Ö | Ö | | Ö |

NIST SP 800-53 Table J-1 Privacy Controls by Family ID: AP, Authority and Purpose; AR, Accountability, Audit, and Risk Management; DI, Data Quality and Integrity; DM, Data Minimization and Retention; SE, Security; and, UL, Use Limitation.

1 This control can also be Corrective if enabled

### 2.3.4 Intellectual Property Issues

USDOT is sponsoring and funding the CV Pilot and, as such, Intellectual Property (IP) claims will be minimal and pertain to methods of collection, algorithms utilized for analysis and products or processes which are existing or patent pending. Any such claim will be noticed in writing to USDOT Agreement Officer and Agreement Officer Representative in a timely manner, but minimally prior to any release outside of the THEA team. In the absence of such notice, all data released for publication including data shared with the IE or uploaded to the ITS DataHub, will be considered as free of IP.

1. Right to manage the data. Raw CV Data collected and utilized within the Pilot is owned by the participant on whom the data is collected. The right to use, manage, analyze and report on this data shall be conferred upon the THEA Pilot Team via informed consent. No release of data in this unfiltered state shall be permitted.

2. Any such IP claim shall indicate who holds the intellectual property rights to the data.

3. Copyrights. Notwithstanding the IP discussed above, there will be no copyrights associated with the data distributed by the THEA CV Pilot. It is understood that the purpose of the Pilot is to contribute to data sharing and other efforts to study and advance CV Technology in the effort to improve safety, mobility, agency efficiency and a sustainable environment.

4. Rights be transferred to a data archive. Once the data has had all PII removed and been formatted for safe release to the IE and ITS DataHub and for other public use, the right to use and manage the resulting data is no longer held by the THEA Pilot. The data is then open for use in accordance with the repository or entity furthering its release. Upon this release of rights to manage, the THEA team can no longer warrant the quality or integrity of the data as it is no longer within their control.
5. For HART drivers, THEA will not be collecting personal information. For auto drivers, the Informed Consent Document states that records of you (the participant) will be kept private. The study investigator and staff, sponsor agency, USDOT, and its designees and the IRB will have access to study records. Personal information will not be used at scientific meetings or in publications. A court of law may request study records. Thus, total privacy cannot be guaranteed.

The InCD also States: Any data that personally identifies you or could be used to personally identify you will be held under a high-level of security at one or more data repositories. Your data will be identified with a code rather than your name. Only qualified researchers will be authorized to have access to data that personally identifies you, or can be used to personally identify you, and the level to which they have access will be based on their level of authorization.

## 2.4 Data Quality Control

The team will implement automated and manual data quality controls to ensure quality data that is consistent and complete. The routine measures to ensure data quality will include scheduled and unscheduled data quality audits to be performed by CUTR project team. The Data Custodian is the owner of the data quality process and responsible for overall data quality. The Data Custodian’s roles and responsibilities are defined as:

- Oversight of the data management process from data received at the Master Server to data shipped to CUTR.
- Periodic spot checking of the data.
- Oversight of data backups and archives
- Implementing the data quality controls
- Implementing the Data Management Plan (This document).

HNTB, THEA's General Engineering Consultant (GEC) consultant and program management lead for the pilot creates a Project Quality Plan for every project and assigns a Project Quality Manager (PQM). In the case of the CV Pilot, THEA submitted a standalone Quality Control Plan (Component of Phase 1 Submittal, Project Management Plan). Additionally, HNTB assigned a PQM as a resource to assist the data custodian with guidance on the overall quality management process. The HNTB PQM should be considered a support resource and not a formal manager role. Overall data management and quality falls to the data custodian.

Data quality controls focus on 5 key data quality attributes[1]:

1. Validity – is the data, message, data set or message set valid according to the standard or framework that governs the data.
2. Reliability – is the data being captured, collected, transmitted and/or received in a reliable way that meets the requirements of the applications that use the data.
3. Precision – is the spatial information accurate enough for the applications that use the data.
4. Integrity – is the data, message, data set or message set complete.
5. Timeliness – was the time stamp of the data within the required limits of the applications that use the data.

Discovering and identifying contributing factors to erroneous data is a function and component of the data quality audits to be performed by the CUTR project team. Given the vehicle and pedestrian safety applications planned for deployment, erroneous data is a significant concern. The CUTR team will work with the Data Custodian to manage erroneous data. Once discovered and assessed, there will be 3 ways to process erroneous data.

1. Delete – If it is determined that the data has no significant value or impact to the overall data set, application or performance measure, the data may be deleted.
2. Flag – If it cannot be determined that the data has significant value, cannot be parsed from its erroneous components, or impact to the overall data set, application or performance measure is undetermined, the data may be retained, but flagged to be omitted from certain analysis.
3. Correct - If it is determined that the erroneous data has significant value, can be parsed from its erroneous components, or impact of the erroneous component to the overall data set,

[1] Quality controls 2, Reliability and 3, Precision do not currently have established baseline standards. One of the stated purposes of the Pilot is to guide future standards. During the preliminary application and systems testing conducted toward the end of phase 2, the preliminary data collected and associated test results will be utilized to establish the baseline data quality standards for these two areas of data quality focus. Controls 1, 4 and 5 already have established baselines vis-a-vis the SAE and IEEE message and communications standards. The findings of the Pilot will be shared with these standards bodies for future refinement of those standards.

application or performance measure is negligible, then the data may be retained, and flagged that the erroneous data was corrected or further cleaned. Corrected data can be included or may be omitted from certain analysis.

During system integration, data quality audits will be conducted on data streams from all sites. Once the system is accepted and in operation, some data may only need to be audited periodically, while other data (safety-critical data) may need to be audited in real-time, constantly.

The Data Quality Control process detailed above is documented in the PMESP [2]. Accuracy of participant data is accomplished via the participant portal. This web-based portal provides a venue for participants to verify their data and update contact information that may change throughout the Pilot. The portal also provides push notification as well so that participants can be advised of changes to procedure or need to return to the Service Center for adjustments or firmware updates.

Another critical consideration of data quality is configuration management, (CM). CM is governed for the Pilot by the Configuration Management Plan (CMP). The CMP is a component of the Phase 1 PMP and includes procedures under which changes to design is submitted and approved by a change control board (CCB). Further an ongoing log of current and historical configurations are documented. This log includes software and firmware versions/dates, device serial numbers, maintenance/repair activities, etc. The Charter of the CCB has been added to the CMP along with meeting minutes and voting records from the first meeting of the CCB. A minimum of five CCB meetings are planned with the provision for more on an as needed basis.

## 2.5 Archiving and Preservation

All data associated with the project will be preserved in the THEA Master Servers. For the purposes of this CV Pilot, the lifecycle of all relevant data collected, as referenced in Section 2.1.3, and stored will be for the life of the project or longer as determined to be required by National Archives and Records Administration (NARA). Once federal funds are exhausted for maintenance and operations, THEA will permanently archive any data that is more than one-year- old to portable media drives and store it in THEA’s off-site secure storage facility.

All data associated with the project will be protected from loss in the THEA Master Server farm using a High-Availability Cluster. Data is replicated over the cluster using Virtual Machine technology. Back-ups will include both on-site and off-site media. On-site/On-line back-ups will be performed on the THEA Master Server VM-HA cluster. Additionally, on-site/off-line back-ups will be performed on portable media drives. Finally, off-site back-ups will be performed on portable media drives. Portable media drives will be rotated thru this process – when a new on-site/off- line back-up is complete, the last on-site/off-line back-up being securely stored on-site moves to THEA’s off-site secure storage facility. The last off-site back-up is then put back into the rotation. This process requires several sets of portable media drives.

Back-ups will be performed on the following schedule.

1. Daily back-ups. Daily back-ups will take place within the THEA Master Servers and are conducted automatically as a scheduled job. The Data Custodian is responsible for ensuring the nightly back-up process completed without error or corruption.
2. Weekly back-ups. Weekly back-ups will take place on-site/off-line and utilize portable media drives. The Data Custodian is responsible for ensuring the weekly back-up process completed without error or corruption. Weekly back-ups will be securely stored on-site for 5 weeks.

3. Monthly back-ups. Monthly back-ups will take place once per month. These back-ups will be stored off-line on portable media drives in THEA's off-site secure storage facility. The Data Custodian is responsible for ensuring the monthly back-up process completed without error or corruption. The Data Custodian is also responsible for ensuring the proper rotation of portable media drives throughout this process.
4. Ad-Hoc back-ups. Ad-hoc back-ups will take place as needed. Ad-hoc back-ups are typically required when performing system or software upgrades, security patches, passing decision gates or phases of the project, etc.
5. Yearly back-ups. Yearly back-ups will take place annually towards the end of THEA's fiscal year. These back-ups will be stored off-line on portable media drives in THEA's off- site secure storage facility. The Data Custodian is responsible for ensuring the yearly back-up process completed without error or corruption. The Data Custodian is also responsible for ensuring the proper rotation of portable media drives throughout this process.

In the event of a failure, the following restoration process and constraints apply:

1. Restoration from daily back-ups in the VM-HA cluster are conducted within the THEA Master Server farm. This type of architecture yields very low, if any, downtime depending on type of failure. There are no time constraints associated with restoring from a daily back-up.
2. Restoration from weekly back-ups is conducted using on-site portable media. This media is connected to the THEA Master Server farm. This type of restoration is usually required if the data is lost or corrupted. This type of restoration is also applicable if the VM-HA cluster experiences a fatal failure or event and the entire cluster requires a re- build. Restoration times are constrained to 4 hours, given the estimated size of the back- up.
3. Restoration from monthly back-ups is conducted using off-site portable media. This media is connected to the THEA Master Server farm. This type of restoration is applicable if the VM-HA cluster experiences a fatal failure or event and the entire cluster requires a re-build. Restoration times are constrained to one day, given the estimated size of the back-up.
4. Restoration from yearly back-ups is conducted using off-site portable media. This media is connected to the THEA Master Server farm. This type of restoration is applicable if the VM-HA cluster experiences a fatal failure or event and the entire cluster requires a re-build. Restoration times are constrained two days, given the estimated size of the back-up.

The project quality manager is required to perform scheduled check-ups on the back-up and restoration process. Once per year, the Data Custodian, THEA IT Staff, and Project Quality Manager shall perform loss and restoration drills to ensure the team is well-practiced in the event of a failure. Data Change Control and Management is a systematic approach to managing all changes made to the data a system generates. The purpose is to ensure no unnecessary changes are made, changes are approved through collaboration, all changes are documented and communicated, that services are not unnecessarily disrupted, and resources are used efficiently.

The concept of the data change control and management process is when new data or modifications to existing data are required, there is a process that governs how the change is documented, when and how the change is communicated to others; and, who collaborated for approval of the change. The purpose is to identify, alert, and assess the impact of the data change to consumers of the data. A data consumer could be a person or an application. The process will engage enterprise-users as well as cross-functional groups. The Data Custodian would assess the data change's impact to other applications and reporting.

The change control process is usually conducted as a sequence of steps proceeding from the submission of a change request. Typical change request logs provide the step-by-step chronological history of each change, including the addition of features to software applications, the installation of patches, and upgrades to network equipment. These requests are reviewed and approved by the CCB.

It is important that the right group of individuals is identified as the Data Change Committee. The Data Custodian is responsible for setting up the committee, however, may elect a committee chairperson to manage change requests.

An example of a data change would be if a new version of the SAE J2735 message set standard was to have a significant change in the structure or syntax of the message set.
This project will employ a six-step process for a data change request:

1. Documenting the change request: When the change request is made, it is categorized and recorded, along with informal assessments of the importance of that change and the level of difficulty of implementing it. Initial assessments will also identify what systems, subsystems, applications and users would be impacted by the proposed change.
2. Formal assessment: The justification for the change and risks and benefits of making/not making the change are evaluated by the CCB and documented in the change log. If the change request is accepted, the Data Custodian will assign staff, who are documented in the change log, to provide a solution/resolution to the change. If the change request is rejected, that fact is documented in the log and communicated to the appropriate parties.
3. Planning: The team responsible for the change creates a detailed work plan for its design and implementation, as well as a plan for rolling back the change should it be deemed unsuccessful.
4. Designing and testing: The team designs the solution/documents the resolution for the data change and tests it. If the change is deemed successful, the team requests approval from the Data Custodian and a date for implementation is agreed.
5. Implementation, review, and outreach: The team implements the change, stakeholders review the change and provide feedback. The Data Custodian communicates with all users documenting the change.
6. Final assessment: If the Data Custodian is satisfied that the change was implemented satisfactorily, the change request is closed, and the log is updated. If the client is not satisfied, the change is reassessed and steps in this process will be repeated until satisfactory completion.

The Data Change Control process detailed above will be added to the next iteration of the Configuration Management Plan section of the Quality Control Plan for the project. Please refer to the Quality Control Plan for further guidance.

# 3 Application Data Management Plan

CV Data are captured and transmitted at the roadside by the RSU. RSUs are located at signalized intersections, along the roadside of the Reversible Express Lane (REL), and at the mid-block pedestrian crossing at the Courthouse.

Detailed use case-location-hardware-software-message set matrices and other reference information for data flows is available in Sections 1 and 3 of the System Architecture Document for this project as well as the Application Deployment Plan.

For the purposes of this document, CV Data is defined as raw and metadata associated with message sets as part of the latest editions of SAE standards J2735 [9] and J2945 [10]. The proposed payload for BSM, TIM, SPaT and MAP were submitted by THEA for interoperability with the other pilot sites (Table 2).

**Table 2. CV Pilots SAE J2735 Message Sets**

| Message | Description |
|---|---|
| BSM | Basic Safety Message |
| MAP | Map Message |
| SPaT | Signal Phase and Timing Message |
| TIM | Traveler Information Message |
| SRM | Signal Request Message |
| SSM | Signal Status Message |
| PSM | Personal Safety Message |

SAE J2945_201603 specifies the system requirements for an on-board V2V safety communications system for light vehicles, including standards profiles, functional requirements, and performance requirements. The system is capable of transmitting and receiving SAE J2735-defined BSM over a DSRC wireless communications link as defined in the IEEE 1609 suite and IEEE 802.11 standards [10]. This standard is the first edition of on-board system requirements for V2V safety communications. It provides the information necessary to build interoperable systems that support select safety applications, which rely on the exchange of Basic Safety Messages.

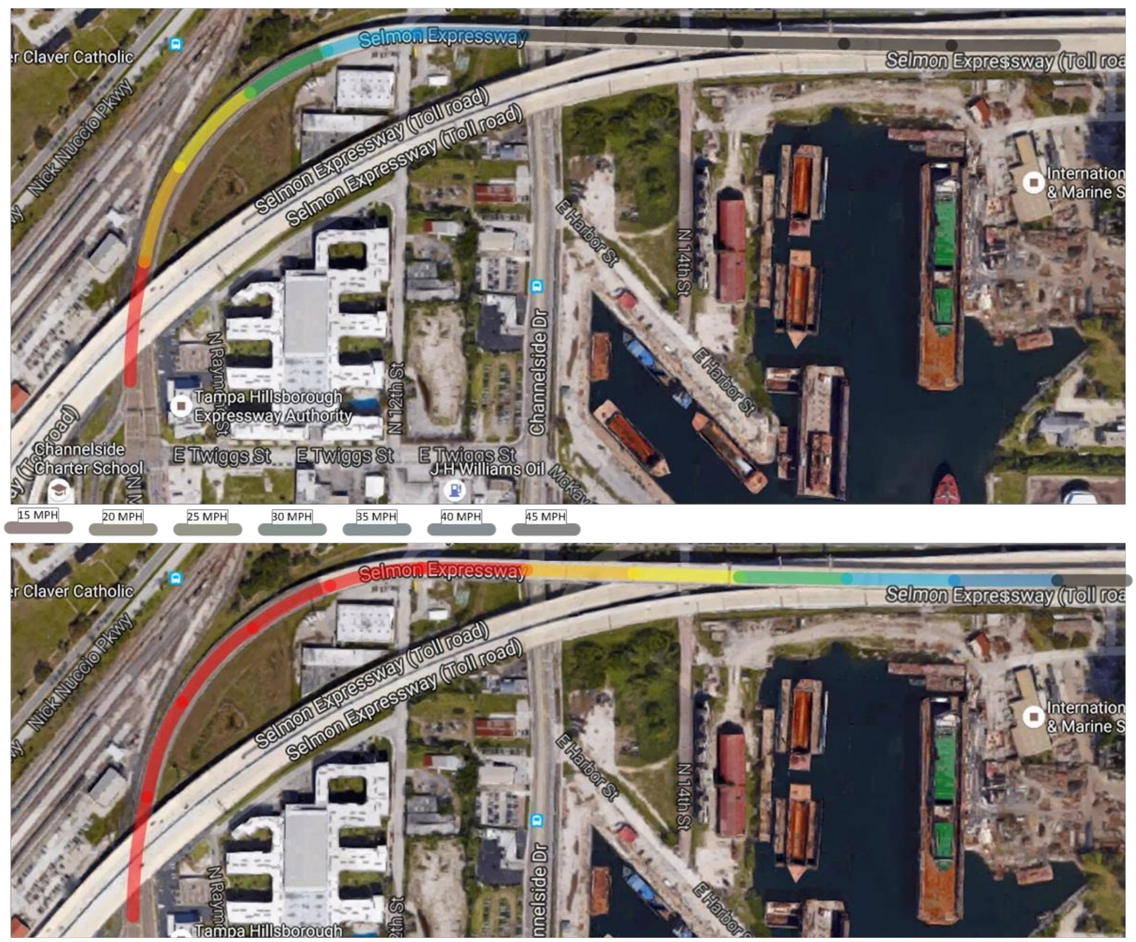
## 3.1 Application Overview

### 3.1.1 End of Ramp Deceleration Warning (ERDW)

#### *3.1.1.1 Overview*

As drivers near the end of the Reversible Express Lane, they enter a curve where the speed limit is reduced from 70 miles per hour (MPH) to 40 mph. During morning rush hour, vehicles back up from the exit into and around the curve creating long queues. This is especially true for the right turn only lane. Vehicles trying to turn right off the REL onto Twiggs Street will back up beyond the designated right turn lane onto the shoulder up onto and around the curve. These queues can grow to a mile or more in length.

The ERDW application provides speed advice to drivers, based on the longest queue length of any lane, who are approaching or are in the curve leading to the REL exit. The RSU application estimates the queue length of each lane. The ERDW application determines the longest queue using data-driven methods that rely on participants BSMs. When a connected vehicle sends a BSM, the streaming data are processed to update the speed information at the location of the vehicle. The value is stored at the index associated with that location. The vector/array is updated if the vehicle's speed is lower than the speed observed earlier at the same location within the same minute. A queue detection algorithm then parses the vector/array and extracts the queue information that is needed to determine the tail end of the vehicle queue. These queue length estimates are then fed back to the RSU to broadcast the Traveler Information Messages indicating the speed advice for each zone approaching the end of queue.




Source: Siemens (on Google Maps), 2017

**Figure 4. Exit Curve of the Reversible Express Lanes**

#### *3.1.1.2 Data Types and Sources*

For the ERDW application, there are two CV device types involved: the RSU and the OBU. Consequently, there are two ERDW applications: one running on the RSU and one running on the OBU.

The RSUs deployed at Twiggs and the REL exit and near the physical speed limit reduction sign for the REL exit curve will be generating the data. The queue length data will be provided once a minute. The recommended speed advice will be provided once a minute to coincide with any change to the queue length.

#### *3.1.1.3 Data Collection and Transmission*

The RSU application will store the following information:

- Timestamp
- Queue length
- Recommended Speed Advice
- OBU data received for upload to the Master Server

The RSU Speed Advice RSU will store the following information:

- Timestamp
- Recommended Speed Advice
- OBU data received for upload to the Master Server

The OBU ERDW will store the following information:

- RSU ERDW Recommended Speed Advice
- OBU ERDW Specific Vehicle Speed Advice

#### *3.1.1.4 Adopted Resolution to Queue Length Estimation*

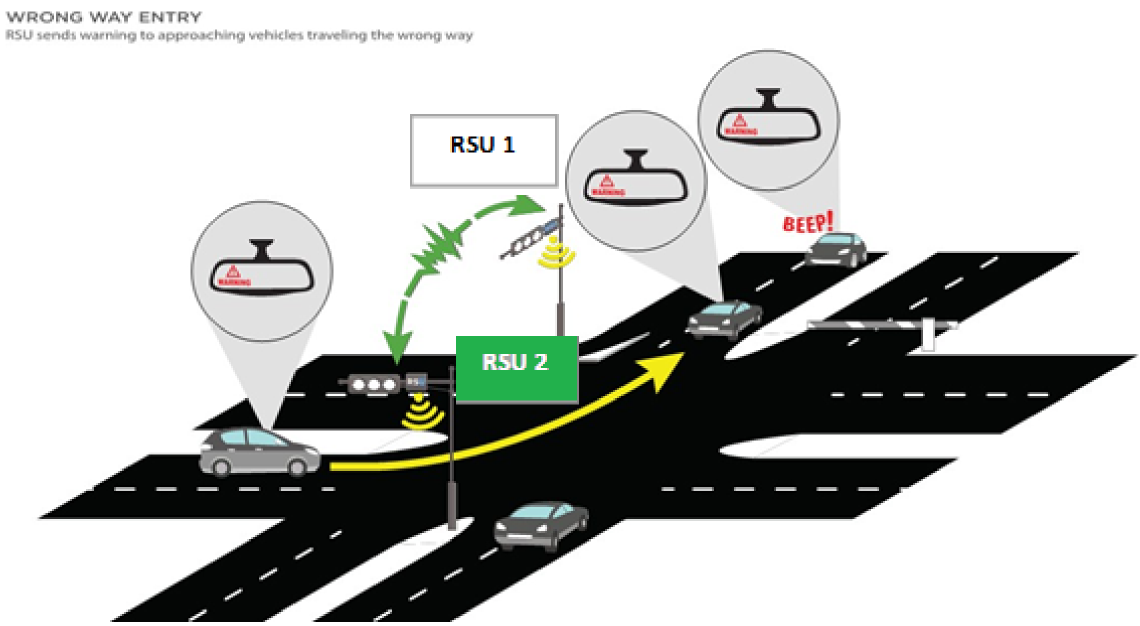
The I-SIG implementation for this use case was designed to feed information to the MMITSS application to estimate queue length (along other traffic delay measures) so that advisory speeds would be sent to participants via the ERDW application as they approached the REL exit. Subsequent performance tests revealed that the MMITSS application was not correctly estimating queue length. During Phase 3 of the deployment, Siemens Mobility provided the following queue length estimation solution, which was implemented in February 2020. Data collection on participants vehicles started in February 2020.

### 3.1.2 Wrong Way Entry (WWE)

#### *3.1.2.1 Overview*

When an OBU equipped vehicle is approaching the REL exit/entrance, it receives a MAP message from the Twiggs at REL RSU. This MAP message provides information about each REL exit/entrance lane including whether the lane is revocable or not. WWE OBU application uses the MAP data in conjunction with continuous snapshots for the vehicle’s location (latitude/longitude), elevation, speed, and travel direction to determine if the vehicle is approaching the REL exit in an attempt to enter the REL going the wrong way. If the WWE OBU application determines that the vehicle is advancing to enter the REL going to wrong way, the application issues a warning to the driver. If the application determines the vehicle has corrected course, there are no further warnings issued. If the application determines the vehicle has continued up the REL the wrong way based on the MAP message, a second warning is issued to the driver. The WWE RSU application receives detector input via the local traffic controller from a wrong-way driver detection system and communicates it to the RSU near the physical speed limit sign on the REL and sends a message to the Master Server. The WWE RSU application on both RSUs broadcasts a TIM message to the approaching equipped vehicles. The WWE OBU application receives the TIM message and warns the driver that a wrong way vehicle is headed toward them. Also, if the vehicle enters a no travel lane, the OBU will issue a warning to the driver. Figure 5 provides a functional view of the WWE application.

**Figure 5. WWE Functional View**

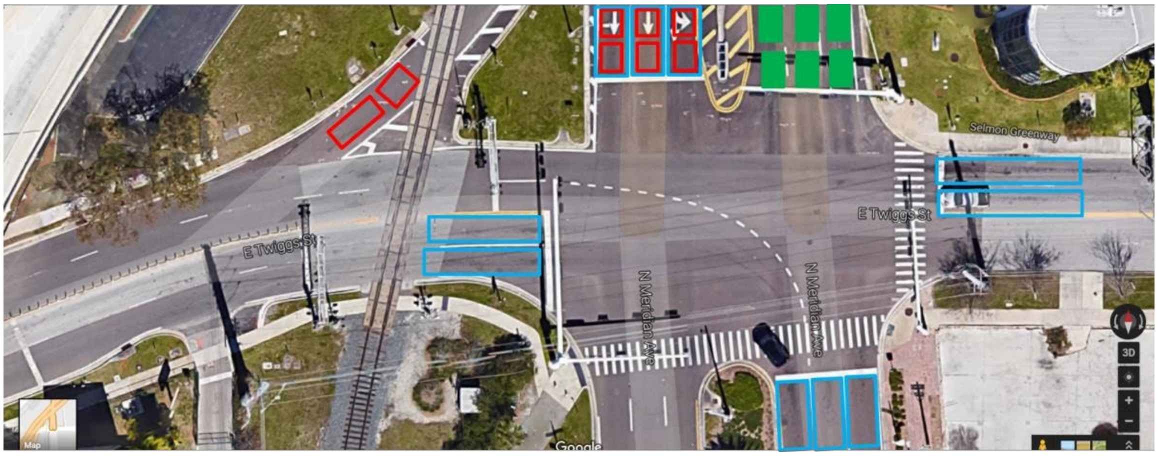



*Source: System Architecture Document, Publication FHWA-JPO-17-459*

**Figure 6. WWE Functional View**

Figure 6 depicts the overlap between Wrong Way application and the Traffic Progression application.

- RSU #1 operates without a signal controller but transmits MAP and SPaT.
- Red indicates the ingress MAPs of the REL with a fixed (inbound) direction
- Green indicates the reversible MAPs of the REL that are either egress or revoked, based on time of day indicated by the SPAT message.
- RSU #2 operates with a signal controller
- Blue indicates MAPs associated with the signal control of Twiggs and Meridian.
- MAP and SPaT are transmitted for improved traffic progression.




*Source: System Architecture Document, Publication FHWA-JPO-17-459*

**Figure 7. WWE Physical Overview**

### *3.1.2.2 Data Type and Sources*

For the WWE application, there are two CV device types involved: the RSU and the OBU. Consequently, there are two WWE applications: one running on the RSU and one running on the OBU.

The WWE OBU application has two functions: one function for the wrong way driver and one function for all other drivers. The WWE OBU for the wrong way driver receives the following data:

- Timestamp
- MAP message for all lanes
- SPAT message for revoked lanes coinciding to the barrier status

The WWE OBU application generates its own

- Local vehicle location, elevation, speed, and travel direction

The WWE OBU for all other equipped vehicles listens for a TIMs broadcast from the WWE RSU application, which contains:

- Wrong Way Driver approaching
- Last known wrong way driver location, heading, speed
- Timestamp

The RSUs for the WWE application are deployed at Twiggs and the REL exit and near the physical speed limit reduction sign for the REL exit curve will be generating the data. The WWE RSU application has two functions: one function to detect and process a wrong way driver and one function processing wrong way warnings. The WWE RSU function to detect and process a wrong way driver receives the following data:

- Timestamp
- WWE OBU message containing wrong way vehicle data

The function broadcasts the following data:

- Timestamp
- WWE RSU message containing wrong way vehicle data

The WWE RSU function to process wrong way warnings receive the following data:

- Timestamp
- WWE RSU message containing wrong way vehicle data

The WWE RSU function to process wrong way warnings broadcasts the following data:

- Timestamp
- TIM containing wrong way vehicle data (location, elevation, speed, and travel direction)
- WWE RSU Master Server message containing wrong way vehicle data (location, elevation, speed, and travel direction)

The Master Server receives the WWE RSU Master Server message and logs the message.
The Twiggs at REL Exit RSU broadcasts the MAP message, broadcasts the TIM, and transmits the Master Server Message. The RSU near the physical speed limit sign broadcasts the TIM.

The TIM will be broadcast 10 times a second while the wrong way vehicle is within the MAP.

#### *3.1.2.3 Data Collection and Transmission*

The WWE RSUs application will store the following information:

- Timestamp
- Wrong Way Data and Warnings
- MAP
- BSMs received
- OBU data received for upload to the Master Server

The WWE OBU will store the following information:

- Timestamp
- WWE OBU offending wrong way vehicle data
- Wrong Way Warnings
- WWE RSU TIM

The Master Server will store the following data:

- Timestamp
- WWE RSU Master Server message
- WWE RSU data
- WWE OBU offending wrong way vehicle data
- WWE RSU TIM
- Wrong Way Warnings
- MAP

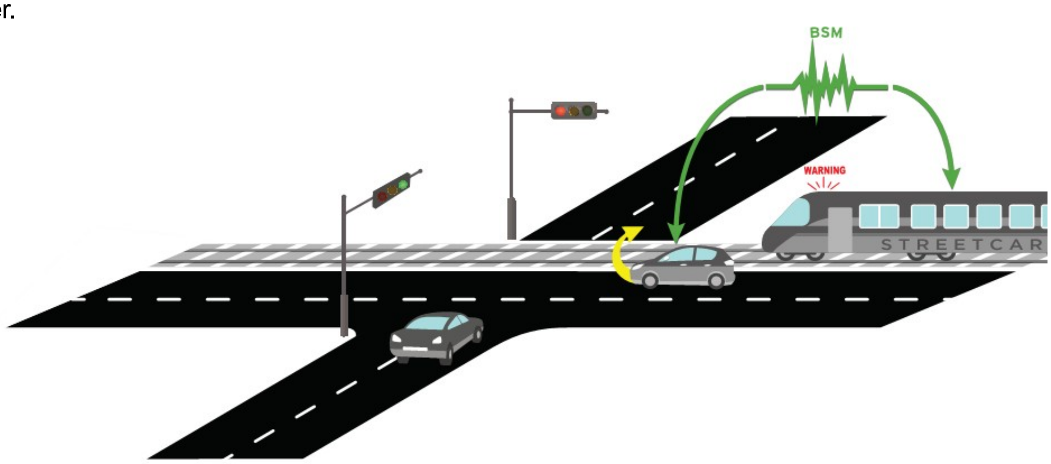
The WWE OBU and WWE RSU will communicate using DSRC. The WWE RSU and the Master Server will communicate via a wireless (LTE) connection.

This data collection was planned to be prototyped in November 2017 for six months. In June 2018, the pilot will begin collecting test data. In March 2019, the pilot started collected data for 18 months from participant vehicles.

### 3.1.3 Vehicle Turning Right in Front of Transit Vehicle (VTRFTV)

#### *3.1.3.1 Overview*

As an OBU equipped vehicle approaches an intersection wanting to make a right turn across the streetcar tracks, the OBU is broadcasting BSMs via DSRC. As an OBU-equipped streetcar approaches the intersection, it receives the BSMs. Using these BSMs, the streetcar OBU VTRFTV application (Figure 7) determines if the vehicle's right-hand signal is activated. If the OBU VTRFTV application determines the streetcar and vehicle are on a collision course, the application sends a warning to the streetcar operator. A similar trajectory calculation is made on the vehicle OBU and if a potential collision course is determined, the vehicle OBU application sends a warning to the vehicle driver.




*Source: System Architecture Document, Publication FHWA-JPO-17-459*

**Figure 8. Functional View of VTRFTV**

#### *3.1.3.2 Data Types and Sources*

The OBU VTRFTV streetcar application will store the following information:

- Timestamp
- Received BSMs
- Streetcar's location and speed
- Vehicles location and speed
- Warning to operator
- Warning to driver

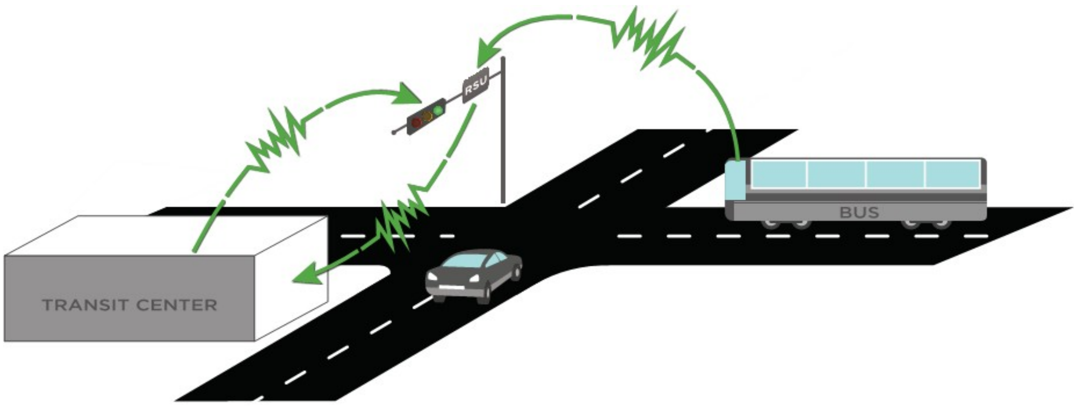
This data collection was planned to be prototyped in November 2017 for six months. In June 2018, the pilot will begin collecting test data. In March 2019, the pilot started collected data for 18 months from participant vehicles.

### 3.1.4 Transit Signal Priority (TSP)

#### *3.1.4.1 Overview*

According to the THEA CV Pilot System Design Document (SDD), the Transit Signal Priority app provides signal priority to transit at intersections along arterial corridors only if the bus is behind schedule. TSP is part of MMITSS. If the bus is behind schedule, priority will be granted for the bus. The OBU sends a Signal Request Message to the RSU. The RSU forwards that to the Transit Server at the Traffic Management Center. The Transit Server determines if the bus is behind schedule. If the bus is behind schedule, the SRM is returned from the Transit Server to the RSU. If the signal is green in the bus's travel direction, the RSU selects the controller phase via National Transportation

Communications for Intelligent Transportation System Protocol (NTCIP) objects to extend the green, allowing the bus to proceed through the intersection. If the signal is yellow or red in the bus's travel direction, the RSU requests the shortest cycle via NTCIP objects to provide a green to the bus as quickly as possible. At the same time, the RSU sends the Signal Status Message to the approaching bus to inform the driver of priority received. Figure 8 shows the functional overview and flows of the application.




*Source: System Architecture Document, Publication FHWA-JPO-17-459*

**Figure 9. Functional View of TSP**

#### *3.1.4.2 Data Types and Sources*

For the TSP application, there are two CV device types involved: the RSU and the OBU. The transit server also contains an application. Consequently, there are three TSP applications: one running on the RSU, one running on the OBU, and one running on the transit server.
The OBU TSP application broadcasts the following data:

- SRM

The OBU TSP application listens for the following data:

- SSM

The RSU TSP application broadcasts the following data via DSRC:

- SSM

The RSU TSP application sends the following data:

- Transit Server Priority Request to the Transit Server
- Priority request to the signal controller

The RSU TSP application listens for:

- SRM
- Transit Server Priority Response (priority granted/priority denied)
- Confirmation from the signal controller

The TSP application listens for the following data:

- Transit Server Priority Request

- The Transit Server sends the following data:
- Priority granted/Priority denied)

The RSUs are deployed along Marion Street, Jackson Street, and Kennedy Street. The SRM is broadcast ten times a second. The SSM is a one-time transmission.

#### *3.1.4.3 Data Collection and Transmission*

The RSU TSP application will store the following information:

- Timestamp
- SRMs
- SSMs

The OBU TSP will store the following information:

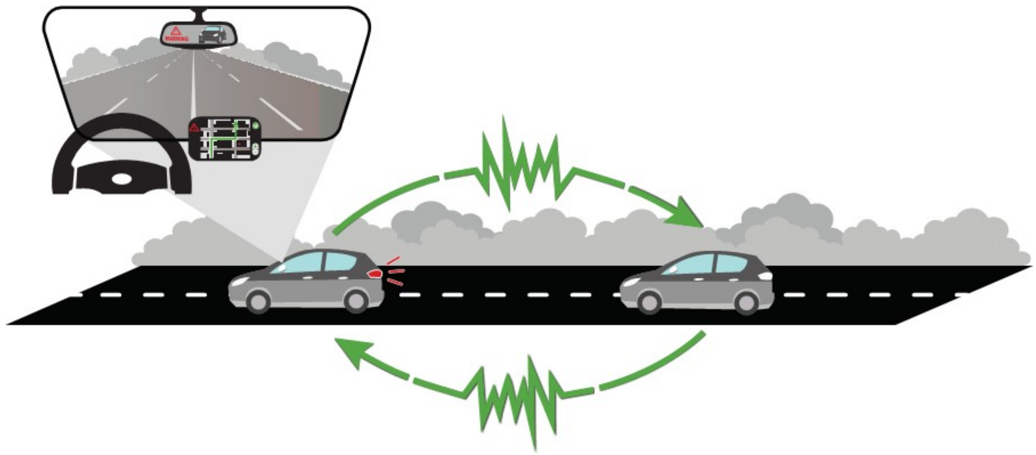
- SRMs
- SSMs

This data collection was planned to be prototyped in November 2017 for six months. Test carried out through 2018. Due to functional implementation issues associated with MMITSS, the TSP apps was not deployed to the bus fleet. Tests are currently ongoing to implement a proposed resolution.

### 3.1.5 Forward Collision Warning (FCW)

#### *3.1.5.1 Overview*

An OBU FCW-equipped vehicle approaches another OBU FCW-equipped vehicle in the same lane. The two vehicles are exchanging BSMs containing data such as location, elevation, travel direction and speed. The rear vehicle FCW application (Figure 9), using the lead vehicle's BSM data, determines if the rear vehicle is about to crash into the lead vehicle. If the FCW application determines the rear vehicle is about to crash into the lead vehicle, the FCW application sends a warning to the driver.



Source: System Architecture Document, Publication FHWA-JPO-17-459

**Figure 10. Functional View of FCW**

#### *3.1.5.2 Data Types and Sources*

For the FCW application, the OBU is the only CV device used by the application on equipped vehicles.

The OBU FCW vehicle application listens for the following data:

- Other equipped vehicles' BSMs

The OBU FCW vehicle application sends a warning to the driver: The BSMs are provided ten times a second.

#### *3.1.5.3 Data Collection and Transmission*

The OBU FCW vehicle application stores the following information:

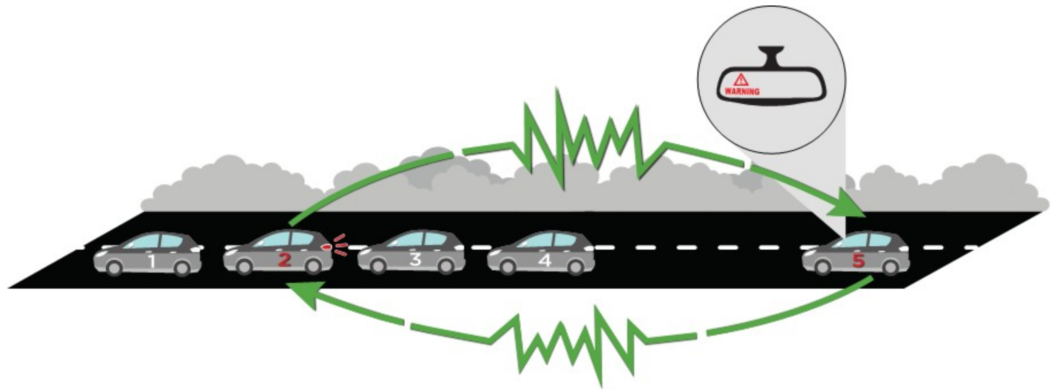
- Timestamp
- Received BSMs
- Vehicle's location and speed
- Alert sent to the driver

This data collection was planned to be prototyped in November 2017 for six months. In June 2018, the pilot will begin collecting test data. In March 2019, the pilot started collected data for 18 months from participant vehicles.

### 3.1.6 Emergency Electronic Brake Light Warning (EEBL)

#### *3.1.6.1 Overview*

An OBU EEBL-equipped vehicle approaches a line of vehicles in the same lane. Another OBU EEBL-equipped vehicle in the line broadcasts its BSMs via DSRC. This vehicle brakes suddenly. The approaching vehicle receives the BSMs of the vehicle in line. The OBU EEBL application (Figure 10), using these BSMs, determines whether the vehicle in line brakes suddenly. When the application determines a sudden braking, a warning is sent to the driver.




Source: System Architecture Document, Publication FHWA-JPO-17-459

**Figure 11. Functional View of EEBL**

#### 3.1.6.2 Data Types and Sources

For the EEBL application, the OBU is the only CV device used by the application on equipped vehicles.

The OBU EEBL vehicle application listens for the following data:

- Other equipped vehicles' BSMs

The OBU EEBL vehicle application sends a warning to the driver: The BSMs are provided ten times a second.

#### 3.1.6.3 Data Collection and Transmission

The OBU EEBL vehicle application stores the following information:

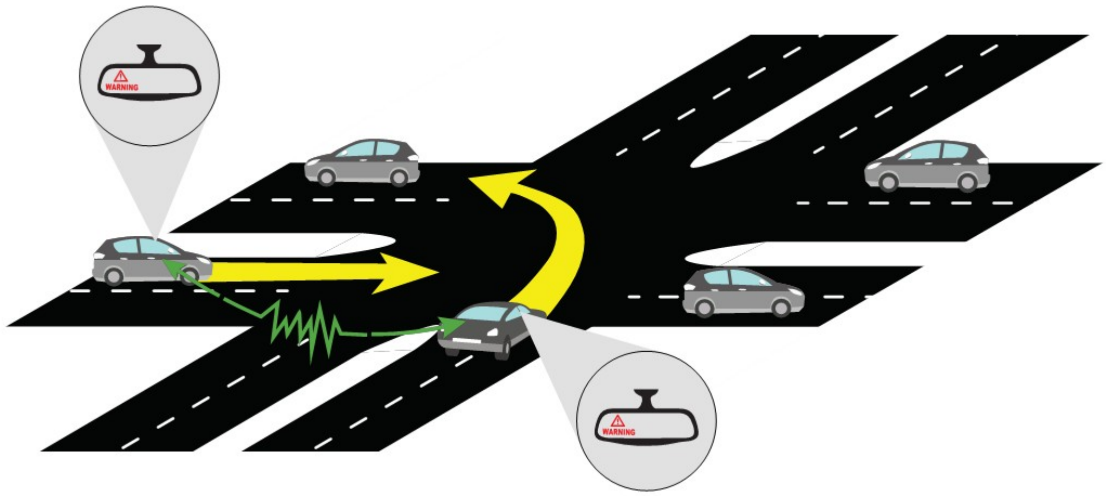
- Timestamp
- Received BSMs
- Vehicle's location and speed
- Alert sent to the driver

This data collection was planned to be prototyped in November 2017 for six months. In June 2018, the pilot will begin collecting test data. In March 2019, the pilot started collected data for 18 months from participant vehicles.

### 3.1.7 Intersection Movement Assist (IMA)

#### 3.1.7.1 Overview

Two vehicles perpendicular to each other approaching or stopped at an intersection are each broadcasting BSMs received by the other. The OBU IMA application (Figure 11) receives the other vehicle's BSMs and calculates the trajectories. If the application determines the vehicles are on a collision course, a warning is sent to the drivers.




*Source: System Architecture Document, Publication FHWA-JPO-17-459*

**Figure 12. Functional View of IMA**

#### 3.1.7.2 Data Types and Sources

For the IMA application, the OBU is the only CV device used by the application on equipped vehicles.

The OBU IMA vehicle application listens for the following data:

- Other equipped vehicles' BSMs

The OBU IMA vehicle application sends a warning to the driver: The BSMs are provided ten times a second.

#### 3.1.7.3 Data Collection and Transmission

The OBU IMA vehicle application stores the following information:

- Timestamp
- Received BSMs
- Vehicle's location and speed
- Alert sent to the driver

This data collection was planned to be prototyped in November 2017 for six months. In June 2018, the pilot will begin collecting test data. In March 2019, the pilot started collected data for 18 months from participant vehicles.

### 3.1.8 Intelligent Signal System (I-SIG)

#### 3.1.8.1 Overview

The RSU MMITSS-I-SIG application uses OBU equipped vehicles' BSMs as well as traditional detector devices to detect vehicles and estimate queue lengths on intersection approaches. I-SIG uses this information in order to adjust signal timing for a series of intersections along Meridian Avenue and Florida Avenue.

#### *3.1.8.2 Data Sources and Types*

For the I-SIG application, the RSU is the only CV device used by the application on equipped intersections. The RSU I-SIG infrastructure application listens for the following data from vehicles:

- Equipped vehicles' BSMs

The RSU I-SIG infrastructure application sends the following data to the signal controller:

- NTCIP 1202 v2 standard SET phase control
- NTCIP 1202 v2 standard GET phase status

The RSU I-SIG infrastructure application receives the following data from the signal controller:

- NTCIP 1202 v2 standard phase status response

The RSU I-SIG infrastructure application sends the following data to the ERDW application:

- Queue length from the REL

The BSMs are provided ten times a second.

#### *3.1.8.3 Data Collection and Transmission*

The RSU I-SIG infrastructure application stores the following information:

- Queue length
- BSM counts

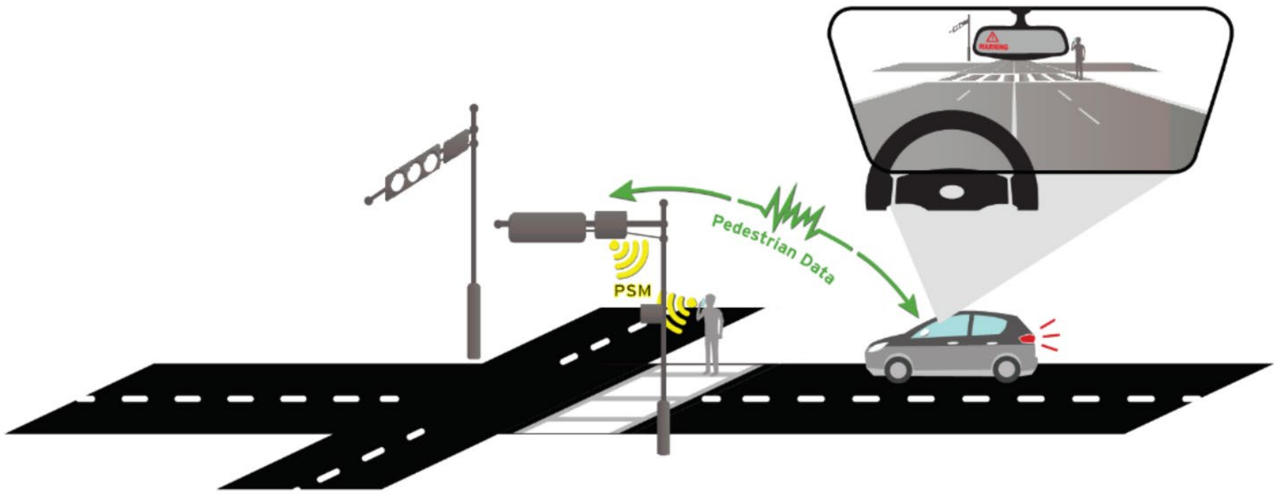
This data collection was planned to be prototyped in November 2017 for six months. Test carried out through 2018. Due to functional implementation issues associated with MMITSS, the I-SIG app was not deployed to collect participant data. Tests are currently ongoing to implement a proposed resolution

### 3.1.9 Pedestrian Collision Warning (PCW)

#### *3.1.9.1 Overview*

As a vehicle is approaching the courthouse mid-block crossing, the vehicle receives PSMs from the RSU. The OBU Pedestrian Collision Warning (PCW) app using the PSMs and the vehicle data, calculates if there is a potential crash course with a pedestrian. If a crash course is determined, the PCW app alerts the driver.

The PCW application was designed to work at the midblock crosswalk on East Twiggs Street at the Hillsborough County Courthouse to improve pedestrian safety. Initially, two Light Detection and Ranging (LiDAR) sensors installed at the crosswalk located the pedestrians in the area and translated the information to PSMs and broadcast them over DSRC to the OBUs. The RSU-PCW application broadcast the PSMs over DSRC to the OBUs. OBU equipped vehicles, using the PCW application, warned the drivers that are on a collision course with pedestrian on the crosswalk. Figure 12 shows the application's functional overview, which originally included a LiDAR that was replaced with thermal imaging sensors because upon deployment of the LiDAR system, it was concluded that the operational reliability of the LiDAR sensors was not adequate to support the UC3 goals within the timeframe of the project.




*Source: System Architecture Document, Publication FHWA-JPO-17-459*

**Figure 13. Functional View of PCW**

### *3.1.9.2 Data Types and Sources*

For the PCW application, the OBU is the only CV device used by the application on equipped vehicles at the mid-block crossing at the courthouse. The OBU PCW application listens for the following data from vehicles:

- PSMs from the courthouse RSU

The OBU PCW application sends the following data to the RSU:

- Detected crash courses with a pedestrian

The OBU PCW application receives the following data from the courthouse RSU:

- PSMs

The OBU PCW application sends the following data to the driver.

- Alert of a crash course with a pedestrian.

### *3.1.9.3 Data Collection and Transmission*

The OBU PCW application stores the following information:

- PSMs
- Crash course alerts

During the Phase 2, data analysis revealed the LiDAR system was not producing accurate data to perform the safety assessment. Eventually, it was concluded that the operational reliability of the LiDAR sensors was not adequate to support the research goals within the limited time remaining in project Phase 3 and thus the LiDAR sensors were replaced with video and thermal imaging sensors. The new system was installed in May 2020, and subsequent testing using test vehicles was conducted in June, July, and August 2020. On August 5, 2020, the system started full operation and deployment to participants. The change to the thermal camera system did not affect the PSM data generation process and data structure.

# 4 Data Collection Analysis

## 4.1 BSM Data Generation Simulation

In the early stages of Phase II, an analysis of the volume of data likely to be generated over the course of the study was performed. The goal was to assess the potential for roadblocks to the intention of collecting and retaining as much CV data as possible. This effort resulted in a baseline scenario that served as the starting point for the refinements presented in this section.
Table 3 reports the baseline assumptions and the initial data generation estimate.

**Table 3. Baseline BSM Data Estimation Scenario**

| *Scenario* | *Unit* | Baseline Scenario |
|---|---|---|
| BSM size (80 + some extra)(payload 80 bytes, incl. cert 240 bytes, incl. headers 320 bytes) | bytes | 100 |
| BSM/sec/veh | | 10 |
| Host Vehicle (HV) | | 1 |
| Remote Vehicle (RV) | | 10 |
| BSM Tx/Vec/sec | bytes | 1,000 |
| BSM Rx/Veh/sec | bytes | 10,000 |
| BSM data rate per HV (Tx+Rx) | KB / sec | 11 |
| Number of equipped cars | | 1,600 |
| Time spent per vehicle in study area | minutes | 30 |
| **BSM data Tx from cars / day** | **GB** | **3.0** |
| **Duplicate BSM data Rx from cars / day** | **GB** | **29.0** |
| **Data collected from cars/day** | **GB** | **32.0** |
| Number of equipped buses and streetcars | | 20 |
| Time spent for buses and streetcars in study area | hours | 12 |
| **Data collected per day from buses and streetcars** | **GB** | **10.0** |
| RSUs within range of an OBU | | 10 |
| BSM data Rx by RSUs per HV | KB / sec | 10 |
| **BSM data Rx by all RSUs per day from cars** | **GB** | **29.0** |
| **BSM data Rx by all RSUs per day from buses and streetcars** | **GB** | **9.0** |
| **Duplicate BSM data Rx from cars/day-bus and streetcars** | **GB** | **9.0** |
| **BSM data Tx from cars / day--bus and streetcars** | **GB** | **1.0** |
| Number of RSUs | | 40 |
| 1 Map (1500) Sent every 2 sec / RSU / sec | bytes | 750 |
| 1 Spat (160) Sent every 1 sec / RSU / sec | bytes | 160 |
| **Spat data Tx by all RSUs per day (don't count MAP)** | **GB** | **1.0** |
| **Dup Spat/Map data Rx by all cars per day** | **GB** | **27.0** |
| **Duplicate (BSM, SPAT, Map) data Rx stored at master server per day** | **GB** | **103.0** |
| **Non duplicate data** | | **5.0** |
| **Data stored per day at master server** | **GB** | **108.0** |
| **Data collected over project duration (1.5 years = 548 days)** | **GB** | **59,184** |
| **Average Data transferred via LTE per RSU per day** | **GB** | **2.7** |
| **Average Data transferred via LTE per RSU per month** | **GB** | **81.0** |

### 4.1.1 Refinement to the Baseline Scenario

The baseline scenario was further refined to account for uncertainty about the magnitude of key input parameters, such as:

- Overall data collection timeframe
- Expected time study participants will spend within the study area
- Expected number of OBUs in range of a RSU
- Expected number of CV equipped vehicles in the study area at any time
- The expected size of BSMs, SPAT, and MAP messages

To this extent, a new spreadsheet was developed that replicates the original baseline scenario and introduces a new one that that uses empirical information on the study area and treats the most sensitive input parameters as random variables generated by known parametric processes. The spreadsheet runs a MonteCarlo simulation to generate a distribution of the total data stored at the master server on a daily basis. The spreadsheet also estimates the probability that the data generated will be less than a given target maximum data load.

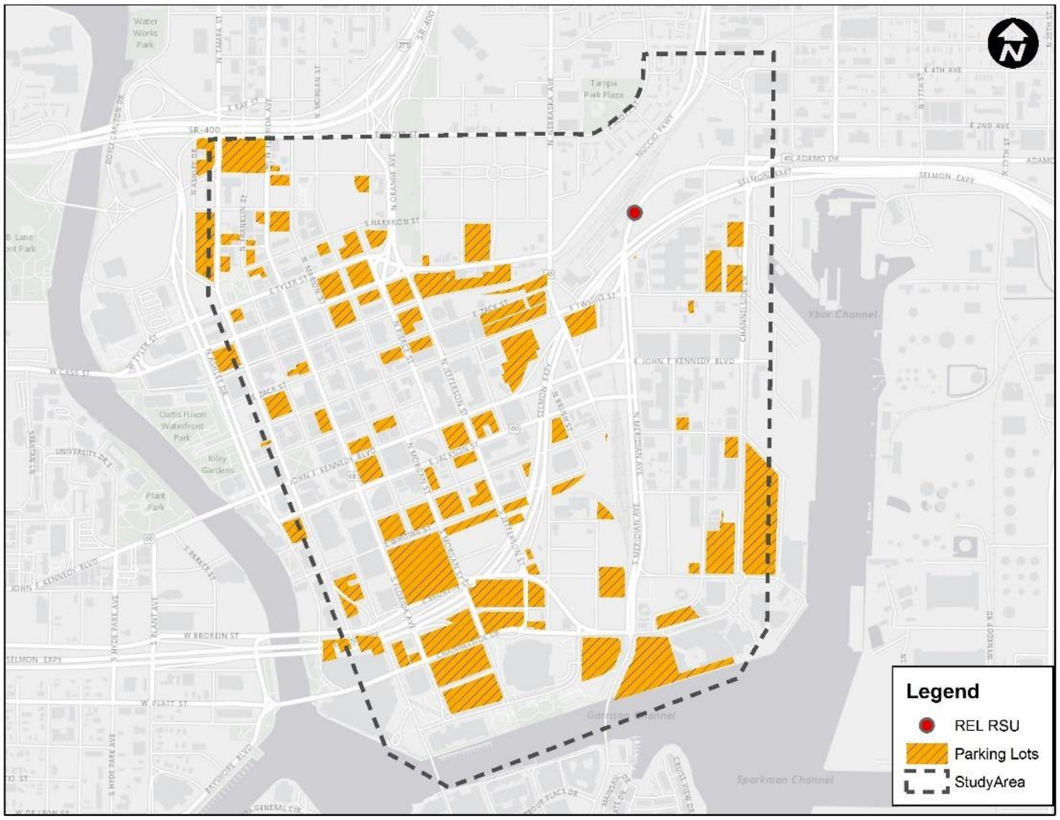
#### *4.1.1.1 Study Timeframe Parameters*

The initial baseline scenario data collection timeframe was revised as more information about the pool of potential participants became available. Given that the vast majority of study participants will be commuters and particularly REL users, most of the data collection is likely to occur over the course of a working week. Therefore, the project duration assumption of Table 3 was modified by substituting working days (21/month) over the course of 1.5 years of data collection. This reduced the original estimate from 59.2 terabytes (TB) to 40.8 TB.

#### *4.1.1.2 Study-specific parameters*

One the most sensitive parameters is the assumed time vehicles will spend, on average, in the study area. For example, reducing travel time from the original estimate of 30 minutes to 15 minutes reduces the amount of data stored to 24.9 TB, using only working days over the study period.

The revised scenario assumes that the time spent per vehicle in the study area is equivalent to the average travel time from/to the first RSU located on the REL to/from all parking lots located in the study area (Figure 13). This is equivalent to hypothesizing that the vast majority of participants will enter the study area via the REL in the a.m. peak only to get back to it in the afternoon commute. Estimates of travel times are available using Google's Directions automated program interface (API). Google Directions API allows batch-processing origin/destination travel distances and travel times using the RSU and parcel centroids coordinates.8 A data collection produced 15-minute interval estimates for the 7:00-10:00 a.m. peak and 4:00-7:00 p.m. period, weekday travel conditions. The assumed travel pattern is from the RSU to all parking lots in the a.m. and travel from all parking lots to the RSU in the p.m. peak period.




*Source: CUTR*

**Figure 14. Study Area Parking Lots**

Table 4 provides basic descriptive statistics of the estimated travel times. A test on the equality of means of a.m. and p.m. peak periods did not find a statistically significant difference in means. These estimates were used to treat the assumed time vehicles spent in the study area as a normally distributed random variable truncated at the minimum and maximum travel times.[2]

**Table 4. Travel Times to/from RSUs**

| *Period* | *mean* | *min* | *max* | *sd* |
|---|---|---|---|---|
| Peak a.m. | 4.26 | 2.00 | 10.00 | 1.68 |
| Peak p.m. | 4.24 | 2.00 | 9.00 | 1.69 |
| Overall | 4.25 | 2.0 | 10.0 | 1.69 |

#### *4.1.1.3 Number of vehicles*

The baseline scenario assumes 1,600 vehicles traveling in the study area every day. The refined scenario assumes that there is no certainty that every day over the course of the study there will be 1,600 vehicles in the study area for the estimated amount of time. Thus, the number of vehicles is treated as uniformly distributed with a minimum of 900 to a maximum of 1,600 vehicles. There is no

[2] Note that the maximum travel time reported in Table 3 refers to the largest observation in the sample. The 95th percentile maximum is 7 minutes. As a conservative approach, the simulation uses the 10-minute maximum.

specific reason or historical information to justify these ranges, other than testing how sensitive the results are to this parameter. More information to hone these parameter estimates will become available as recruitment takes place.

#### *4.1.1.4 Number of buses/streetcars*

The baseline scenario assumes a total of 20 equipped buses and streetcars servicing the area 12 hours a day. The update considers that this number could vary between 18 and 20 vehicles following a uniform distribution. According to TECO, the streetcar runs from noon to 10:00 pm Monday through Thursday and from 11:00 a.m. to 1:30 a.m. on Friday.[3] At this time, the number of bus/streetcar service hours/day has not been estimated. Note that treating this input as a random variable allows testing how sensitive the results are to this parameter.

#### *4.1.1.5 Number of RSU within range of an OBU*

The baseline scenario uses 10 RSUs in range of an OBU. It is now assumed that this number will vary from 9 to 10 following a uniform distribution. There is no specific reason to do so other than testing how sensitive the results are to this parameter.[4]

### 4.1.2 Simulation Results

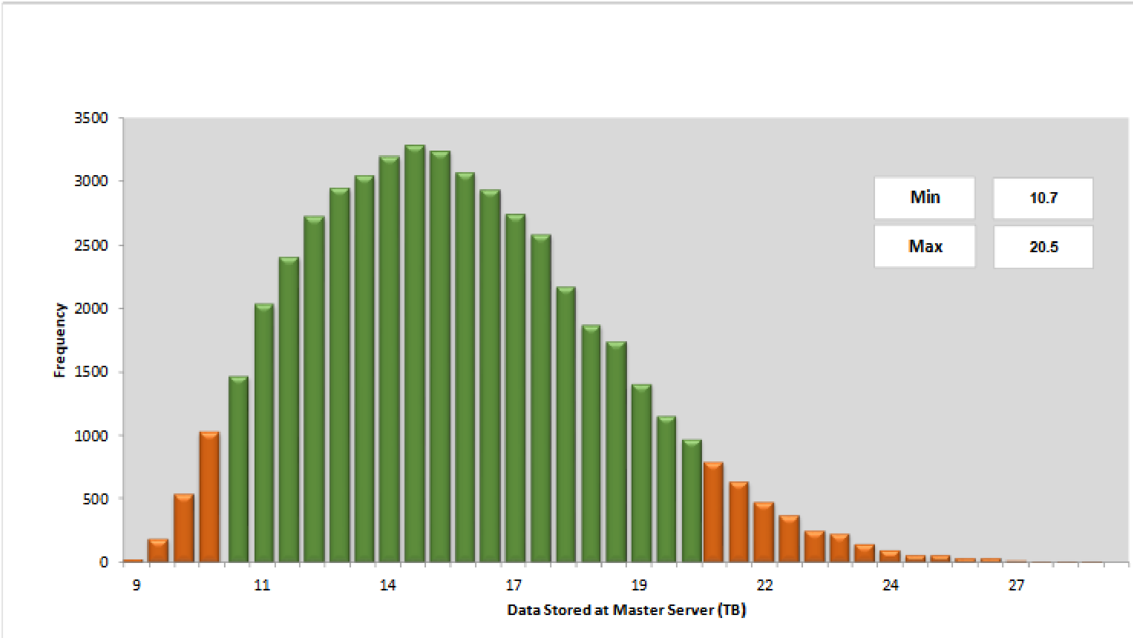
Table 5 reports the results of the simulation with 50,000 iterations. Being the output of a simulation, the best interpretation of these results is in terms of ranges over the distribution of the outcome interest, the total amount of data generated over the course of the entire study, as shown in Figure 14. Given a maximum threshold of data that can be stored on a server, the simulation can estimate the probability of being at or below that threshold. Figure 14 indicates the estimated probability that the total amount of data generated will below 20 TB is 93.4 percent.

**Table 5. Simulation Results**

| Data Collected/Transmitted | Estimated Output | Unit |
|---|---|---|
| Duplicate (BSM, SPAT, Map) data Rx stored at master server/day | 38.9 | GB |
| Non duplicate data/day | 2.2 | GB |
| Data stored at master server/day | 41.1 | GB |
| Average Data transferred via LTE per RSU per day | 1.0 | GB |
| Average Data transferred via LTE per RSU per month | 21.3 | GB |
| Data collected over project duration (1.5 years = 378 workdays) | 15.6 | TB |

[3] http://www.tecolinestreetcar.org/#/home#cs5

[4] While the RSU map (not reported here) shows almost a complete as-the-crow-flies overlap of the study area, Tampa CBD's building and other infrastructure will likely negatively affect this overlapping and thus data transmission/reception. Therefore, one can assume there is no perfect identity between overlapping and data redundancy.




| | |
|---|---|
| Max target data (TB) | 20.0 |
| Probability less than target | 93.4% |

*Source: CUTR*

**Figure 15. Estimated Data Storage at Master Server**

## 4.2 Data Collection Approach

### 4.2.1 Data Collection Within the Study Area

The original analysis presents a baseline scenario under which OBUs log all sent and received BSMs and RSUs log all received BSMs at a 10 Hz rate. This scenario produces about 59.2TB of data over the course of the study timeframe. The baseline scenario's assumptions were revised and augmented with traffic data to reflect historical travel conditions in the study area. The revised scenario treats key input parameters as random variables conducive to a MonteCarlo-based simulation exercise. The simulation estimates that over the course of the study the total data generated will be on average about 15.1 TB or 25.5 percent of the initial baseline scenario estimate of 59.2 TB. On a daily basis, the amount of data transmitted by each RSU via LTE will be about 1.1 GB. Based on these results, the Tampa team will take an approach where all OBUs log all sent and received BSMs and RSUs log all received BSMs at a 10Hz rate.

### 4.2.2 Data Collection Outside the Study Area

The scenario of Figure 14 can be further refined by populating input parameters with more precise estimates as data become available. The scenario points to the likelihood of using excess storage capacity to collect data outside the study area. Collection of participant data outside the study area would enrich the performance evaluation effort by gathering data to control for 1) confounding factors that are individual-specific, and 2) populating the data to compensate for issues such as sufficient number of events recorded in the study area. To collect data outside of the study area, an approach has been developed that will use a combination of the path history and critical events logging in conjunction with critical event logging and before/after data. The data collection approach will employ the J2945/1 Path history algorithm, which logs BSM data for vehicle breadcrumb if the perpendicular distance is greater than 1 meter.

## 4.2.3 BSM Data Collection Bottleneck Approach

The Tampa site has developed an approach in the event that during operational testing there are breakdowns in the transmission, receipt and collection of BSM data. A hierarchy has been established that is specific for the first few contingencies.

#### *4.2.3.1 Communication Overload at "First" RSU*

One frequently discussed potential problem is that as equipped vehicles encounter the first RSU installed in the study area, the numbers of OBUs and time that will be spent within the range of the roadside unit will be insufficient to download data that has been collected outside of the study area (particularly during the morning peak). If this issue is encountered in the late testing in Phase 2, the plan is to install additional RSUs along the Reversible Express Lanes. These additional units would provide an opportunity to receive data at several intervals prior to a vehicle encountering the "first" RSU and collect data generated outside of study area. It has also been proposed that perhaps these additional units could be used to facilitate any software updates that may be required during the deployment.

#### *4.2.3.2 General Data Bottleneck or Overload*

In the event that the projections of the overall amount of BSM and other data prove to be underestimated, creating transmission, storage or other issues, the first contingency will be to employ the modified data logging approach for the data collected outside of the defined study area.
If these problems were to persist, then the next step would be to filter all out of study area data. Based on the most recent analysis included in this document, additional steps are not anticipated at this time and would need to be developed in a manner that have the least impact to the needs of US DOT and the site evaluation team. Lastly, as a part of the final design of the OBUs and RSUs, a prioritized data filtering process is being developed to account for data overload issues not solved based on the steps detailed here.

## 4.2.4 Alert and OBU Data Log Collection

Beyond attempting to capture all BSMs, alerts and other logging, data will be required to be collected, stored and transmitted to the Master Server and ultimately to CUTR and the IE for evaluation purposes. As with the BSM data approach, a prioritization method is being employed as a decision support mechanism to deal with the potential of too much data.

For the eight vehicle applications that can generate alerts, there are both an audio and a visual cue provided to the driver. Other message data will be critical to receive as well. Table 6 details the data types that may be collected by the OBUs and the priority assigned to those items for storage and transmission.

**Table 6. OBU Message and Alert Data Priority**

| Data Description | Priority Rating |
|---|---|
| | |
| Display activation (graphics change) | Medium |
| WWE screen activation (graphics change) | Medium if Audio alert is captured |
| WWE Audio alert activation | High |
| ERDW screens activation (graphics change) | Medium if Audio alert is captured |
| ERDW audio activation | High |
| VTRFTV screen activations (graphics change) | Medium if Audio alert is captured |
| VTRFTV audio alert activation | High |
| IMA screen activation (graphics change) | Medium if Audio alert is captured |
| IMA audio alert activation | High |
| PED-X screen activation (graphics change) | Medium if Audio alert is captured |
| PED-X audio alert activation | High |
| EEBL screen activation (graphics change) | Medium if Audio alert is captured |
| EEBL audio activation | High |
| FCW screen activation (graphics change) | Medium if Audio alert is captured |
| FCW audio alert activation | High |
| TSP Screen activation (graphics change) | Medium if Audio alert is captured |
| TSP audio alert | High |
| Display - system activated indicator | Low |
| Other OBU output activated, TBD | Low |
| Speed data logged, TBD sampling | High - Other methods available |
| CAN data (not planned) | N/A |
| MAP logging | High |
| RSA logging | High |
| TIM logging | High |
| BSM logging | High |
| SPaT Logging | High |
| PSM logging | High |
| TSP logging | High |
| SSM logging | High |
| SRM logging | High |
| USB data transfer | Medium |
| OTA transfer activation (data transferred) | Medium |
| All Antenna status | Medium |
| Turn signal activation (graphics change) | High |
| Ignition state | Low |
| Reverse state | High |
| SD card activation | Low |
| Tampering/security activation | Medium |
| Firmware download/install | Medium |
| SCMS connection &download time | Medium |

# References

1. Johnson, S., et al., *Connected Vehicle Pilot Deployment Program Phase II Data Privacy Plan - Tampa Hillsborough Expressway Authority (THEA) (Report No. FHWA-JPO-17-461)*. 2017, U.S. Department of Transportation Intelligent Transportation Systems (ITS) Joint Program Office: Washington, DC.
2. Concas, S., A. Kourtellis, and S.L. Reich, *Connected Vehicle Pilot Deployment Program Performance Measurement and Evaluation Support Plan, Phase 2 UPDATE – Tampa (THEA) (Report No. FHWA-JPO-16-314)*. 2019, U.S. Department of Transportation Intelligent Transportation Systems (ITS) Joint Program Office: Washington, DC.
3. Novosad, S., et al., *Connected Vehicle Pilot Deployment Program Phase 1, System Requirements Specification (SyRS) – Tampa (THEA) (Report No. FHWA-JPO-16-315)*. 2016, U.S. Department of Transportation Intelligent Transportation Systems (ITS) Joint Program Office: Washington, DC.
4. Cordahi, G., et al., *Connected Vehicle Pilot Deployment Program Phase 1, Application Deployment Plan – Tampa (THEA) (Report No. FHWA-JPO-16-316)*. 2016, U.S. Department of Transportation Intelligent Transportation Systems (ITS) Joint Program Office: Washington, DC.
5. Johnson, S., et al., *Connected Vehicle Pilot Deployment Program Phase 2 - System Architecture Document – Tampa (THEA) (Report No. FHWA-JPO-17-459)*. 2017, U.S. Department of Transportation Intelligent Transportation Systems (ITS) Joint Program Office: Washington, DC.
6. Beresheim, S., et al., *Connected Vehicle Pilot Deployment Program Phase 1, Safety Management Plan – THEA (Report No.FHWA-JPO-16-313).* 2016, , U.S. Department of Transportation Intelligent Transportation Systems (ITS) Joint Program Office: Washington, DC.
7. Gordon, A. and J. Malik, *Official (ISC)2 Guide to the CISSP CBK*. 2015, Boca Raton, FL: CRC Press.
8. *Security and Privacy Controls for Information Systems and Organizations.* 2020, National Institute of Standards and Technology, U.S. Department of Commerce: Washington, DC.
9. Society of Automotive Engineers, *J2735: Dedicated Short Range Communications (DSRC) Message Set Dictionary.* 2016, SAE International.
10. Society of Automotive Engineers, *J2945: On-Board System Requirements for V2V Safety Communications*. 2016, SAE International.

U.S. Department of Transportation
ITS Joint Program Office-HOIT
1200 New Jersey Avenue, SE
Washington, DC 20590

Toll-Free "Help Line" 866-367-7487
www.its.dot.gov

FHWA-JPO-17-462