# Southeast Michigan Test Bed Advanced Data Capture Field Testing

## Task 4: Operational Data Environment – Concept of Operations

www.its.dot.gov/index.htm

**Draft — July 28, 2015**

**FHWA-JPO-15-227**

U.S. Department of Transportation

Produced by Booz Allen Hamilton
U.S. Department of Transportation
Federal Highway Administration (FHWA)

## Notice



## Quality Assurance Statement

**Technical Report Documentation Page**

| 1. Report No. | 2. Government Accession No. | 3. Recipient's Catalog No. |
|---|---|---|
| FHWA-JPO-15-227 | | |

| 4. Title and Subtitle | 5. Report Date |
|---|---|
| Southeast Michigan Test Bed Advanced Data Capture Field Testing<br>Task 4: Operational Data Environment - Concept of Operations | November 12, 2015 |
| | 6. Performing Organization Code |
| **7. Author(s)**<br>Dwayne Henclewood, Brien Miller, Paul Valdetaro, Hamid Musavi, Ismail Zhody, and Ram Kandarpa | **8. Performing Organization Report No.** |
| **9. Performing Organization Name And Address**<br>Booz Allen Hamilton<br>8283 Greensboro Drive<br>McLean, VA 22102 | **10. Work Unit No. (TRAIS)** |
| | **11. Contract or Grant No.** |
| **12. Sponsoring Agency Name and Address**<br>U.S. Department of Transportation<br>Federal Highway Administration<br>1200 New Jersey Avenue, SE<br>Washington, DC 20590 | **13. Type of Report and Period Covered** |
| | **14. Sponsoring Agency Code**<br>FHWA |

**15. Supplementary Notes**

**16. Abstract**

This document presents the concept of operations (ConOps) of the implementation of the Operational Data Environment (ODE) for the Southeast Michigan Test Bed. The ODE is a smart data router that processes vehicle-to-infrastructure (V2I), as well as infrastructure-to-vehicle (I2V), data collected from connected vehicles, intersections, and traveler advisories by the Southeast Michigan Test Bed's Data Clearinghouse and Data Warehouse, along with other non-connected vehicle sources. Through this process, the ODE provides valuation, aggregation, integration, sanitization, and propagation functions, allowing downstream client applications to receive data on request. The ODE is intended to complement a connected vehicle infrastructure by functioning as a smart data router brokering processed data from various data sources, including connected vehicles, to a variety of data users. Data users include client transportation software applications, such as those that may be used by a transportation management center (TMC). This document provides an overview of the operations and capabilities of the ODE.

| 17. Key Words | 18. Distribution Statement |
|---|---|
| ODE, Operational Data Environment, aggregation, integration, valuation, propagation, v2i, connected vehicles, architectures, big data, real-time processing. | No Restrictions |

| 19. Security Classif. (of this report) | 20. Security Classif. (of this page) | 21. No. of Pages | 22. Price |
|---|---|---|---|
| Unclassified | Unclassified | 44 | N/A |

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

# Table of Contents

## List of Figures

# CHAPTER 1. SCOPE

## 1.1. Identification

This document presents the concept of operations (ConOps) of the implementation of the Operational Data Environment (ODE) in the Southeast Michigan Test Bed. In short, the ODE processes vehicle-to-infrastructure (V2I), as well as infrastructure-to-vehicle (I2V), data collected from connected vehicles, intersections, and traveler advisories by the Southeast Michigan Test Bed's Data Clearinghouse and Data Warehouse, along with other non-connected vehicle sources. Through this process, the ODE provides valuation, aggregation, integration, sanitization, and propagation functions, allowing downstream client applications to receive data on request.

In the recent past, the U.S. Department of Transportation (USDOT) invested in a separate system, which is referred to as the Prototype-Operational Data Environment (P-ODE). While the ODE and P-ODE are similar in name and have a few common functionalities, they are two different systems, developed with different objectives—the main objective of the P-ODE being the foundational concept of the ODE. Mention of the P-ODE in this document is limited to aspects regarding the government's continued efforts toward supporting connected vehicles and the few common capabilities between the two systems.

## 1.2. Document Overview

This document provides a ConOps for the ODE. The document's goals are to define the system requirements, describe the development of functionalities to fulfill those requirements, highlight the key functional components and connections throughout the system, and show how various system/user requirements are met.

This document adheres to the Institute of Electrical and Electronics Engineers (IEEE) guidelines for ConOps documentation as much as possible; however, a few sections deviate from what is prescribed by IEEE to best present some of the components and features of the ODE.

The remaining chapters of this document are:

- **Chapter 2. Referenced Documents:** Lists several documents that are referenced throughout this document to describe of the operation of the ODE.
- **Chapter 3. Current System/Situation:** Presents a high-level summary of the current situation as it relates to the USDOT's efforts in developing connected vehicle technology and applications. Two of the more recent efforts highlighted include the P-ODE and the Southeast Michigan Test Bed project.

- **Chapter 4. Justification for Changes/Modifications:** Highlights the two core drivers for the development of the ODE:
  - Need to further capture and realize the opportunities of connected vehicle technology
  - Ability of the ODE to effectively leverage other connected vehicle technology development efforts supported by the USDOT.
- **Chapter 5. Concepts for the Proposed Operational Data Environment:** Details the core concepts of the functionalities and capabilities of the various components of the proposed ODE.
- **Chapter 6. User-Oriented Operational Scenarios:** Describes a series of step-by-step processes that allow readers to envision how the various components of the system, as well as the entire system will operate.
- **Chapter 7. Analysis of the Proposed System:** Provides a high-level summary of the benefits of the ODE and opportunities to increase the ODE's service offerings.

## 1.3. System Overview

The ODE is intended to complement a connected vehicle infrastructure by functioning as a smart data router brokering processed data from various data sources, including connected vehicles, to a variety of data users. Data users include client transportation software applications, such as those that may be used by a transportation management center (TMC). This document provides a high-level overview of the operations and capabilities of the ODE.

As a smart data router, the ODE routes data from disparate data sources to software applications that have placed data subscription requests to the ODE. These subscribing applications may include connected vehicle applications. While provisioning data from data sources for data users, the ODE will also be performing necessary security and credential checks and, as needed, data valuation, aggregation, integration, sanitization, and propagation functions. These are core functions to the ODE and will be detailed in later sections. However, in summary:

- Valuation is the process of making a judgment about the quality or value of the data and taking the appropriate action.
- Aggregation is the creation of composite or summary information from more granular data.
- Integration is the process of combining different data from multiple sources to provide more complete information.
- Sanitization is the processing of data, in its original form, to prevent the discovery of personal identifiable information (PII).
- Propagation is the acquisition and distribution of data to client applications requesting data services.

Figure 1 represents a high-level contextual representation of the ODE, presenting the connection between the data sources and data users. This figure shows a number of key components, functionalities, and capabilities of the system, each of which will be detailed in subsequent chapters. At the core of the ODE's operation is its subscription to the Southeast Michigan Test Bed Situational Data Clearinghouse (SDC), its connection to the Southeast Michigan Test Bed Situational Data Warehouse (SDW), and its ability to ingest data from other data sources. These other data sources may include repositories that contain data such as weather (the state of the atmosphere), road weather (the state of roadways given the impact of atmospheric weather events), traffic volume counts, incident information, special event calendars, and transportation management strategies. Depending on the data type and source, the data that are ingested by the ODE are subject to a series of data valuation, aggregation, and integration processes, as well as data sanitization. The ODE will receive data requests from subscribing client applications and will transmit the requested data to these applications to support their operations. Later chapters of this document provide additional details on the operation of the ODE, including its various processes.

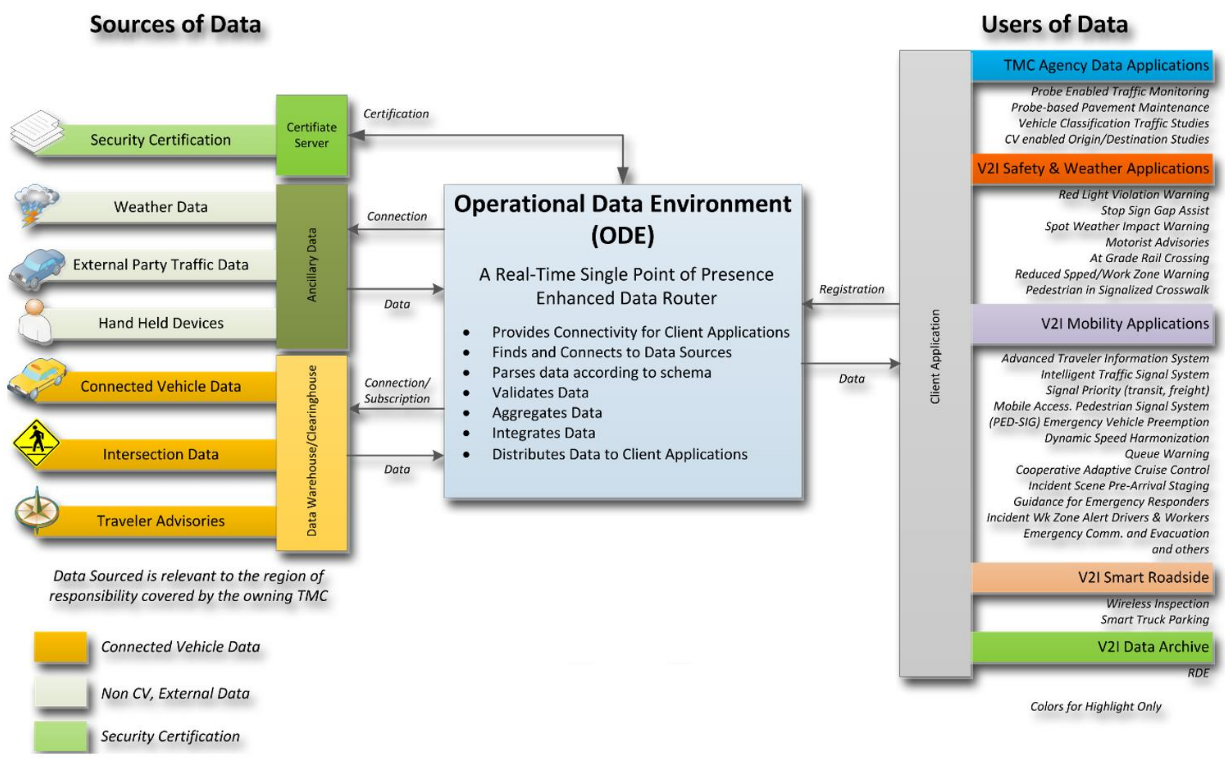



**Figure 1. High-Level Contextual Representation of the ODE, Connecting Data Source to Data Users**

This page intentionally left blank

# CHAPTER 2. REFERENCED DOCUMENTS

The following resources were used to support the development of this ConOps document. This list includes both published and unpublished works.


- *Operation of the Situation Data Clearinghouse and Warehouse at the ITS World Congress in September 2014 – Task 2 Report*, November 2014
- *Southeast Michigan Test Bed Advanced Data Capture Field Testing  Task 3: Aggregator/Integrator Enhanced ODE White Paper*, February 2015
- Fehr, Walt, *Southeast Michigan Test Bed 2014 Concept of Operations*, September 17, 2014
- Leidos, *Prototype Operational Data Environment Data Concept of Operations*, September, 2014
- Leidos, *Prototype Operational Data Environment - DMA Emulator Concept of Operations*, August, 2014
- IEEE, *IEEE Guide for Information Technology System Definition - Concept of Operations (ConOps)*, March 1998.

This page intentionally left blank

# CHAPTER 3. CURRENT SYSTEM/SITUATION

## 3.1. Background, Objectives, and Scope

The promise of connected vehicle technology has been the focus of many research efforts, both public and private in nature. These efforts span various types of institutions including academia, government, and private industry—many of which belong to the automotive and communication sectors. Given the complex nature of the infrastructure that will eventually be needed to support this technology, a series of connected vehicle test beds have been created. These test beds allow researchers to further advance the state of the art and the state of the practice for connected vehicle technology. Research and development performed using the test beds allows for the identification and distribution of lessons learned to support the continued development of this technology. To enable the use of connected vehicle applications, it has become clear through such research efforts that having a mechanism to ingest, process, and output pertinent data from connected vehicles and connected infrastructure to client (transportation) applications is imperative. It is with this concept in mind that the ODE is being developed.

To date, extensive work is ongoing to develop a connected vehicle test bed in Southeast Michigan. Emphasis has been placed on the secure and trusted communication of broadcast safety messages and peer-to-peer encrypted information exchanges. *Hypercurrent* and *hyperlocal* messages, which are high-frequency messages transmitted at approximately 10Hz, are broadcast to other vehicles for real-time use by safety applications. This includes basic safety messages (BSMs) that are received by roadside units (RSUs), as well as signal phase and timing (SPaT) messages sent by signal controllers. BSMs are captured as enhanced vehicle situation data (EVSD), and SPaT messages are captured as intersection situation data (ISD). Both EVSD and ISD are sent to the SDC for real-time distribution. Travel advisory messages are captured as traveler situation data (TSD) and are available from the SDW, for a configurable time period. In addition to storing TSD, the SDW also stores archived EVSD and ISD, again for a configurable time period. The data clearinghouse and data warehouse (where data from the test bed are stored) do not perform any quality checking or data bundling operations. Connected vehicle applications or data aggregators may subscribe to the real-time data clearinghouse or query from the data warehouse to obtain data on time and/or location.

Due to the work being done in the Southeast Michigan Test Bed, the ODE presented here not only aims to support the general development of connected vehicle technology; more specifically, it also provides an initial implementation of quality checking, aggregation, and provision (real-time and archived) of data from the test bed (and other sources) to various client applications.

The ODE is a data provisioning system that is being developed for integration with the Southeast Michigan Test Bed. This construct is in a similar vein to the government's ongoing effort in augmenting the Northern Virginia Connected Vehicle Test Bed with the P-ODE (which is described below in more detail). The remainder of this chapter provides a synopsis of the current situation. It includes a high-level description of the P-ODE's functionality, as well as describes some of the key components of the Southeast Michigan Test Bed, specifically those components with which the ODE will interface to obtain data from the test bed. Key points regarding the operation of the Southeast Michigan Test Bed, particularly those that will impact the development and operation of the ODE, follow.

## 3.2. Description of the P-ODE and the Southeast Michigan Test Bed

### 3.2.1. The P-ODE

As stated in the P-ODE ConOps document, the P-ODE is:

> *"intended to extend the scope of the DCM program to include near real-time data sources and publish-subscribe interfaces. The ODE collects data in real-time from multiple sources and publishes quality checked data to subscribers as well as the Research Data Exchange (RDE), removing any PII contained in the data prior to publication."*

- P-ODE Concept of Operation 09/2014, Leidos

The P-ODE is currently being developed and operated in conjunction with the Northern Virginia Test Bed. The Northern Virginia Test Bed provides a variety of data sources to the P-ODE to support its mission. Data that are currently being collected in support of the P-ODE include those provided by the Saxton Transportation Operations Laboratory (STOL) at the Turner Fairbank Highway Research Center (TFHRC), the Virginia Department of Transportation (VDOT), and the Regional Integrated Transportation Information System (RITIS)—all for the northern Virginia area.

These data are then transmitted to the subscribing Dynamic Mobility Application (DMA) Emulator. As written in the DMA Emulator ConOps, the emulator is intended to:

> *"demonstrate an application subscribing to multiple data types and multiple data sources from the P-ODE to support Traffic Management Operations. The intent is not to define rigorous, ironclad algorithms, but to demonstrate the ability of applications to use multiple data types from multiple data sources. To demonstrate the power of the P-ODE, the emulator will calculate travel times along I-66, outside of the Northern Virginia*

> *Beltway, by subscribing to BSMs, Speed\Travel Time, Volume, Occupancy, and Weather related data"* [which is being outputted from the Northern Virginia Test bed area].
>
> - DMA Emulator Concept of Operation 09/2014, Leidos

This concept of having applications supporting transportation operation subscribe to a single repository that is a sink for data from multiple sources is similar to the concept of the ODE. The ODE's ultimate scope is beyond the P-ODE functionality described here; however, one of its first sources of data will be generated from the Southeast Michigan Test Bed, versus the Northern Virginia Test Bed.

### 3.2.2. The Southeast Michigan Test Bed

The Southeast Michigan Test Bed was first established in 2007 to serve as a test environment for the USDOT and the automotive industry to examine the feasibility of dedicated short-range communication (DSRC) in the 5.9 GHz spectrum. In recent years, the Southeast Michigan Test Bed has undergone a number of upgrades and now serves as a testing facility for the latest innovations in connected vehicle technology. This test bed is constantly evolving to support the rapidly developing connected vehicle technologies to better meet the needs of the emerging connected vehicle community and provide support for the future pilot deployments.

This section provides a high-level overview of the Southeast Michigan Test Bed and some of its capabilities, particularly those that are relevant to the development of the proposed ODE. As the name suggests, the Southeast Michigan Test Bed is located in the southeastern region of the state of Michigan. Figure 2 presents the geographic boundary of the test bed.


**Figure 2. Southeast Michigan Test Bed Geographic Boundary**

Figure 3 presents the high-level physical layer of the Southeast Michigan Test Bed architecture. It provides a high-level depiction of the test bed components. For a detailed description of each of the components and their sub-components, view the *Southeast Michigan Test Bed Concept of Operations* document and other descriptive documents, some of which are available at [http://cps-vo.org/node/8937/browser](http://cps-vo.org/node/8937/browser) (Note: This website is not publicly accessible and requires access authorization from the USDOT.) This document defines a few key components of the test bed as they pertain to the development and operation of the ODE.

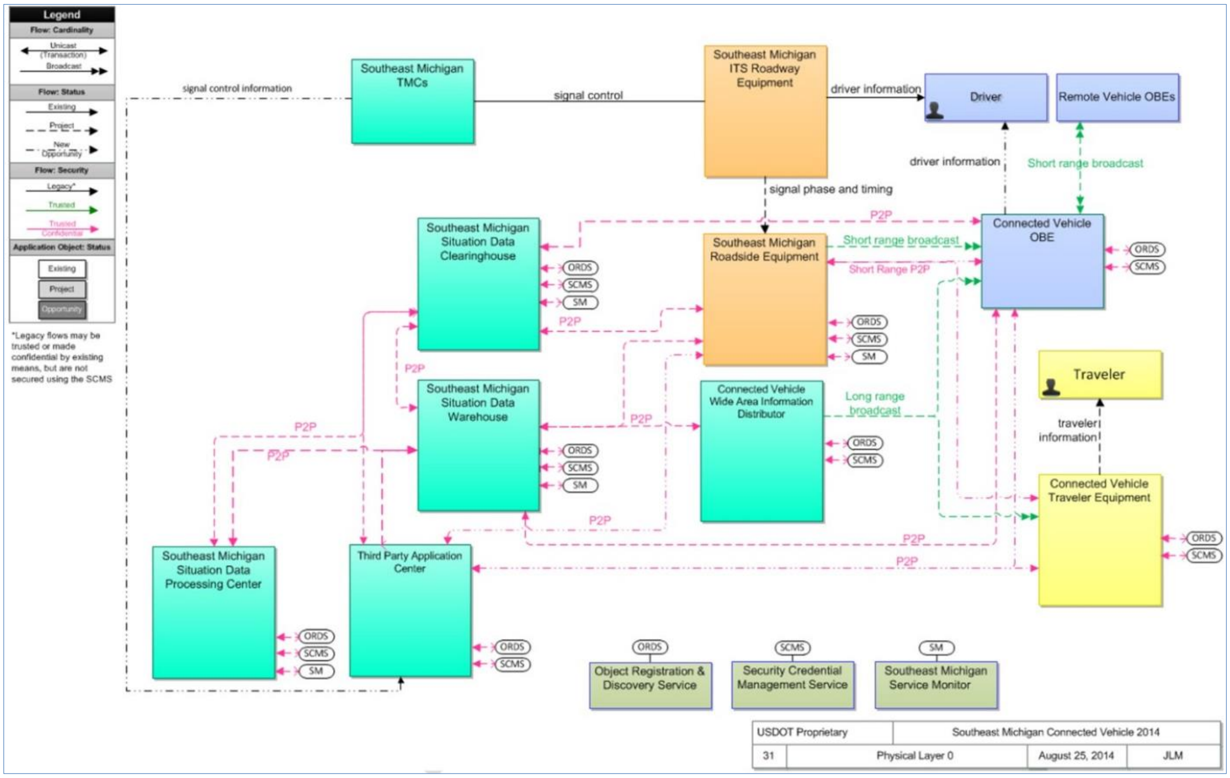



**Figure 3. The Physical Architecture (Level 0) of the Southeast Michigan Test Bed**

The ODE will interact with Southeast Michigan Test Bed through three of its primary components via their provided application-programming interfaces (APIs): 1) the SDC, 2) the SDW, and 3) the Situation Data Processing Center (SDPC), collectively referred to herein as the USDOT Data Distribution System (DDS) based on Connected Vehicle Reference Implementation Architecture (CVRIA) nomenclature. The following section offers a brief description of each of these components and the means through which they may be accessed.

### *Test Bed Components to which the ODE Connects to Support Its Operation*

As mentioned above, the ODE will interact with the test bed via the following three primary components:

1. **USDOT SDC**: The SDC collects, processes, and distributes connected vehicle data of short-term utility; that is, data that are useful now, and typically over a small area. It consumes data from data producers, such as connected vehicles, and provides that data to interested data consumers using a publish-and-subscribe mechanism. It does not store data longer than is necessary to satisfy the publication mechanism.
2. **USDOT SDW:** The SDW is a data distribution system that collects, processes, and distributes connected vehicle data, connecting data producers with data consumers and facilitating data exchange in the connected vehicle environment. It focuses on data that is relevant for a period beyond the immediate (10 minutes or more), and distributes that data to

interested parties using a query mechanism. The SDW data retention window is designed to allow data consumers to query for relevant data whenever the supporting use case does not justify a constant stream of data as is provided by the SDC. It may discard data when it is no longer relevant.

3. **USDOT SDPC:** The Southeast Michigan SDPC monitors and controls traffic and the road network. It represents centers that manage a broad range of transportation facilities including freeway systems, rural and suburban highway systems, and urban and suburban traffic control systems. It communicates with Southeast Michigan ITS roadway equipment and Southeast Michigan roadside equipment (RSE) to monitor and manage traffic flow and monitor the condition of the roadway, surrounding environmental conditions, and field equipment status. It manages traffic and transportation resources to support allied agencies in responding to, and recovering from, incidents ranging from minor traffic incidents through major disasters.

The ODE will connect to each of the above components through the appropriate API to submit subscription requests and queries, and receive data from the Southeast Michigan Test Bed.

***Available Data from the Southeast Michigan Test Bed***

The data available via Southeast Michigan Test Bed is distributed in the form of three message types:

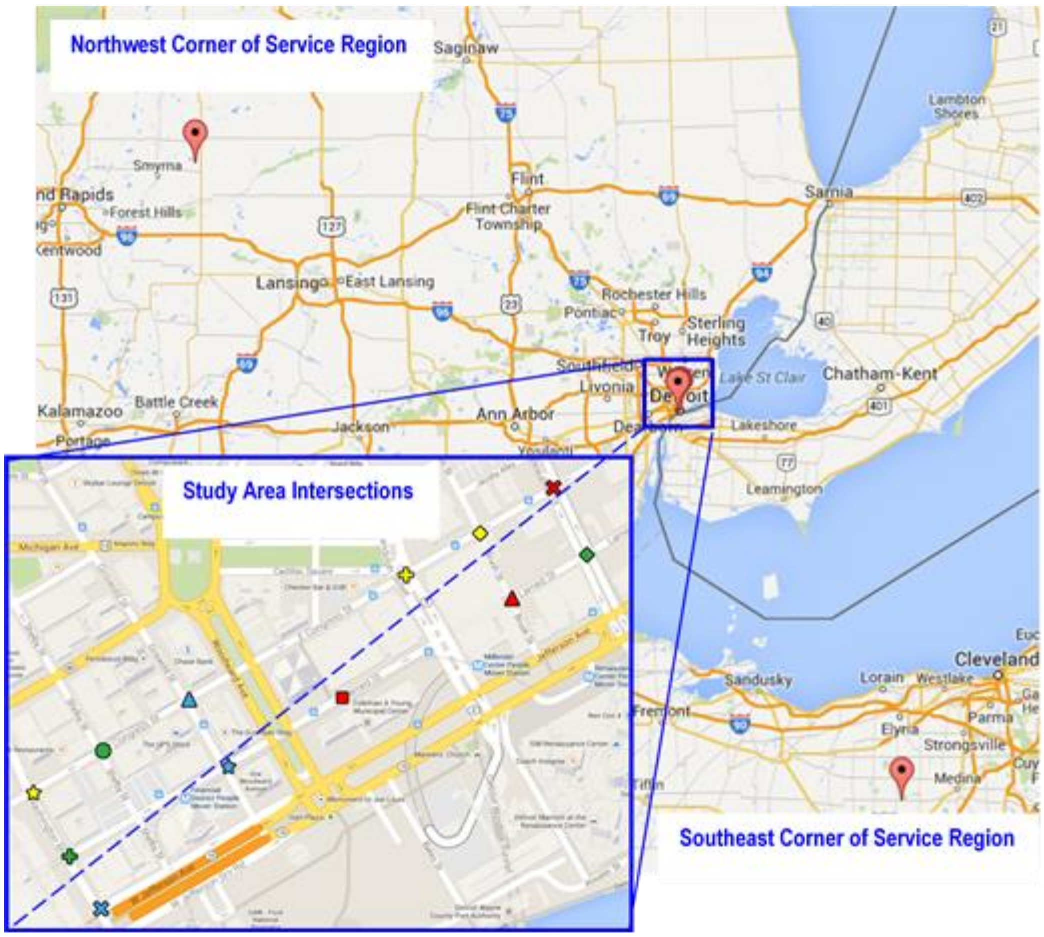
1. EVSD, which are provided by participating vehicles
2. ISD, which include SPaT and MAP for a given intersection, and are provided by participating RSEs (and by TMCs acting as a "proxy RSE" for intersections, signalized or otherwise, without an RSE)
3. TSD, which are generated by the SDW based on individual TSD messages that are provided by the SDPC, and can be provided by other TMCs and transportation information centers.

The following sections briefly describe each of these message types.

**EVSD:** Nine vehicles equipped with onboard equipment are generating BSM data elements while navigating the Southeast Michigan Test Bed area. An EVSD message is composed of data elements that have been mapped from BSMs. The EVSD message primarily contains data elements that are members of the BSM Part I message set. These data elements include vehicle position (latitude, longitude, and elevation), speed, acceleration, and brake system status. The EVSD also include a vehicle "path history" (breadcrumb trail), when available, which is usually a part of the BSM Part II message set. All the data elements and their definitions are consistent with the details presented in the SAE J2735 Standard.

These EVSD messages are broadcast, from an equipped vehicle, to the SDC, with the RSU serving as a gateway to pass along the messages to the SDC.

**ISD:** Twelve intersections are equipped with RSUs that collect and transmit ISD in the downtown Detroit area. Figure 4 shows the locations of these 12 intersections, along with their IDs. (Note: Each colored symbol in the figure depicts a signalized intersection from which data was obtained during an observation period.) ISD messages are composed of MAP and SPaT information. MAP information contains an intersection's location (latitude and longitude), elevation, and geometric features such as number of approaches and lane configurations. SPaT data contains the current state of the intersection's signal indication(s) and the timing of the next phase. Transmitted ISD messages are received by the SDC and, as such, are available via subscription to the SDC.




*The various symbols represent different intersections*

**Figure 4. Geospatial Representation of the Southeast Michigan Test Bed and Its Instrumented Intersections**

**TSD:** TSD is obtained from the SDW and not the SDC. To access TSD from the SDW, queries are submitted to the SDW, and the results are returned via a secured connection. This message type contains messages to be delivered to roadway users to aid in their navigation of the network.

These messages are similar to messages on variable message signs communicating information such as curve speed warning and road weather alerts.

## 3.3. Current Mode of Operation of the Southeast Michigan Test Bed

The ODE interacts with the Southeast Michigan Test Bed through secure connections to the SDC and the SDW. As previously mentioned, these are data repositories from which three message types are obtained—the EVSD, ISD, and TSD. The operational mode of the Southeast Michigan Test Bed may be best described in terms of when these data types are generated. Currently, EVSDs are produced by the limited number of equipped vehicles. These vehicles are only sent out into the field at select times. The specific times are largely driven by current publicity activities to help make the ITS community aware of the capabilities of the Southeast Michigan Test Bed and subsequently invite them to use the Test Bed facilities to help develop their connected vehicle applications and/or technologies. These vehicles also traverse the Southeast Michigan Test Bed during arranged intervals in support of field experiments or tests. While the availability of EVSD is limited at the moment, this will not constrain either the development or operation of the ODE. It is anticipated that the ODE project team and Southeast Michigan Test Bed project team will coordinate efforts so that when EVSD is needed by applications subscribed to the ODE, the necessary arrangements will be made to facilitate this.

ISD are consistently being transmitted from equipped, signalized intersections to the SDC and subsequently to the SDW, and may be accessed from the SDW via queries. Similarly, TSD messages are available by querying the SDW. To date, there are approximately 20 sample TSD messages available to be queried at any point in time. This perpetual availability of ISD and TSD enables the ODE to supply its subscribing applications with these data, at any point, to support their missions.

## 3.4. End Users

This section of a ConOps is traditionally meant to communicate with the set of users that is interacting with the current system. Since there is no currently existing ODE, this section highlights three potential user classes that are envisioned to interact with both today’s Southeast Michigan Test Bed and the developing ODE. These three user classes are:

1. Connected vehicle technology developers
2. Connected vehicle application developers
3. Connected vehicle/ITS research community (potentially including the Connected Vehicle Pilot Deployment team).

The connected vehicle technology developer class is intended to reflect entities interested in conducting real-world field testing of various connected vehicle technologies being developed. Such technologies may vary from the application of other wireless technologies that complement or supplement DSRC to improved transmission antennas that may increase the range at which messages maybe sent and received. While connected vehicle technology developers may not directly interact with the ODE, subscribers to the ODE will want to be aware of the technologies implemented throughout the Southeast Michigan Test Bed to ensure that the requested data are being interpreted and incorporated correctly.

As for connected vehicle application developers, EVSD, ISD, and TSD from the Southeast Michigan Test Bed are currently available to these entities to support the development of their various applications. These may or may not include an application that is aligned with any of the already defined connected vehicle application program areas. While these data may sufficiently support applications that are currently (or soon to be) under development, as the connected vehicle field evolves, more requirements may be placed on these applications to further support transportation operation and management. Thus, this user class should be considered when building out the future of connected vehicle technology, as these demands may warrant additional data or information to support their goals.

Similar to the demands of connected vehicle application developers for additional data to support their work, the research community may also demand more of the data being collected to extract greater insights to advance the transportation practice. Asking for more of the data can take many forms including obtaining a more diverse and/or larger sample size with greater accuracies. This user class' potential data needs in the connected vehicle realm may serve as a driver of means to continue to extract and provide as much data as possible from real-world connected vehicle deployments.

This page intentionally left blank

# CHAPTER 4. JUSTIFICATION FOR CHANGES/MODIFICATIONS

In accordance to the IEEE standard for developing ConOps documentation, this chapter describes shortcomings of the current system or situation that are driving the development of the proposed system. As previously stated, currently there is no deployed ODE. Three of the more notable drivers for development of the ODE are:

1. The need for providing connected vehicle data to applications in a reliable, useful, and uniform manner during field operational testing of connected vehicle technologies and applications
2. The need for providing connected vehicle data to applications in a reliable, useful and uniform manner to enhance and transform current surface transportation maintenance and operational practices
3. The need to support the ongoing conversation regarding connected vehicle data sharing policies, security concerns, and other related institutional policies per the implementation of connected vehicle technologies.

## 4.1. Providing Data for Connected Vehicle Development Efforts

The USDOT and many private and public entities have been exploring the feasibility of a connected transportation system. The efforts of these entities have demonstrated the feasibility of real-world implementation of connected vehicle and infrastructure technologies as well as their benefits, albeit in simulated environments or through relatively small-scale studies. With this promise of connected vehicle technology, the next series of explorations are trending toward larger-scale field tests, implementations, and simultaneous evaluations of multiple technologies and connected vehicle applications. As connected vehicle technology and application deployments grow in scope and scale, a uniform, reliable, and useful method of ingesting and providing data is needed. This is one of the motivational factors for developing the ODE.

## 4.2. Leveraging Connected Vehicle Data to Support Transportation Operations and Maintenance

In addition to establishing the Southeast Michigan Test Bed as a real-world connected vehicle implementation, the USDOT is investing in several other efforts, including the upcoming Connected Vehicle Pilot Deployments, that will result in the development of other real-world instances of a connected vehicle infrastructure. The expectation is that the development efforts underway now will eventually yield products, systems, and services that will enhance and transform current surface transportation operation and maintenance practices.

In line with these efforts and continued preparation to extract the benefits from a *fully* connected vehicle infrastructure, there is a need to develop the appropriate mechanisms to help provide these benefits. The proposed ODE will likely provide the foundation for traffic management centers (TMCs) to extract value-added data from a connected transportation network to effectively support their operation and maintenance needs.

To meet these needs, the ODE:

- Is modular in development to support transferability and scalability
- Includes data valuation, integration, aggregation, quality checking, and sanitization capabilities
- Embeds privacy and security measures that are needed to protect data providers
- Has the capability to perform real-time processing of large data volumes.

## 4.3. Support the Development of Institutional Policies per Connected Vehicle Technology

As the transportation community marches toward advancing the state of the implementation of connected vehicle technologies, opportunities to leverage the use of the generated data are becoming prolific. As these opportunities present themselves and various entities either attempt to use the data or provision for others' use in supporting additional advances in the connected vehicle domain, discussions are occurring to define appropriate use of these data and how they should be shared.

As one of the core functions of the ODE is to provision the data to support the leveraging of connected vehicle data, it is envisioned that building this system and facilitating user interactions will in turn add to the conversation regarding the best practices to provision data for reuse while curtailing inappropriate use of the data. The ODE will be uniquely positioned to facilitate this conversation. The ODE will not only represent the state of the art of a connected vehicle data provisioning system, but it is also envisioned to be the foundation upon which future connected vehicle data provisioning systems will be built to support full scale, real-world implementation of a connected vehicle infrastructure. This system will in turn support the USDOT, and it's various partners, to help inform policies surrounding the distribution and use of connected vehicle data and the requisite overarching security measures that will be needed to protect all involved in materializing the promise of connected vehicle technologies.

# CHAPTER 5. CONCEPTS FOR THE PROPOSED OPERATIONAL DATA ENVIRONMENT

## 5.1. Background, Objectives, and Scope

The ODE is a real-time, data router. It processes vehicle and infrastructure data collected from the SDC and the SDW, along with other non-connected vehicle sourced data. The ODE provides data valuation, aggregation, integration, sanitization, and propagation functions. These functionalities of the ODE allow downstream client applications to receive forwarded data streams, integrated data, and aggregated data of interest upon request. The current scope of the ODE's development is to augment the operation of the Southeast Michigan Test Bed with additional data processing functionalities to support quality checking and sanitization of connected vehicle data, as well as provide data aggregation, integration, and distribution capabilities. It is anticipated that additional capabilities will be added to the ODE in the near future. This will in turn extend the Southeast Michigan Test Bed's ability to further support connected vehicle applications, while simultaneously bolstering the ODE's ability to provide enhanced data routing capabilities.

On a more granular level, the ODE is a single data source for client applications that enables ultimate end users to access and gain insights from connected vehicle and other related data to support their efforts. Figure 5 presents a high-level illustration of how the ODE provisions data from various sources to a variety of client applications. The ODE provides its client applications (shown on the right as a list of connected vehicle applications) with a standardized interface to which they can securely connect and request data. While responding to the request, the ODE will perform additional data processing, such as data valuation and data aggregation.

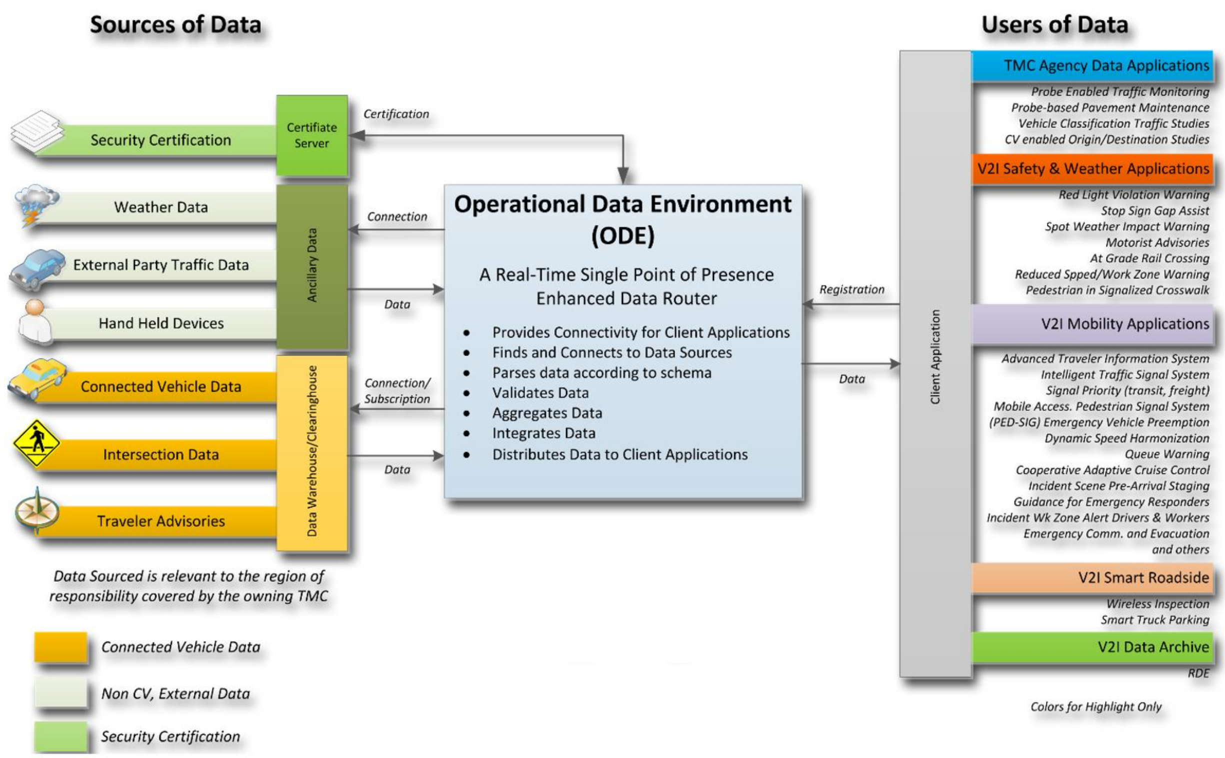



**Figure 5. High-Level Contextual Representation of the ODE, Connecting Data Source to Data Users**

The ODE processes provide the capability to parse its client applications' requested data needs, configure connections, and source data to service all requests efficiently. The ODE has secure mechanisms for connection to data sources and data collection from each source. The components of this system run real-time processes aggregating and integrating data, matching up users and their data, and performing data valuation to verify receipt of valid data (e.g., the inbound messages come in as expected and the data appears useable).

The ODE has the ability to scale up and down, by adding or removing computational resources, to meet the data demands of client applications. The design approach of this system will fulfill all client application requests. The ODE will employ a series of methodologies and techniques, including those that are often associated with processing *large* volumes of data, to ingest and distribute data as rapidly as possible to meet the needs of its client applications.

The remainder of this chapter provides additional details regarding some of the central components and features of the ODE.

## 5.2. Description of the Proposed System

This section provides an overview of the general operational concepts of the ODE by:

1. Highlighting how the ODE integrates with the Southeast Michigan Test Bed
2. Presenting the high-level requirements of the ODE
3. Providing a walkthrough of an application's interaction with the ODE.

Some of the details are informed by the *Southeast Michigan Test Bed Advanced Data Capture Field Testing Task 3: Aggregator/Integrator Enhanced ODE White Paper*[1]. Refer to this white paper for additional information regarding these concepts.

### 5.2.1. ODE Integration with the Southeast Michigan Test Bed

As previously mentioned, the ODE will predominantly engage with three primary components of the Southeast Michigan Test bed:

1. USDOT SDC
2. USDOT SDW
3. USDOT SDPC.

To show this interaction, Figure 6 illustrates the ODE system architecture at level 1. The diagram identifies "application objects" that are involved with the operation of the system as it pertains to the ODE and the ODE itself.

[1] *Southeast Michigan Test Bed Advanced Data Capture Field Testing Task 3: Aggregator/Integrator Enhanced ODE White Paper*, February 2015

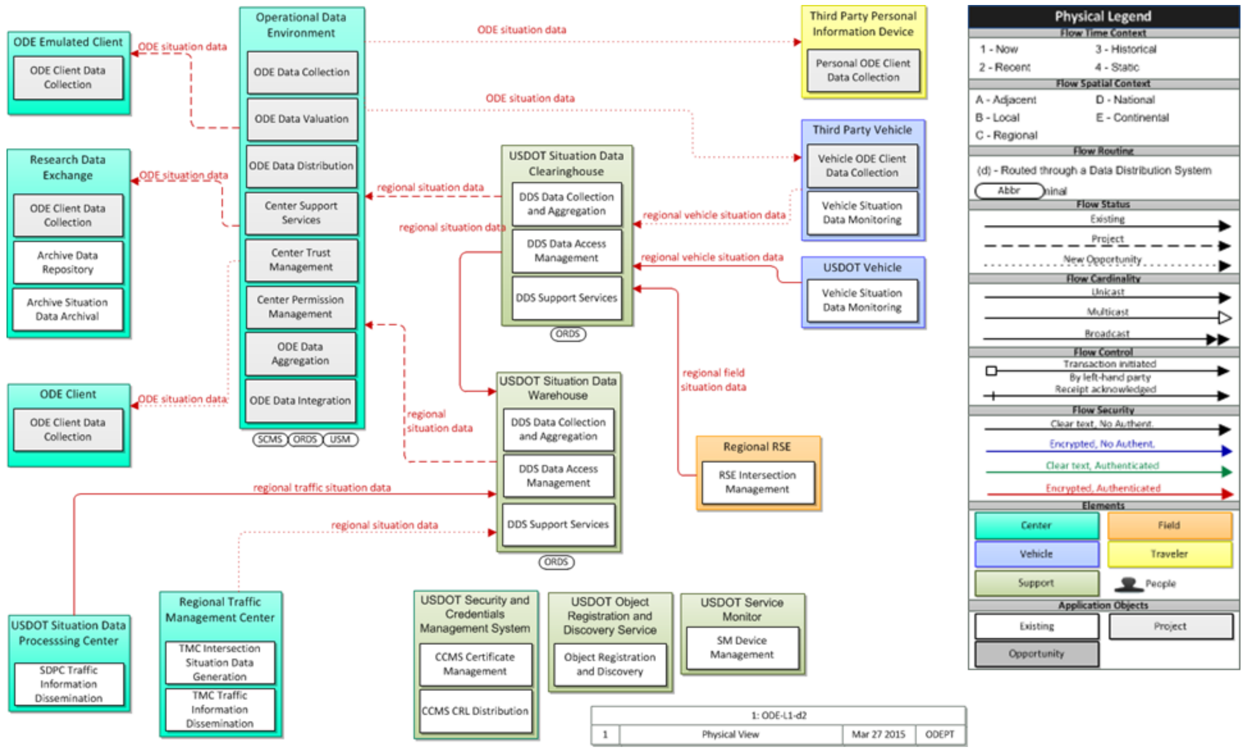



**Figure 6. System Architecture Physical View – Level 1**

## 5.2.2. High-Level Requirements of the ODE

This section presents the high-level requirements for the operation of the ODE, grouped into three categories: 1) general ODE requirements, 2) data processing requirements, and 3) analytics requirements. The data processing category includes requirements aligned to the four core functions of the ODE, termed the VISA (Validation-Integration-Sanitization-Aggregation) functions. Given the criticality of these functions to the operation of the ODE, the definition and implementation for each are presented prior to listing the function's requirements. Each function's definition is directly informed by its definition and description in the Task 3 White Paper[2]. For additional information regarding these functions and other implementation features, please view the Task 3 White Paper.

Before listing the requirements for each category, the following are some key assumptions governing the development and operation of the ODE under this effort:

- The system as defined by these requirements is a proof-of-concept system. It will not be used in a real-life operational setting.

[2] *Southeast Michigan Test Bed Advanced Data Capture Field Testing Task 3: Aggregator/Integrator Enhanced ODE White Paper*, February 2015

- All data sources are known at design time. All the requirements in this document refer explicitly to these sources. Adding other data sources may introduce additional requirements.
- The applications that consume data from the ODE are emulated. They are not real-life applications used in any operational situation.
- Few, if any, performance requirements are being captured here. One of the goals of the ODE is to understand the performance issues related to supplying data to subscribing applications.
- Connected vehicle data will be available from the Southeast Michigan Test Bed.

### *General ODE Requirements*

General requirements include:

5.2.2.1 The ODE shall allow a system administrator (super user) to set the parameters for data validation (e.g., ranges of validity for ingested data).

5.2.2.2 The ODE shall allow a client application to register for data access.

Provision and authentication requirements include:

5.2.2.3 The ODE shall require all client applications to authenticate using a user name and password combination.

5.2.2.4 The ODE may require all client applications to connect over HTTPS.

5.2.2.5 The ODE shall authenticate with all its data sources using whichever authentication the data sources require.

Data requirements include:

5.2.2.6 The ODE shall interface with the SDC in the Southeast Michigan Test Bed.

5.2.2.7 The ODE shall interface with the SDW in the Southeast Michigan Test Bed.
5.2.2.8 The ODE shall ingest streaming EVSD from the SDC.

5.2.2.9 The ODE shall query the SDW or SDPC for archived EVSD.

5.2.2.10 The ODE shall ingest streaming ISD from the SDC.

5.2.2.11 The ODE hall query the SDW or SDPC for archived ISD.

5.2.2.12 The ODE shall ingest streaming TSD from the SDW.

5.2.2.13 The ODE shall query the SDW for archived TSD.

5.2.2.14 The ODE shall allow importing of road network data.

5.2.2.15 The ODE may ingest simple weather data (e.g., atmospheric temperature, barometric pressure).

5.2.2.16 The ODE may allow applications to deposit EVSD, ISD, and ASD to the SDW.

5.2.2.17 The ODE shall disseminate data to consumer application in the form of JSON payloads.

Streaming subscription requests include:

5.2.2.18 The ODE shall allow consumer applications to specify the geographic area for which they wish to receive data.

Query requests include:

5.2.2.19 The ODE shall allow consumer applications to provide the following parameters when requesting archived data—data source, data type, geographic region, and time interval.

### *Data Processing Requirements*

#### *Validation (V)*

**Definition:** Validation is the application of identifiable rules to data to ensure they meet minimum levels of acceptable quality, and are therefore suitable for use in a specific application. When data does not meet the established minimum levels of acceptance, this data will be flagged so that subscribing applications are aware of the data's validity.

**Implementation:** To implement the Validation function, logic will be developed to identify inconsistencies in the data requested by an application. This logic will identify two types of inconsistencies:

1. *Subscription:* Verifies that the requested data are consistent with a client application's subscription
2. *Performance and Operation:* These attributes, mostly aligning to EVSD records, will be validated against a set of reasonable ranges for each vehicle performance measure.

Once data inconsistencies are identified, the ODE will propagate these data to client applications. However, these data will be flagged, indicating their erroneous values.

**Requirements:**

5.2.2.20 The ODE shall remove any ingested data that was generated in a geographic location outside the geographic area specified by the subscription request.

5.2.2.21 The ODE shall remove any ingested data that was generated outside the temporal window specified by the subscription request.

5.2.2.22 The ODE shall mark as invalid any ingested data for which the value falls outside a predefined range.

5.2.2.23 The ODE shall distribute all invalid data to the requesting applications.

5.2.2.24 The ODE shall mark invalid data with an error code that indicates why the data was deemed invalid.

5.2.2.25 The ODE shall mark invalid data with a specific and human-readable description of why the data was deemed invalid (e.g., “vehicle speed of 207 mph exceeds upper range limit of 200 mph”).

### *Integration (I)*

**Definition:** Integration is the ability to combine data from multiple sources to provide more complete information to subscribing client applications.

**Implementation:** The ODE will take streamed EVSD and integrate a static file containing atmospheric and, when available, road weather data in a format that is consistent with the Weather Data Environment (WxDE). This will support having weather, temperature, and other available WxDE data appended to the EVSD record, when a vehicle enters a defined area for which we have data.

**Requirements:**

5.2.2.26 The ODE shall integrate EVSD with temperature data from at least one meteorological source.

5.2.2.27 The ODE shall perform the integration for all subscriptions.

5.2.2.28 The ODE shall NOT integrate data that was marked as invalid during validation.

### *Sanitization (S)*

**Definition**: Sanitization is the processing of data, in its original form, to prevent the discovery of PII.

**Implementation:** While it is envisioned that transmitted data from the Southeast Michigan Test Bed will not contain PII, this function will demonstrate the ODE’s capability to purge some data, defined as sensitive, for pre-identified vehicles. A list of vehicle IDs and corresponding pre-defined bounding boxes is required to deploy this sanitization method. Once a specific vehicle enters one of its sensitive areas, that portion of its trip trajectory will not be propagated to the subscribing client. In addition, the ODE shall also have the capacity to remove or obfuscate/ encrypt vehicle IDs in the EVSD record to further sanitize data from the Southeast Michigan Test Bed.

**Requirements:**

5.2.2.29 The ODE shall remove all data that falls within predefined exclusion areas.

5.2.2.30 The ODE shall not remove data for vehicles within a predefined group of vehicle IDs, because these vehicles are considered exempt from sanitization (there is no expectation of privacy for these vehicles).

#### *Aggregation*

**Definition:** The ability to compute summary information from more granular data (e.g., count, average), triggered by client application requests or the ODE's default setting to provide a specific transportation/roadway system performance metric(s).

**Implementation:** The ODE will make available two aggregation methods:

*Cardinal Aggregation:* Without the association of road segment information to EVSD records, this method provides summary information (e.g., minimum, maximum, mean, and median) based on a coarse list of compass directions (N, NW, W, SW, S, SE, E, NE). Summary information will be calculated by snapping EVSD record to a cardinal direction based on real-time heading.

*Segment Aggregation:* Given the association of road segment information to EVSD records, the ODE is able to provide summary information such as minimum, maximum, mean, and median for a specified road segment.

**Requirements:**

5.2.2.31 The ODE shall allow consumer applications to request aggregate data.

5.2.2.32 The ODE shall calculate aggregate data within a configurable pre-defined time window (aggregation window).

5.2.2.33 The ODE shall disseminate the aggregate data to the requesting applications at the end of the aggregation window.

5.2.2.34 The ODE shall calculate average vehicle speed for all configured road segments, with a frequency of every 5 minutes.

5.2.2.35 The ODE shall not use data that was marked as invalid during validation as input for aggregation.

#### *Other Processing Requirements*

5.2.2.36 The ODE shall perform validation for all data sources where applicable.

5.2.2.37 The ODE shall perform sanitization for all data sources where applicable.

5.2.2.38 The ODE shall perform aggregation as required by the ODE configuration.

### ***Analytics Requirements***

#### *Measuring Performance*

5.2.2.39 The ODE shall monitor the volume of data ingested over time.

5.2.2.40 The ODE may monitor the volume of data distributed to each consumer application over time.

5.2.2.41 The ODE shall measure and monitor the amount of delay introduced by its processing (i.e., the time elapsed between ingestion and dissemination).

5.2.2.42 The ODE shall capture all errors related to its operation, including but not limited to data ingestion errors, data processing errors, and data dissemination errors.

### 5.2.3. Interacting with the ODE

To continue describing the operation of the ODE, this section details a typical scenario of a client application's interaction with the ODE, requesting data for its operations.

To begin this interaction, the client application must have registered with the ODE by submitting a registration request. During this registration process, the ODE records, stores, and subsequently manages the registration data, for current and future support. Registration data includes elements such as application/user identification and certification and credentials, possibly including credentials for obtaining unsanitized data if requested.

Upon satisfying the necessary prerequisites to acquire data from the various data sources, the ODE ingests data streams in real time and simultaneously performs any necessary conversions. In addition to performing the conversions, the ODE also performs the necessary data valuation, aggregation, and integration to support the data request.

After the application's data request is decoded and the ODE appropriately sources the data, possibly from multiple sources, the ODE efficiently packages the data. Finally, the ODE transmits the packaged data to the requesting client application.

Throughout its various processes, the ODE maintains a log file, which is a collection of event logs from the various sub-processes and sub-functions. This log function is a vital debug capability that provides developers with the ability to observe the inner workings of the ODE, evaluate its performance, and identify opportunities to improve said performance.

## 5.3. Client Application Interface

The client application interface is the primary portal through which users interact with the ODE. This interaction is intended to be through an API, which is composed of a set of routines and protocols that enable client software applications to connect to the ODE. This connection enables client applications to configure data requests. Once a secure connection is established with the ODE and data requests are successfully submitted, the client applications ingest these data for their applications.

## 5.4. ODE Operational Framework

Since the ODE is fundamentally a data management tool, its various operational modes are best determined by understanding the general characteristics of the data that it processes. Mobility data managed by the ODE is fundamentally associated with movements in time and space, and its greatest economic value is therefore derived when used as a real-time[3] asset. Another aspect of ODE data is that loads, as identified by network traffic, closely match the utilization of the road network. Typical rush-hour traffic will result in higher connected vehicle loads, thereby creating daily peaks in network traffic and demands on the ODE for processing. Another aspect of mobility data worth mentioning here is that despite the rush-hour peaks, mobility is a 24x7 activity, and therefore the ODE must be available on a similar basis. This is especially true as connected vehicle data becomes increasingly adopted for use by various end users and dependence builds on this data. It is unacceptable to say that data is not available because of planned "system maintenance." Fortunately, advances over the last decade in information technology (IT) make this last requirement feasible to meet.

Based on the above data characteristics, the following list identifies the necessary properties of all scalable "internet capable systems," which are important concepts to understand when reviewing the solution architecture.

- **Scalability** – The need to scale results from increased loads that arise from more processing demand. There are several types of scalability, all of which can be best understood from an evolutionary IT perspective. Older computing architectures ensured scalability via the acquisition of hardware whose capacity was deliberately greater than current demand and was therefore scalable by tapping into the excess capacity. ***Vertical scaling*** is perceived as an ineffective use of resources because it requires large capital outlays, especially as the marginal costs of acquiring this idle capacity tends to increase significantly with larger capacity servers due to increasingly exotic technologies at the margin. These relatively inflexible computing architectures require high levels of effort (LOEs) to develop, deploy, and maintain. ***Virtualization technology*** can be perceived as a way to better utilize the resulting oversized hardware resources by allowing many applications to simultaneously execute on the same server with a heightened level of isolation from each other to increase reliability.

  Around 15 years ago, when virtualization allowed many processes to share a single server, data volumes from internet startups spiked at startling rates, resulting in the Big Data problem. The principal factors driving Big Data were log file monitoring and clickstream analysis used to determine user behavior and web crawlers that provided web indexing and searching. Single machine architectures were quickly overwhelmed by these workloads as data I/O (input/output) became the sudden bottleneck. ***Horizontal scaling*** solved this latter bottleneck by a "divide and conquer" approach to the compute tasks at hand. Large numbers (i.e., hundreds or even thousands) of low-cost commodity servers were coordinated, with the aid of a "master" server, to perform parts of the same task and return their partial results to a single machine. This master, which received relatively low volumes of interim calculations, could then summarize the results of all of the individual efforts carried out at each server. This computational technique became embodied in Map-Reduce. Map-Reduce, which

[3] In this context, we will define real-time asset as data that has been collected, transmitted, and transformed, with minimal difference in time from when the data was collected to when it provides information that improves the utilization of the transportation network.

leverages the Hadoop Distributed File System (HDFS), also provided the critical benefit of being fault tolerant, or the ability to continue to perform despite the failure of hardware and/or software. This is achieved by replicating the data to at least three servers for every compute node.

The last scalability technique to mention is ***elastic scaling***. This provides a way to use automated methods to dynamically provision computing capacity as demand approaches preset maximum capacity utilization thresholds in order to expand computing capacity and, equally important, to decrease the provisioned computing capacity as demand wanes. Elastic scalability is available in relatively advanced and large-scale compute clusters/data centers where various consumers at any point in time will scale up and down compute resources to provide good hardware resource utilization. If elastic scalability is deployed in a compute cluster used by a single entity, it provides minimal economic benefits in much the same way vertical scalability is achieved with considerable economic inefficiency. For instance, excessive capital investment in hardware, deliberately targeted to meet the occasional peak demand, is an over-investment of compute resources at non-peak hours.

- The scalability considerations above, which are available to users of the more advanced Infrastructure as a Service (IaaS) vendors, also achieve another important benefit—namely, the option to deploy various ***multi-tenancy based data sharing*** approaches. These result in significantly ***reduced latency*** whenever multiple consumers of data within the cloud reside in the same data center. Often, the data can flow between two computer racks that are in the same room. The reduced latency can also achieve cost savings since the data stays within a single vendor's cloud, thereby reducing or even avoiding altogether data transmission costs. Another important aspect is that multi-tenancy in a publicly accessible cloud creates a perception of a greater level of transparency in terms of data governance.
- Lastly, ***load balancing*** provides a "traffic management" capability, which can monitor the resource utilization across several different providers in a cluster to route new computing requests to the lowest utilized members of the cluster, which is a group of machines or nodes that are configured to cooperatively process incoming traffic. Load balancers can also serve to implement fail-over, which occurs whenever specific compute resources become unavailable and computing flow is re-directed to other similar computing resources.

## 5.5. End-User Classes

End-user classes identify categories or groups of users that interact with the ODE. It is anticipated that each user class will interact with the ODE differently, depending on the functions and goals of the user class. However, given the primary role of a data provider, all end-user classes will in fact interact with the ODE in a very similar fashion, irrespective of the goal(s) of the user.

Each user will interact with the ODE via a secure subscription, requesting a series of data elements to support the task at hand. The fundamental difference among subscriptions is the specifics, such as the geographic expanse and the data type requested. Due to the lack of distinguishable user-based ODE functions, it is challenging to define end-user classes, which are largely based on functional interactions with a system. Thus, end user profiles will communicate how different users may interact with the ODE, further demonstrating the operation of the ODE.

Three end-user profiles have been developed to demonstrate how different users interact with the ODE:

1. Passive user profile
2. Active user profile
3. Hyper-active user profile.

In short, a user that fits the **passive user** profile minimally interacts with the ODE and the data that is being transmitted per that user's request. On the other hand, a user that is more aligned with the **active user** profile interacts with the ODE more frequently than a passive user. This active user also interacts more with the data from the ODE. This interaction may take the form of additional data processing, which typically involves transforming the data to support the user's task. These interactions tend to be more systematic, repeatable, and less ad hoc than how a hyper-active user would interact with data from the ODE. The **hyper-active** user profile fits a user that not only interacts with the ODE frequently but also interacts with it in an ad-hoc fashion. This user does a significant amount of post processing of the data to achieve the task's objectives. It is also not uncommon for this user to have multiple, simultaneous subscriptions to the ODE, with data requests differing from one subscription to the next. The differences in data requests may range from requesting the same data elements at different levels of granularity to requesting a completely different set of data elements.

Chapter 6. User-Oriented Operational Scenarios provides additional details on each user profile. The chapter describes a few situations to demonstrate how a user aligned to each of these user profiles interacts with the ODE to achieve a task.

As the USDOT's Connected Vehicle program evolves, so too will the ODE and its users. It is anticipated that once connected vehicle technology and the ODE are more mature, clearer ODE user classes will present themselves. These user classes will fit in some of the more traditional user classes of such systems. One of the more traditional sets of user classes that may appropriately classify the users of the ODE is the Private-Public-Hybrid (Public-Private) set. In the case of the ODE, these user classes are defined as follows:

1. **Private Users** – These are groups of companies/individuals that would use the subscription to the ODE to obtain data that would typically be used in a mobile or stationary application to drive the bottom line of a business
2. **Public Sector Users** – An example of this user class is the department of transportation in any state where connected vehicle data are generated. Other entities that may belong to this user class include other local governments, counties, and cities that own and maintain their TMCs. These entities will use the data to support the management, operation, and maintenance of their transportation facilities.

3. **Hybrid Users (Public-Private)** – In the case of an infrastructure management team, this group might, for example, want to measure the relative speed variances between high-occupancy toll (HOT) lanes and non-HOT lanes to create more adaptive pricing models—to better manage the facility.

As user classes become more distinguishable, the ODE development team is aware that additional functionalities will be needed to better serve its diverse users. One such functionality is a tiered data provisioning capability that limits the amount of data that a user group can receive from the ODE. For example, there could be particular data elements that are more sensitive than others (in terms of personal information and/or commercially sensitive data), which may only be accessed by a vetted set of users.

Irrespective of the ultimate set of end user classes, members of these classes will still be aligned to one of the aforementioned user profiles.

## 5.6. Operational Policies and Constraints

Operational policies and constraints that may affect the development and testing of the ODE include:

- Privacy and security concerns will need to be addressed to protect participants and to establish safe and secure data connections and client subscriptions.
- The availability of the Southeast Michigan Test Bed (given prescheduled demonstrations and activities) and its operators, which include test vehicle drivers, may impact the development and operation activities of the ODE.
- The communication and messaging standards are not yet stabilized. The ODE development team will aim to implement measures to keep abreast of these evolving standards and take the necessary development actions to allow the ODE to ingest any changes.

# CHAPTER 6. USER-ORIENTED OPERATIONAL SCENARIOS

The description of user profiles presented in Chapter 5.6 served as a guide in developing the following user-oriented operational scenarios. The goal of these scenarios is to communicate how users are envisioned to interact with the ODE. Presenting some of the specificities of the interaction with the ODE should provide a greater understanding of the ODE's operation and use to support transportation demand, operation, maintenance, and management.

## 6.1. Passive User Profile: Real-Time Data Visualization

A passive user profile fits a user that minimally interacts with the ODE. This user will subscribe to the ODE and request a set of data elements that undergo very little to no processing or transformation to achieve the user's mission. In such a case, the user may simply have a web-based dashboard to visualize the streaming data.

A passive user may be an individual or team at a TMC that wants to visualize vehicle location on a map, in real time, say as a means to monitor the system. To accomplish this, the user will place a subscription to the ODE. This subscription will include a set of desired data elements that the user will use to accomplish the task at hand. The requested data elements may include vehicle ID, vehicle position (latitude and longitude), vehicle speed, and bounding area coordinates. The ODE will verify/authenticate the subscription and source and stream the requested data via the user's port.

Upon receiving this data stream, the user will then use the stream as input to a geospatial tool to visualize the raw, unprocessed, un-transformed data. This may take the form of a map of the area, with a refresh rate of 1 second, on which dots represent vehicles traversing the transportation network of interest. The dots' color may vary according to the speed of the vehicle. This display requires some transformation, albeit minimal.

## 6.2. Active User Profile: Watch and Dispatch

A user that fits the active user profile is one that, as the name suggests, actively engages with the ODE. Actively engaging with the ODE may take the form of resubmitting subscriptions at regular or irregular intervals or redefining subscriptions depending on the user's goals and the application requirements to achieve them. Another defining characteristic of the active user profile is actively engaging with the data. This characteristic is meant to capture users who request data from multiple sources that the ODE polls as well as users who conduct additional data processing and transformation to get the data in a form suitable for their applications.

Additional data processing and transformation may include the application of a myriad of statistical packages and machine learning algorithms to support mission-critical goals.

As an example of an active user, consider a company that has been contracted to rid a transportation network region of roadside distractions, which reduce roadway capacity, during the morning and evening rush hours. These roadside distractions range from accidents and vehicle breakdowns to obtrusive debris. To effectively rid the network of these distractions, the company relies not only on real-time traffic information and advisories but also on a series of output from its algorithms to aid in determining the likely locations for these distractions, the type of distraction, and the best route to these locations for the removal of these distractions. Toward this end, this company will have a recurring subscription to the ODE requesting data between 5:00 AM and 9:00 AM and between 3:00 PM and 7:00 PM for weekdays. The data requested within these subscriptions remains constant. Requested data elements range from raw vehicle kinematic data and a series of aggregated roadway performance measures to travel advisories and weather data.

After the ODE fulfills these subscriptions, the requesting team will input the data into its various procedures to best identify/predict the location, potential cause, and type of these distractions and subsequently provide real-time route guidance to field units to remove these distractions.

## 6.3. Hyper-Active User Profile: The Explorer

The hyper-active user profile aims to capture how the (transportation) research community interacts with the ODE. The essential characteristic of the hyper-active user is variation in interactions—both in subscription frequency and details. The drivers of the variance in subscription are changing user goals and mission. At times, this user group may use data from the ODE to perform macroscopic analyses of the study area; at other times, this group may conduct more microscopic analyses to evaluate the minute details of a given phenomenon. For instance, a research team attempts to study traffic behavior before and after crashes. For such an investigation, it is important to: 1) determine whether there is anything to be uncovered from driver behavior before an accident that could predict the occurrence of an accident, and 2) determine whether there are traffic flow/driver behavior based measures that may be applied to return the traffic flow to a state of normalcy with greater efficiency than what currently exists.

There are many different ways a research team may engage the ODE to help tackle this challenge. One such way is through multiple, parallel, and dynamic subscriptions to the ODE. Each subscription will request a different set of data elements with different levels of granularity. Of these multiple subscriptions, one will play the role of main subscription that will continuously ingest data from the ODE to monitor the entire expanse of the traffic network. The subscription will largely subscribe to aggregated network performance measures, such as average speeds on the various links. The main subscription will also include other contextual data such as local weather and road weather to communicate the conditions under which data are being collected. The main usage of the data that is being streamed via this subscription is to efficiently monitor the entire network without processing and storing a lot of data.

There are two secondary subscriptions that support this main subscription. These subscriptions request more granular vehicle kinematic data, network operation data (e.g., traffic signal states), and weather data. They are responsible for collecting data for analysis to: 1) learn whether pre-crash data can help predict the occurrence of a crash and 2) determine what measures may be implemented to return traffic flow back to a state of normalcy in the shortest amount of time. Once the data from the main subscription indicates an abnormal change in performance, the tool that ingest this data triggers the storage of the last 30 minutes of data obtained from one of the secondary subscriptions. Simultaneously, the other secondary subscription will be triggered to begin its subscription to the ODE to collect detailed microscopic data, confined to the general expanse of the area where the drop in performance occurred.

In this scenario, the interplay of the subscriptions to the ODE is an example of a hyper-active user, as the user dynamically manages multiple subscriptions, requesting different data elements, with subscription for varying durations. Another component of this scenario that aligns with the hyper-active user profile is the various processing that the collected data element will undergo to support the mission at hand.

This page intentionally left blank

# CHAPTER 7. ANALYSIS OF THE PROPOSED SYSTEM

This chapter provides a high-level summary of some of the benefits that this ODE provides over the current state of the art. The chapter also provides a brief overview of opportunities to increase the offering and capabilities of the ODE.

## 7.1. Summary of Key Benefits

There are four key benefits of the ODE:

1. **Scalability and Transferability:** The ODE is being developed with features and components that are scalable and transferable. It is envisioned that the future of the ODE will include a number of different implementations in diverse environments. Some of these implementations will not only be different geographically, but will also be different in the transportation operations that they support. The ODE is being developed to be able to support the full spectrum of transportation network environments—from the rural network with hundreds of vehicles to the urban environment with hundreds of thousands of vehicles. At the core of the ODE's scalability and transferability features are its modular approach in developing each of its functions and its cloud-ready framework that will allow it to easily scale, both in terms of compute power and cloud storage.
2. **Single Repository for Disparate but Related Data:** One of the central tenets of the ODE is to provision data from disparate data sources. In doing so, the ODE becomes a single repository of data—including data not only on the performance of the transportation network, but also regarding the context under which these data were collected. While it is important to extract and derive transportation performance measures to understand how the network is performing, it is also important to collect and account for contextual data, such as weather and road weather, to evaluate the network and apply necessary measures to improve its operation. Operating and managing a system blind of contextual data may lead to the application of ineffective solutions, which may result in the system performing worse than before.
3. **Availability of Pre-Processed Data (to Support User Goals):** The ODE will perform two levels of data processing to support user goals: 1) source-based data processing and 2) user-driven data processing. Source-based data processing is the processing of source data before distribution to users or subscribing applications. This level of processing is reflected in the four key operational roles of the ODE—data valuation, data integration, data sanitization, and data aggregation. User-driven data processing refers to processes that were executed based on the details of a user's subscription. The resulting data from these processes are then

stored to be transmitted to other users upon their request for similar data. Both of these levels are meant to improve user efficiency when interacting with the ODE.

4. **Application-Oriented Platform:** In addition to using data provisioning as one of the capabilities that guides the development of the ODE, the interaction of the ODE with client applications also guides its development. From the perspective of application interactions, the ODE development team envisions various ways an application may interact with the ODE and is implementing features that support these interactions. A few of these features include:
   - An application interface that is used to request a data set to support an application mission(s)
   - Measures to reduce lag between data ingestion and data distribution in the event that the application requires real-time data to support its operation
   - The ability to derive the full picture of the operation of the transportation network to better communicate to its users the most appropriate actions and responses
   - The capability for an application to specify aggregation parameters such as roadway segments and time window, over which to aggregate data.

   Developing the ODE with applications in mind will better serve current and future applications—versus only having a system centered on provisioning connected vehicle data.

## 7.2. Opportunities to Extend ODE Service Offerings

The connected vehicle landscape is evolving and so are its supporting technologies and opportunities to leverage the data generated from this landscape. This evolution is also applicable to the ODE and its offerings. While the current ODE is envisioned to be a complete implementation, given the current infrastructure of the Southeast Michigan Test Bed, a few aspects of the ODE will not be able to undergo complete evaluation, and some of its service offerings will not be realized under the current deployment plan. The following list of opportunities further showcase the capabilities of the ODE in supporting connected vehicle applications:

1. **Increased Number of Vehicles:** Currently, there are nine equipped vehicles that are a part of the Southeast Michigan Test Bed. While these vehicles will supply critical data to the ODE, and subsequently the ODE's subscribing applications, this limited number of vehicles may not enable these applications to function as effectively as they were intended. The limited number of vehicles may also impact the level of accuracy of the ODE when reporting aggregated values such as key network performance metrics. This will be particularly true when these vehicles' behaviors are inconsistent with that of the rest of the traffic throughout the test bed.
2. **Emulation of Additional Applications:** To date, only a small number of applications will be emulated. While this is in part due to the availability of the algorithms of some of these

applications, the constraints of the physical network of the Southeast Michigan Test Bed also limit the selection of applications for emulation. To encourage additional application emulation, a simulated instance of the test bed may be developed to provide additional sensors and data to support the ODE’s interaction with a greater number of applications.

3. **Inclusion of an Integrated Microscopic Traffic Simulation Platform:** The ODE is capable of supporting connected vehicle technologies from both the data brokerage and the connected vehicle application perspectives. Under the current implementation plan, the Southeast Michigan Test Bed is limited in its ability to provide a robust volume of data as well as to facilitate the incorporation of a set of sufficiently diverse connected vehicle applications.
4. **More Diverse Set of Data Sources:** While the ODE does subscribe to a number of different data sources to support connected vehicle applications, a few other data sources are able to add value to some of these applications. Some of these additional data sources include mobile devices, social media, and other transportation applications used to support travelers today.

U.S. Department of Transportation
Federal Highway Administration
1200 New Jersey Avenue, SE
Washington, DC 20590

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

FHWA-JPO-15-227

U.S. Department of Transportation