RITA Intelligent Transportation Systems Joint Program Office

Concept of Operations: Data Capture and Management Research Data Exchange

www.its.dot.gov/index.htm

DRAFT Research Data Exchange Concept of Operations – August 2011
[publication number]

U.S. Department of Transportation
Research and Innovative Technology Administration

Produced by
ITS Joint Program Office
Research and Innovative Technology Administration
U.S. Department of Transportation

Notice

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE (DD MM YYYY)	2. REPORT TYPE	3. DATES COVERED
8 1 2011	DRAFT	Month YYYY – Month YYYY

4. TITLE AND SUBTITLE
Concept of Operations: Data Capture and Management Research Data Exchange

5a. CONTRACT NUMBER
DTFH61-06-D-00002

5b. GRANT NUMBER
Xxxx

5c. PROGRAM ELEMENT NUMBER
Xxxx

5d. PROJECT NUMBER
Xxxx

5e. TASK NUMBER
MI02047

5f. WORK UNIT NUMBER
Xxxx

6. AUTHOR(S)
Mike McGurrin
Anna Corrigan
James Larkin
Karl Wunderlich

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Noblis 600 Maryland Avenue, SW, Suite 755, Washington, DC 20024

8. PERFORMING ORGANIZATION REPORT NUMBER
Xxxx

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Department of Transportation
Research and Innovative Technology Administration
ITS Joint Program Office
1200 New Jersey Avenue, S.E.
Washington, DC 20590

10. SPONSORING/MONITOR'S ACRONYM(S)
U.S. DOT RITA JPO

11. SPONSORING/MONITOR'S REPORT NUMBER(S)
Xxxx

12a. DISTRIBUTION/AVAILABILITY STATEMENT
This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

12b. DISTRIBUTION CODE
Xxxx

13. SUPPLEMENTARY NOTES
Xxxx

14. ABSTRACT (Maximum 200 words)
Xxxx

15. SUBJECT TERMS
Xxxx

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE	Xxxx	Xxxx	Xxxx
Xxxx	Xxxx	Xxxx			19b. TELEPHONE NUMBER: (xxx) xxx-xxxx

Executive Summary

The USDOT's ITS Joint Program Office is researching the use of mobile data communications in surface transportation to improve safety, mobility, and the environment. The effort includes multiple research projects, field tests, and demonstrations. In addition, many other organizations are also conducting research and tests in the same field. Data sets are being collected and developed in all of these programs, but at the present there is no single resource that collects and provides information about these data sets or provides access to them. The *Data Capture and Management Research Data Exchange,* hereinafter referred to as the "Research Data Exchange" or "RDE," will provide such a resource.

The stakeholders in this project include:

- The Intelligent Transportation Systems Joint Program Office of the USDOT, which is the sponsor of this program;
- The Research and Innovative Technology Administration (RITA);
- The Federal Highway Administration (FHWA);
- The Bureau of Transportation Statistics (BTS);
- The Federal Transit Administration (FTA);
- The Federal Motor Carrier Safety Administration (FMCSA);
- The National Highway Traffic Safety Administration (NHTSA);
- State and local government agencies;
- Automobile manufacturers;
- Equipment manufacturers;
- Universities;
- Transit agencies;
- Commercial vehicle operators;
- Connected vehicle information providers;
- Connected vehicle application developers.

The Research Data Exchange is a core element of the USDOT's Data Capture and Management Program.[1] The Research Data Exchange is intended to support research, analysis, application development, and testing. As a research tool, it does not directly support operations of traffic management or other operational systems. However the concepts and lessons learned from the project will be useful in developing operational systems exploiting concepts of mobile communications and connected vehicles.

The Research Data Exchange will not be a single, centralized repository, but rather a *system of systems* linking multiple data management systems, some of which will be maintained and controlled

[1] *Connected Vehicle Research,* http://www.its.dot.gov/connected_vehicle/connected_vehicle.htm, accessed March 22, 2011.

outside of the USDOT, through a common web-based Data Portal[2]. Figure ES-1 shows the Research Data Exchange from the user's perspective. The user need not be concerned with the physical system wherein the data resides. Instead, to the maximum extent practical, the Data Portal will present the user with a single, integrated view of the information contained within the Research Data Exchange, regardless of where the data is housed or where a real-time feed originates. The user can access individual data sets and/or real-time data feeds, or can access pre-defined Data Environments that have been prepared. A Data Environment is assembled from multiple data sets and optionally real-time data feeds in order to serve a particular research purpose.

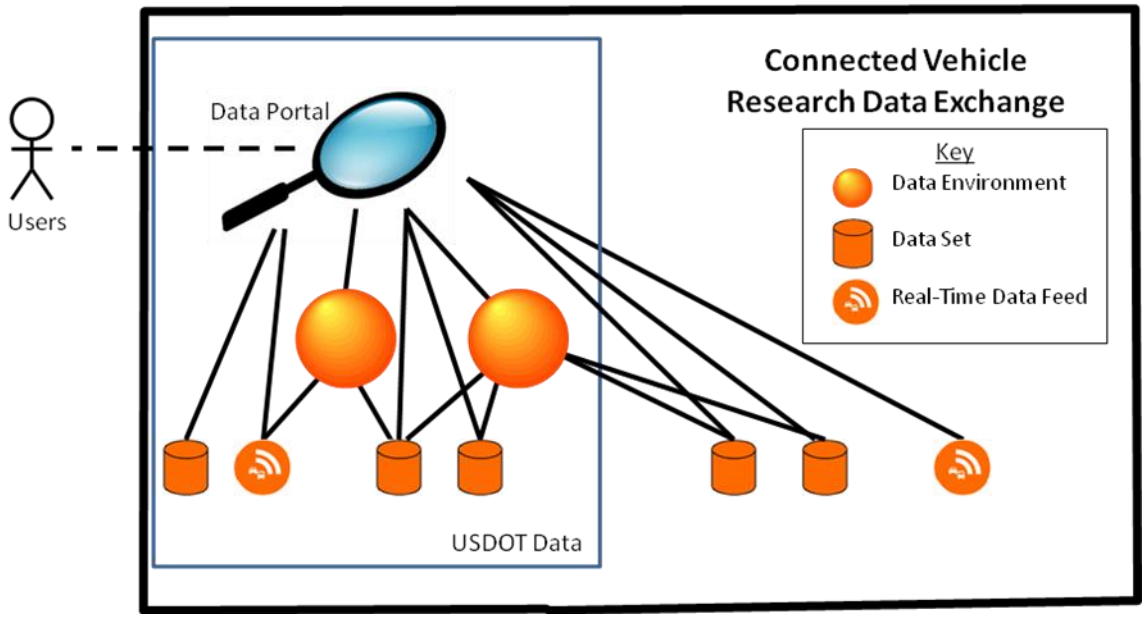

Figure ES-1: View of the Connected Vehicle Research Research Data Exchange

Much of the data will be available to any interested party, with little or no restrictions. Access to some data sets and data feeds will require registration and possibly have additional restrictions depending upon the needs of the data owners or the nature of the data. No personally identifiable information will be stored in the federal portion of the Data Exchange.

The purpose of this Concept of Operations is to communicate an understanding of user needs and to describe how the system will operate to fulfill those needs. This preliminary version is considered a living document that will be updated as new user needs are discovered and operational concepts are refined.

The intended audience for this document includes:

[2] This Data Portal, which is a component of the Connected Vehicle Research Data Exchange, is not to be confused with the Open Source Portal which is an entirely separate system that is being developed by the Data Capture and Management Program. The term "portal" in this document refers to the Data Portal.

- System developers who will create and support the Research Data Exchange based on the user needs and system concepts described in this document;
- Connected vehicle mobility stakeholders to determine whether their needs and desires have been adequately captured for use by the Research Data Exchange developer(s);
- Analysts, researchers, and mobility application developers requiring access to a research data for application development.

Table of Contents

List of Tables

List of Figures

1. Scope

1.1 Identification

This Concept of Operations (ConOps) describes the system characteristics of the *Data Capture and Management Research Data Exchange*, hereinafter referred to as the "Research Data Exchange" or RDE.

1.2 Document Overview

The purpose of this ConOps is to communicate an understanding of user needs and to describe how the system will operate to fulfill those needs. This preliminary version is considered a living document that will be updated as new user needs are discovered and operational concepts are refined. Development of this ConOps follows the IEEE Guide for Concept of Operations.[3] The intended audience for this document includes:

1. System developers who will create and support the Research Data Exchange based on the user needs and system concepts described in this document
2. USDOT mobility program stakeholders to determine whether their needs and desires have been adequately captured for use by the Research Data Exchange developer(s).
3. Analysts, researchers, and mobility application developers requiring access to research data for analysis and application development

The major sections of this documents are: (1) a description of the current situation, (2) justification for changes, (3) operational concepts for the proposed Data Capture and Management Research Data Exchange (the "Research Data Exchange") that includes both the activities of people and supporting technology, (4) idealized operational scenarios, (5) a summary of the impacts of the proposed system, and (6) an analysis of the improvements and limitations of the proposed system.

1.3 System Overview

The Research Data Exchange is a core element of the Data Capture and Management Program, which is part of a USDOT research effort considering mobile data communications in surface transportation to improve safety, mobility, and the environment.[4] This program is sponsored by the Intelligent Transportation Systems Joint Program Office of the USDOT. Key USDOT agencies that have strong interests in this project include the Research and Innovative Technology Administration

---

[3] IEEE Std. 1362-1998, IEEE Guide for Information Technology—System Definition—Concept of Operations (ConOps) Document.

[4] *Connected Vehicle Research,* http://www.its.dot.gov/connected_vehicle/connected_vehicle.htm, accessed March 22, 2011.

(RITA), Federal Highway Administration (FHWA), Burueau of Transportation Statistics (BTS), Federal Transit Administration (FTA), Federal Motor Carrier Safety Administration (FMCSA), and the National Highway Traffic Safety Administration (NHTSA). State and local government agencies, automobile manufacturers, equipment manufacturers, universities, transit agencies, commercial vehicle operators, and connected vehicle information providers and application developers are also important stakeholders.

The objective of the system is to share archived and real-time multi-source and multi-modal data with the transportation community. The Research Data Exchange is intended to support research, analysis, application development, and testing for the Dynamic Mobility Applications and Applications for the Environment Real-Time Information Synthesis (AERIS) programs, as well as other related research. As a research tool, it does not directly support operations of traffic management or other control systems; however the concepts and lessons learned from deploying and utilizing the RDE will be useful in developing operational connected vehicle systems that exploit real-time, multi-source data.

Two documents strongly influenced the development of this Concept of Operations. The *IntelliDrive Data Capture and Management Program Vision: Objectives, Core Concepts and Projected Outcome* document discusses the Data Capture and Management Program objectives, core concepts and high-level program activity plan. The *IntelliDrive Data Capture and Management Program: Transforming Federal Role*[5] document addresses stakeholder needs for the program and what role the Federal Government will play.

Tthe Research Data Exchange is being developed as a transportation data-sharing system that promotes sharing of both archived and real-time data from multiple sources (including vehicle probes) and multiple modes. This new data-sharing capability will better support the needs of ITS researchers and developers, reducing costs while encouraging innovation.

Section 1.3.2 provides a more detailed description of the system, while Section 1.3.1 defines some of the terms used in describing the system.

1.3.1 Definition of Terms

Data Set: A *data set* is a collection of related data, organized into a regular and consistent format. As used in this document, a *data set* also includes the associated metadata and other documentation about the data.

Data Environment: A *Data Environment* consists of *data sets* and/or *real-time data feeds*. A *Data Environment* is:

- a well-organized collection of data of specific type and quality;
- captured and stored at regular intervals from one or more sources;
- systematically shared in support of one or more applications; and
- designed to promote research and decision making.

[5] "IntelliDrive Data Capture and Management Program: Transforming the Federal Role," United States Department of Transportation, Washington, DC, April 2010.

A single Data Environment may include data sets that physically reside in one or several different *Data Management Systems. Data Environments* are a new concept to the surface transportation community, developed for the Mobility Data Capture Program which captures the notion of a the logical collection of data compiled and organized to support research and decision-making regardless of where data elements originate or are stored

Data Management System: A physical system that stores archived data (data sets), and/or *real-time data feeds, and/or Data Environments*. The Research Data Exchange physically consists of interconnected *Data Management Systems* and the *Data Portal.* Data Management Systems is a more general term than Data Warehouse, and in this document, the term Data Management System is used in reference to component systems of the Research Data Exchange.

Data Portal: A component of the Research Data Exchange. It provides a common, web-based interface into the various *Data Management Systems* in the Research Data Exchange. It will also house and provide the user with additional information about the various *data sets, Data Environments*, *and real-time data feeds* that the Research Data Exchange can provide.

This Data Portal is not to be confused with the Open Source Portal which is an entirely separate system that is being developed by the Dynamic Mobility Applications Program.

Data Warehouse: a repository of electronically stored data. It houses data that originated from either another application or an external system. In this document, the term Data Warehouse will always refer to existing systems, not to the planned Research Data Exchange or its component systems. Instead, the more general term Data Management System is used.

Real-time Data Feed: A mechanism for receiving updated data as it is made available. The data may be updated periodically or the updates may be event-driven. The mechanisms to access a real-time data feed can vary, including information that is *subscribed to* and pushed out automatically to subscribers or information can be *pulled*, with the sender transmitting it upon receiving a request from the recipient. The Data Portal is the component of the Research Data Exchange that supports access to real-time data feeds.

System of Systems: A *system of systems* consists of component systems that each serve independent purposes and can accomplish these purposes even if detached from the larger system of systems. However, through collaboration, these systems can serve additional purposes and meet additional needs. [6]

Federated System of Systems: A *federated system of systems* is a *system of systems* characterized by the autonomy of the collaborating partners, heterogeneity, and geographic distribution.[4]

Data Exchange: As used in this document, a *Data Exchange* is a system designed to provide access to a variety of data that is hosted and managed in one or more locations by one or more entities. A *Data Exchange* encompasses many of the attributes and features described in the above definitions.

[4] This definition is based upon "System of Systems: Architecture based Systems Design and Integration," IEEE SMC 2005 keynote speech, Dr. Andrew P. Sage.

Specifically, the Research Data Exchange is a *Data Exchange* designed to support the needs of the US DOT ITS Research Program.

1.3.2 System Description

The Research Data Exchange is a new, decentralized, Internet-based, federated system of systems. Physically it consists of a number of linked (federated) Data Management Systems plus the Data Portal that serves as a common entry point. From a user perspective, however, the Research Data Exchange contains Data Environments (collections of related data sets and possibly real-time data feeds), individual data sets, and real-time data feeds. A Data Environment may contain data sets that are physically stored across multiple Data Management Systems. See Section 1.3.2 for the definitions of key terms used in this document, including a definition of a Data Environment.

Figure 2 shows the Research Data Exchange from the user's perspective. The user need not be concerned with the physical system wherein the data resides. Instead, to the maximum extent practical, the portal will present the user with a single, integrated view of the information contained within the Research Data Exchange, regardless of where the data is housed or where a real-time feed originates. The user can access individual data sets and/or real-time data feeds, or can access Data Environments that have assembled multiple data sets and optionally real-time data feeds to serve particular research or application development and testing purposes.

The Data Environment concept was conceived to address the USDOT research needs on how mobile data communications in surface transportation could improve safety, mobility, and the environment. While it is possible that Data Management Systems external to the USDOT might also have Data Environments, it is more likely that these will all be defined within the USDOT system. Although the data environment definitions and implementations will be contained with the USDOT's portion of the Research Data Exchange, an individual Data Environment may draw data from multiple sources (i.e., from multiple Data Management Systems). This is shown in Figure 2.

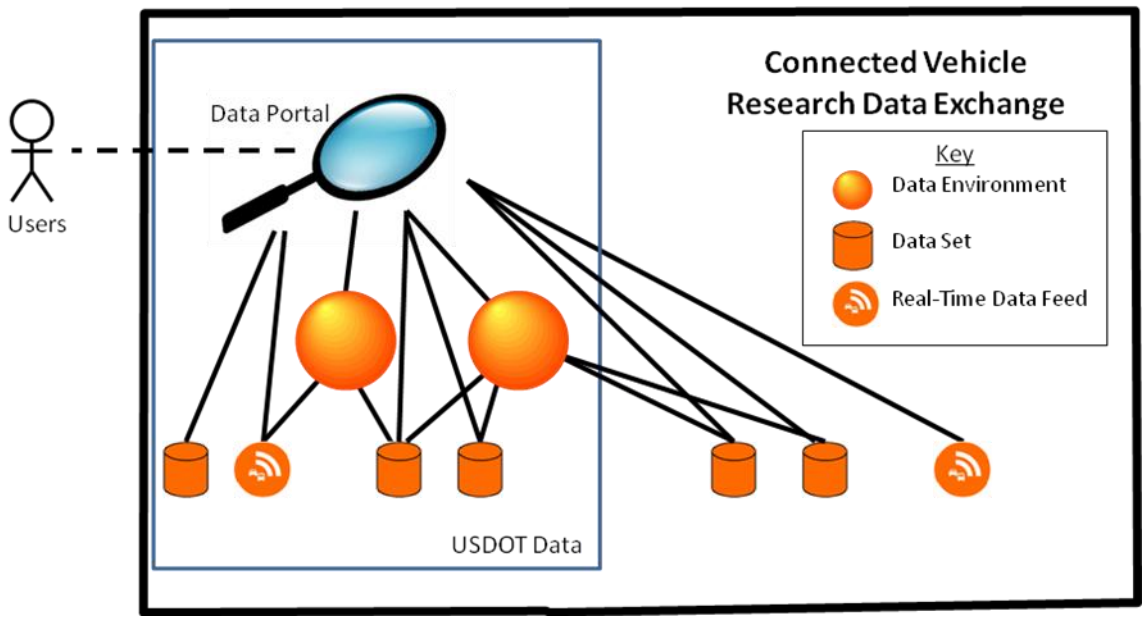

Figure 2: User View of the Connected Vehicle Research Research Data Exchange

Figure 3 shows a conceptual depiction of the physical system. The Research Data Exchange consists of multiple, federated Data Management Systems, all reachable through a common Data Portal. It is anticipated that one or more of the Data Management Systems will be operated and managed by the Data Capture and Management Program, while others will be operated and managed by other federal and non-federal stakeholders. The Data Management Systems that are not operated by the USDOT are shown as residing partially within and partially outside the system boundary for the Research Data Exchange. This is deliberate. Typically, these systems serve additional purposes independent of their role in the Research Data Exchange, and may also have additional external interfaces that are outside the scope of the Research Data Exchange.

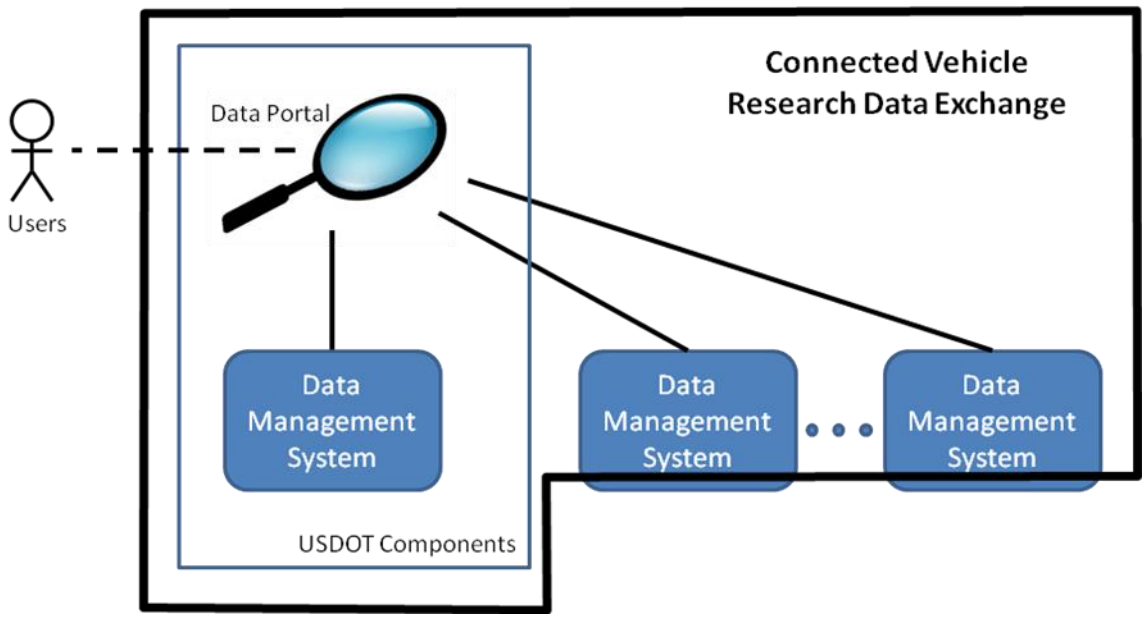

Figure 3: Physical View of the Research Data Exchange

While each Data Management System is an individual entity, the Data Portal is a single website that maintains information and metadata on all the participating data. The Data Portal serves four general functions:

1. The Data Portal serves as a central location for storage and access to summary metadata information about each individual data set, Data Environment, and real-time data feed found in the Data Capture and Management Research Data Exchange, as well as general information about the Data Capture and Management Program as a whole.
2. The Data Portal acts as a single location for users to access all Data Management Systems. This does not preclude some Data Management Systems from also providing a direct access path to their own (local) data.
3. The Data Portal serves as a single location where users (both human and automated) can access the information (presented as a collection of Data Environments) drawn from any of the participating Data Management Systems (provided the user has the appropriate access permissions). To the maximum extent feasible, the portal will provide a common look and feel, regardless of which Data Management Systems are being accessed.
4. As a convenience, the Data Portal may also serve as a single location where users can access and download utility software that can be used to manipulate, analyze, or visualize data from one or more data sets, data environments, or real-time data feeds. However, this is not a primary purpose of this system.

2. Referenced Documents

1. *IntelliDrive Data Capture and Management Program Vision: Objectives, Core Concepts and Projected Outcomes*, Version 1.5, US DOT, April 2010.
2. K. Wunderlich & C. Burnier, *Data Capture and Management Program: Transforming the Federal Role,* Version 5.0. Noblis, Washington, DC, May 2010.
3. K. Wunderlich & C. Burnier, *Data Capture and Management Program DRAFT: Appendix*, Version 2.0. Noblis, Washington, DC, Feb. 19, 2010.
4. *IntelliDrive Dynamic Mobility Applications Program Vision: Objectives, Core Concepts and Projected Outcomes*, Version 1.4, USDOT, April 2010.
5. *IEEE Std. 1362-1998. Guide for Information Technology--System Definition –Concept of Operations (ConOps) Document*.
6. A. Sage, *System of Systems: Architecture based Systems Design and Integration*, IEEE SMC 2005 keynote speech.

3. Current System or Situation

3.1 Background, Objectives, and Scope

The Research Data Exchange is a new stand-alone system that will not replace any of the current data warehousing systems. Since there are no current systems on which to model the connected vehicle Research Data Exchange, this section will provide background about the current transportation data warehousing efforts used by the United States transportation community. The objectives and scopes of the current numerous data warehouses have no bearing on this new system and will not be discussed in this document.

3.2 Operational Policies and Constraints

Because there is no current system, this section is intentionally left blank.

3.3 Description of the Current Situation

Current technologies and archiving warehouses already provide researchers and system operators with a sizeable number of data sets. Recent inventories of several dozen of the many transportation data warehousing efforts throughout the U.S. can be found in two recent reports.[7,8] The inventories include warehouse descriptions, data availability, and update frequency for each of the warehousing sites that were reviewed.

The studies found that most current data warehouses can be characterized as being highly dependent on passive infrastructure-based sensors. Furthermore, most of these data resources are single-source. For example, there may be separate databases for data provided by freeway loops, data provided by arterial occupancy sensors, and data provided by periodic "floating car" travel time runs. Traveler data are limited to behavioral data collected from small samples from infrequently conducted surveys.

Data captured and managed in the current data warehouses tend to be collected over time, assessed for quality and potentially aggregated at some intermediate point, and then at a later date (days, months or even years later) made available to researchers or other interested users. Each of the individual transportation archiving warehouse systems has its own objectives and scope based on its own needs and goals. This also holds true for the overarching operational policies or constraints for

[7] USDOT, *Data Capture Federal Role Appendix: Inventory of Data Warehousing Efforts*, Unpublished.
[8] A. Vandervalk, *Data Capture and Management, Needs and Gaps in the Operation and Coordination of U.S. DOT Data Capture and Management Programs,* Report No. FHWA-HOP-11-004, Cambridge Systematics, November 2010.

each data warehouse or group of data warehouses, whether it is run publicly through a state, local, or federal organization or by a private company or organization.

3.4 Modes of Operation for the Current System or Situation

Because there is no current system, this section is intentionally left blank.

3.5 User Classes and Involved Personnel

This section covers information about the current users of the various web-accessible transportation data warehouses used in the United States. Not all possible classes are included and the class definitions are general in order to encompassthe broad differences in users that make use of the thousand or more different transportation warehouses. Each warehouse has its own objectives and funding, and attracts different types of users.

The current users of the various data warehousing systems can be broken into three categories. The first category is the users of the system that actually interact with the system and use information it provides. The second class category is the owners of the system. These are the individuals and organizations that make overall decisions concerning the policies and plans of the data warehouse. The final class category is the managers of the data warehouse who maintain the system and run the day to day operations of the warehouse. Examples of the three categories are listed below.

- Users of data in the data warehouses include:
  - Applications developers who use the data to develop, test, or run software
  - Researchers from the public sector, universities, non-profit organizations and private consulting firms, who evaluate the data from the data warehouse
  - Decision makers who use the data to evaluate future transportation improvements and develop policy
  - Automated consumer-facing applications that access and utilize the data
- Owners of the data warehouses include:
  - Universities
  - Private organizations (such as consulting firms or trade associations, State, local, and Federal government organizations
- Managers of the data warehouses include:
  - University department support staff
  - Private organization technology support staff
  - In-house government technology support staff or outside support contractors

There are similarities between the future user classes for the proposed system and the current users of the various transportation data warehouses. Users of the current systems access data from the warehouses to develop and support applications, conduct research and support decision making, as will be the case with the proposed system. Also, some of the current data warehouses include detailed metadata, quality checking, and governance rules concerning access to the data, functions which will also be included in the proposed system.

The major differences between current systems and the proposed Research Data Exchange is that users of the proposed system will be able to access data from multiple systems from a single access point and can contribute data, metadata, and information back into system. This process will enhance the data and allow new users of the data to benefit from other users' work.

Another difference is that the new system will promote the open use of the data to encourage more users to access the data, where as in most of the current data warehouses data access is more limited and closed. Finally, while the current situation includes both research and operational data and data warehouses, this new system's purpose will be to support research (including prototype applications development), analysis, and policy decision-making. It does not support ITS operations. As such, it will supplement, not replace, existing systems.

3.6 Support Environment

Because there is no current system, this section is intentionally left blank.

4. Justification For and Nature of Changes

4.1 Justification for Changes

The Data Capture and Management Program plays a key role in supporting other ITS JPO initiatives identified in strategic plans, in the areas of safety, mobility, and environment. Many of these initiatives will require systematic capture and management of data over time to realize their objectives. The cross-cutting Data Capture and Management Program is chartered to coordinate across these initiatives to identify comprehensive data needs. Further, the Data Capture and Management Program is jointly responsible for designing laboratory experiments and field tests that meet these identified data needs in the most cost-effective way. In the past, only the results and findings of these activities were typically made available (typically through papers and technical reports). Because of a confluence of web-based technologies and asynchronous collaboration methods, it is increasingly easy to share base research data and tools, allowing for re-use and lowering the time and cost involved in future research efforts. The value of this open flow of research results has been eloquently stated in a 2009 report from the National Academy of Sciences[9]:

> *The advance of knowledge is based on the open flow of information. Only when a researcher shares data and results with other researchers can the accuracy of the data, analyses, and conclusions be verified. Different researchers apply their own perspectives to the same body of information, which reduces the bias inherent in individual perspectives. Unrestricted access to the data used to derive conclusions also builds public confidence in the processes and outcomes of research. Furthermore, scientific, engineering, and medical research is a cumulative process. New ideas build on earlier knowledge, so that the frontiers of human understanding continually move outward.*
>
> *Researchers use each other's data and conclusions to extend their own ideas, making the total effort much greater than the sum of the individual efforts. Openness speeds and strengthens the advance of human knowledge.*

The resulting well-documented, distributed, data resources will allow data captured from diverse sources to be integrated, shared, and leveraged by a broad range of researchers, private sector partners, and system operators. Program objectives, core concepts and a high-level program activity

[9] National Academy of Sciences, Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age," page 59, Washington, DC, National Academies Press, 2009.

plan are detailed in the *Data Capture and Management Program Vision Document* .Without a cross-cutting data capture program, the result will be *ad hoc* data capture and management by each application area. This traditional approach results in redundancies and increased expenses in data collection activities, supporting only limited applications by analysts working with individual data sets that target a single application.

The need for a reliable, easy-to-access, publicly available system where researchers, application developers, and decision makers can go to download quality multi-source and multi-modal transportation data or enhance existing data is critical to the transportation community as a whole. Below are two examples that provide justification for the new system.

4.1.1 Researcher Example

A researcher from a University Transportation Center is working on a research project focused on extending transit signal priority and real-time transit schedule adjustments to consider transit connection protection. The researcher would like to know if there are existing data sets that he could use to develop and test a new algorithm that uses wirelessly obtained transit passenger data to synthesize origin destination and transfer point information.

The researcher is new to the effort, and begins by searching the Research Data Exchange to see if anyone has already collected the types of data he needs for his research. He is able to find sets of data associated with three different arterial Data Environments with transit data incorporated into them (Detroit, Portland and Memphis test beds). He makes rapid progress understanding and using one of the data sets from Memphis because of the excellent metadata. At some point, however, the researcher thinks he has found inconsistencies between the transit vehicle data and the passenger data. He voluntarily registers on the Research Data Exchange, gaining access to the user forums. He reports this apparent discrepancy in a forum for the Memphis Data Environment and another user of the data points out to him that the discrepancy is not an error, but rather data from passengers waiting for the transit vehicle along the route.

While the Research Data Exchange enabled the researcher to jump-start his work, saving time and money, he still needed to collect new data on traffic flows in Memphis to help him analyze the impact that transit signal priority adjustments would have on other vehicles. Having experienced first-hand the value of the Research Data Exchange, he now cleans up the city's data documenting the correction, and later resubmits it as a new data set. Once approved, the new data set is added to the system. A few months later, he receives an alert from the forum letting him know that someone has posted a question on his data set. He logs in to find that another researcher investigating the use of Signal Phase and Timing data for environmental applications is combining his data with signal timing plan data, and needs some clarification on the definition of one of the parameters. He is able to answer the question quickly, and the new researcher expresses his thanks for the timely support.

As this simple example shows, there is a critical need for a single, easily accessible source of quality connected vehicle transportation data and a mechanism for researchers and others to collaborate on enhancing, understanding, and using the data. The future development of connected vehicle transportation data sets provides an active dynamic environment where members can collaborate about their research and studies.

4.1.2 Application Developer Example

In this example, we will examine an application developer who is funded by the USDOT to develop an application to provide dynamic route guidance for commercial vehicles. First we look at how the process would be conducted without the Research Data Exchange in place and then with the proposed system in place.

Without the use of the Research Data Exchange, the developer would first have to conduct research both on the current state of the art for dynamic route guidance for freight movement and on what data sets might be available to support the development, calibration, and validation of the application. This process could take several months. After finding the appropriate data, the developer would then need to contact the owners of the data to determine what information is available and what the data represents in order to determine if it is indeed applicable to his development effort. The developer would also need to arrange terms for obtaining the information.

Typically, upon the conclusion of the project, the new application will be available, but any data collected or refined for use in the project will be unavailable for use by other researchers and developers.

Now we will look at a situation with the Research Data Exchange in place. The application developer can first use the Data Portal to search for data sets or Data Environments that are relevant to his work. This information can greatly reduce research time depending on how much information is already available in the Research Data Exchange.

For example, suppose the application developer finds a Data Environment with raw GPS vehicle traces for a corridor of interest, another Data Environment that contains detailed weather information for the same area (for use in evaluating inclusion of weather in predictive travel time algorithms) and then goes to the Open Source Portal (a separate system from the Research Data Exchange) and finds an open source application that can process the GPS traces to that they can be used to emulate a stream of connected vehicle probe vehicle data. The combination of the open data and open source software allows the developer to jump start the development of his project.

Once the developer voluntarily registers on the Research Data Exchange portal, he or she can also collaborate with other members to see if they have any suggestions or recommendations for the data collection or application under development.

As stated above, this was a USDOT-funded research project. Per the terms and conditions of the contract, grant, or cooperative agreement, any new or refined data sets are made available under an open data sharing agreement. Thus, the developer benefited from prior work conducted in the area and will in turn contribute to future work in the same area. Depending upon the particulars of each case, the information might be delivered to the USDOT to be housed in a federally operated Data Management System or it may be housed in a Data Management System operated by the developer that meets the terms and conditions of the Research Data Exchange and is already or becomes part of the Research Data Exchange.

4.2 Description of Changes

This section outlines the changes based on the gap between the current state of practice and the envisioned system. The changes detailed in this section will be used to bridge the gap between the current and future systems, and will help to produce system requirements. The changes fall into four categories:

- Capability changes: Functions and features are added, modified, or deleted.
- Environment changes: Both changes in the operational environment that will result in changes to the operation of the system and changes in the operational environment that should take place due to changes in the system.
- Operational changes: Changes to the user's operational policies, procedures, methods, or work routines caused by the above changes.
- Support changes: Changes in the support requirements caused by changes in the system.

4.2.1 Capability Changes:

Capability needs are defined as functions and features to be added and modified in order for the new Research Data Exchange to meet its objectives and requirements. The Research Data Exchange needs the following capabilities:

C1. *Store archived transportation data:* The system needs to archive and store data for a given period of time directly within the sytem. Storing the archived transportation data will allow application developers to have access to the data over an extended period of time and use the archived data to test new applications. Archived data will alos permit historical analysis and performance benchmarking.

C2. Search transportation data: The system needs to have the ability to search the data sets by source, location, contributor, date, type of data, and type of data elements. The system needs to be able to provide a simple and easy-to-use search interface for members of the system to quickly find and sort data.

C3. *Store and display history/context information:* The system needs to store history/context information about each Data Environment and the data sets they contain. The system needs to display this information to all users of the system. This will ensure that users and potential users understand the terms, purpose, scope, and context under which the data was collected. This helps users and potential users evaluate the utility of the information for their intended purpose.

C4. *Support external applications accessing data from the Research Data Exchange:* External applications need automated mechanisms for accessing data from the Reseacrh Data Exchange. It is anticipated that many if not most data sets will be available without restrictions, however some of the participating data stores may have additional restrictions. For Data Environments or data sets with additional access restrictions, the Data Portal must be capable of passing along the appropriate authorization information between the external application and the Data Management System where the information resides. The individual Data Management System, not the Data Portal, is responsible for actually implementing any specific access controls.

C5. *Implement access controls for automated access:* Because of the potential for accidental or deliberate misuse of the system, impacting the systems availability for other users, some sort of registration and authorization will be required for all automated access. Once a user registers the application and it is approved, the external application will need to be authenticated each time it connects. The impact on the system of each automated application will be monitored to ensure that it does not adversely impact the operation of the system.

C6. *Accept real-time or near real-time data feeds from external sources and make those data feeds available to users:* To provide current information to external applications and system users, the system needs to collect real-time data from external systems at a certain rate and with a given maximum allowable lag time (provided that the external systems are operating correctly). The required "real-time" or "near-real-time" provision will need to be established based on the requirements of anticipated applications. The system needs to allow applications and users access to the real-time data feeds as well as store this real-time data as archived data for a period of time required by the various application needs.

C7. Store and share utility software, including data manipulation, analysis, and visualization tools: As a convenience, the system needs to be able to store and allow users to download utility software (which is expected to be primarily or exclusively in executable format). The system shall allow users to search, view descriptions, and download these executable data support tools; However this is not intended to duplicate the DMA Open Source Application Portal, which is a completely separate system.

C8. *Store, update/provide detailed metadata:* The system needs to store detailed metadata specifically describing the contents of the data sets, Data Environments, and real-time data feeds, as well as the utility software. The metadata needs to include information about from the source of the data (e.g. maps, GPS survey data)), what the data measures (e.g., vehicle speed, time of day), and the units of measure associated with each piece of data (e.g., mph, seconds since the beginning of the trip). For utility software, the metadata would include the purpose, required operating environment, inputs, and outputs.

4.2.2 Environment Changes:

There is currently no Research Data Exchange, so the changes from the current environment represent needs for a new system. Environment needs are defined as changes in the operational conditions that will cause changes in the system functions, processes, interfaces, or personnel and/or changes. The environment needs for a Research Data Exchange are:

E1. *Facilitate interactive communication between registered users of the Research Data Exchange:* Users of the Research Data Exchange need to communicate with each other about how they are using the information in the Research Data Exchange. The system will facilitate this collaboration, providing registered users of the system the ability to share insights, review related work, address common concerns, and work with other members. This interaction will keep registered users of the system informed of the happenings within the Research Data Exchange to prevent duplicate work and improve their own work.

E2. *Support the establishment and growth of an active community to build up and sustain the data sets and Data Environments in the Research Data Exchange:* For the exchange to be successful a community must form around it to help promote its use, error check the data, and contribute data and information back into the Research Data Exchange. Quality connected vehicle data stored in the system will help start the community-building process, but it is the active community participation that will sustain the system, potentially beyond the initial period of Federal support.

E3. *Provide a single Internet-based portal for access to all data resources and meta-data:* The Research Data Exchange requires an Internet-based single access point that can be used by multiple user types in multiple locations. This single point of access has the potential to link to multiple Data Management Systems together to allow users to access data from a broad range of research areas. The Internet portal must be easy to navigate, adhere to all government accessibility standards, and provide a central point for using any of the data sets, Data Environments, and/or real-time data feeds within the system. To the maximum extent feasible, the portal will provide a common look and feel for access with each Data Management System.

E4. *Organize and support the use of multi-source data*: Members of the Research Data Exchange community need to work together to solve institutional issues in collecting multisource and multimode data, which requires Federal, local, and state agencies, as well as research institutions such as universities, to collaborate closely in project synchronization, cost/data sharing, and/or any necessary agreements. The Research Data Exchange will provide a platform for sharing multi-source and multimodal data.

E5. *Deliver and support archived data and real-time data feeds*: Participants in the data capture and management program need to work with the rest of the connected vehicle team to enable data collection, especially in collecting real-time data. Collecting real-time data requires the development of automated interfaces and interface standards. This effort involves intense collaboration, in addition to understanding the technical issues, to standardize interfaces to facilitate automated data collection. Member input is also critical to determine how to archive the real-time data for various application needs.

E6. *Utilize (and where necessary, develop) and improve standards in data sharing*: Key stakeholders are responsible for supporting efforts to select, develop, and improve relevant standards for data sharing and interfaces. Standardization is crucial to make the Research Data Exchange viable and useful.

4.2.3 Operational Changes:

There is currently no Research Data Exchange, so the changes from the current environment represent needs for the new system. Operational needs are defined as operational policies, procedures, methods, or daily work routines. The operational needs for a Research Data Exchange are:

O1. *Support quality checking on data from all sources*: To ensure that data in the data sets and Data Environment are useful and meaningful to the users, the data must be checked for errors and inconsistencies. Any errors found in the data need to be flagged so users are aware that the specific piece of data may be inaccurate. This gives users and application

developers the option of using only processed quality controlled data or the raw data depending on their needs. This checking would be the responsibility of the managers or operators of each Data Management System, not the portal. The USDOT may establish common minimum guidelines for quality checking.

O2. *Establish and document governance for the Research Data Exchange:* The Research Data Exchange has to have both a clear set of rules for adding, archiving, or deleting both Data Management Systems as well as individual data sets and Data Environments contained in the federated system. These might include requirements for minimum levels of data quality and minimum required metadata information. In addition, rules for becoming registered users of the system and standard default access policies will be needed. These default policies will govern access to the USDOT's data sets, Data Environments and real-time data feeds within the system and other environments will be encouraged but not required to adopt them. These default rules must comply with all applicable federal laws, rules, and policies. However, generation of additional restrictions specific to the Research Data Exchange should be avoided to the maximum extent possible. All policies, procedures and rules must be easily accessible to all users of the system and must be agreed to by the users of the Research Data Exchange before gaining full access to the system.

O3. *Monitor and quantify Research Data Exchange utilization:* The system needs to be able to monitor key functions including adding/deleting/modifying of users, data sets, and Data Environments. Other critical functions are tracking site usage and system performance. By monitoring the data exchange functions, administrators will become aware of when they need to reanalyze and address support requirements of the system.

O4. *Enforce governance, address user needs, and maintain the system:* For the Research Data Exchange to be successful there must be a single entity that will help to enforce the governance rules to safeguard users' and contributors' data rights within the Research Data Exchange. This entity may also promote the Research Data Exchange and market it to potential users. This is anticipated to be one of the roles of the Data Capture and Management Program's "Data Manager."

4.2.4 Support Changes:

There is currently no Research Data Exchange, so the changes from the current environment represent needs for the new system. Support needs are defined as the system maintenance and support requirements caused by the system functions, processes, interfaces, or personnel. The support needs for a Research Data Exchange are:

S1. *Restore the system from failure:* To ensure the integrity of the system and safeguard the data stored on the system, there must be a way for the system to be restored after an internal or external failure occurs. All information must be backed up on a regular basis in order to minimize any losses of data, metadata, messages, member information, and content information. Full system functionality must be restored.

   a. The system is intended to support research, analysis, and testing, not for actual traffic management operations. The need is for a return to service within about 48 hours. The restoration requirements for individual Data Management Systems are under the control of each responsible party. However the Research Data Exchange

may or may not establish minimum standards concerning backup and restoration that would need to be met by participating Data Management Systems.

S2. *Facilitate and support registered users of the system in documenting their work with the data:* Registered users need a method to voluntarily document how they plan to use or are using the data they access from the Research Data Exchange. This documentation process will also allow members who are improving the data on the system to share their updates and improvements with the community.

S3. *Support and manage multiple access levels for users:* The Research Data Exchange must be able to differentiate users and to assign appropriate permission levels to each user. This requirement allows for different members to access only data they have been approved to view and download. It also allows designated members to have permissions to add, restructure, or delete content and administer the Research Data Exchange. The portal must handle this for portal-related access and services, and the portal must be able to differentiate users and support authentication, but permission levels for access to each Data Management System, beyond the overall default values, will be managed by the individual Data Management Systems.

S4. *Track usage and manage members of the Research Data Exchange:* The Research Data Exchange must be able to track usage in order to continually assess the value of the system, continually improve the system, and help protect against abuse of the system. Potential actions that may be monitored include: who registers with the Research Data Exchange and when, when members log in or log out, user status within the system, and when users submit information to the system. Each Data Management System may also assign access levels to users to control content within that Data Management System. This control prevents misuse of the system and helps promote users who contribute more to the system by giving them recognition within the community. Additionally, this allows the administrator to manage users and assess how individuals are using the Research Data Exchange. General web analytics tools and or user surveys may also be used to assess both usage and utility.

S5. *Provide security for the system:* The Research Data Exchange represents a significant investment from contributors and users of the system, and that investment needs to be protected from outside attacks. The external applications are dependent upon the data which is stored and processed on the Research Data Exchange. To be able to successfully meet this need, the system must adhere to appropriate information technology security practices as well as any applicable federal requirements.

S6. *Provide reliable access to Research Data Exchange assets:* The components that form the physical infrastructure of the Research Data Exchange need to be, at a minimum:

- Reliable – The system must be developed to recover system failure.
- Available – The system must be designed to reduce the amount of down time encountered by users, however some downtime is allowable.
- Maintainable (and well maintained) – The system must have processes and personnel that ensure it is working properly
- Extensible – The system must be designed to adhere to current common standards to increase the chances for extending its capabilities in the future.

- Interoperable - The system must be designed to work with a broad range of applications either by developing an Application Programming Interface (API) or other simple access format.

4.3 Priorities Among Changes

The essential needs are, for the most part, those that must be met in order to provide a secure, reliable, multi-modal and multi-sourced data archive, available to a wide range of users. The system could begin to provide value once these needs are addressed. Desirable features greatly enhance the value of the Research Data Exchange, e.g., by adding real-time data feeds, facilitating communications between users, and monitoring data quality. They are not optional, but do not have to all be present for the system to begin providing value. Optional features would add further value, but are not a core part of the concept, and could be left out.

At this time, it is not possible to assign rank orders within the three priority levels.

Capability Needs (8)
Essential:

C1. Store archived transportation data
C2. Search transportation data
C4. Support external applications accessing data from the Research Data Exchange
C8. Store, update/provide detailed metadata to support the Research Data Exchange

Desirable:

C3. Store and display history/context information
C5. Implement access controls for automated access
C6. Accept real-time or near real-time data from external sources

Optional:

C7. Store and share utility software

Environment Needs (6)
Essential:

E3. Provide a single Internet-based portal for access to all data resources and meta-data
E4. Organize and support the use of mulit-source data

Desirable:

E1. Facilitate interactive communication between registered users of the Research Data Exchange
E5. Deliver and support archived data and real-time data feeds
E6. Utilize (and where necessary, develop) and improve standards in data sharing

Optional:

E2. Support the establishment and growth of an active community to build up and sustain the data sets and Data Environments in the Research Data Exchange

Operational Needs (4)

Essential:

- O1. Support quality checking on data from all sources
- O2. Establish and document governance for the Research Data Exchange
- O4. Enforce governance, address user needs, and maintain the system

Desirable:

- O3. Monitor and quantify Research Data Exchange utilization

Support Needs (6)

Essential:

- S1. Restore the system from failure
- S3. Support and manage multiple access levels for users
- S5. Provide security for the system
- S6. Provide reliable access to Research Data Exchange assets

Desirable:

- S2. Facilitate and support registered users of the system in documenting their work with the data
- S4. Track usage and manage members of the Research Data Exchange

Optional:

- None.

4.4 Changes Considered But Not Included

Consideration was given to having stronger rules requiring registration, greater restrictions on access with a formal approval process, and greater requirements for sharing the intended use of the data before approving access. This approach would ensure that all users of the system, as well as the USDOT, would have a general idea of the projects that the other users were working on. The USDOT's lack of such information was a concern that had arisen in the Next Generation Simulation (NGSIM) program. However, such requirements would also likely reduce the use of the system, and the USDOT has received expressions of concern about registration requirements being too onerous. Therefore, USDOT decided to allow as much access as reasonably possible without requiring registration, while also providing incentives for registration and information sharing (e.g., access to the community discussions). Other means, such as random voluntary user surveys, will be used to collect complementary information of interest. [Note that unregistered users will be restricted to viewing and downloading information. They may not upload data nor post in discussion forums or have any other sort of write access to the system]

5. Concepts for the Proposed System

5.1 Background, Objectives, and Scope

The development of a Research Data Exchange is an element of the Data Capture and Management Program, which is part of a United States Department of Transportation (USDOT) research effort. Through these efforts, the USDOT Intelligent Transportation Systems Joint Program Office (ITS JPO) is engaged in assessing the potential of a multisource, active acquisition data paradigm to enhance current operational practices and transform future surface transportation systems management. The USDOT research program is a collaborative initiative spanning the ITS JPO, the Federal Highway Administration (FHWA), the National Highway Traffic Safety Administratin (NHTSA), the Federal Transit Administration (FTA), and the Federal Motor Carrier Safety Administration (FMCSA). One foundational element of connected vehicle research is the Data Capture and Management Program. Program objectives include:

- Enable systematic data capture from vehicles, mobile devices, and infrastructure;
- Develop Data Environments that enable the integration of data from multiple sources for use in transportation management and performance measurement;
- Reduce costs of data management and eliminate technical and institutional barriers to the capture, management, and sharing of data.

Further details on the Data Capture and Management Program objectives, core concepts and high-level program activity plan can be found in the document *IntelliDrive Data Capture and Management Program Vision: Objectives, Core Concepts and Projected Outcomes*.

The core objective of the Research Data Exchange is to create a system that broadly shares multi-source and multi-modal connected vehicle data with the transportation community for purposes of researching, analyzing, and testing connected vehicle concepts and prototypes. To achieve this objective, the Research Data Exchange will use both current best practices from existing data warehouses and also institute new functional transformations in transportation data storage.

The three fundamental transformations that the Research Data Exchange will support are:

1. *Transformation 1: Move from single source to multisource and multimodal Data Environments.*
2. *Transformation 2: Transition from purely archival data to include real-time data provision.*
3. *Transformation 3: Move from passive acquisition (surveys and roadside sensors) to active interrogative methods (using mobile devices as programmable sensor platforms).*

These three transformations are detailed in the *IntelliDrive Data Capture and Management Program: Transforming the Federal Role.*

5.2 Operational Policies and Constraints

5.2.1 Technical Constraints

Some of the technical constraints on the proposed system are:

- *System must be Internet-based* – The system must use an Internet- based solution. This constraint is critical to make sure the system can connect all Data Management Systems over a broad geographic location and meet all accessibility needs of the users.
- *System must support diverse Data Management Systems* – The system will include multiple *autonomous* Data Management Systems that are managed and under the control of a variety of organizations.

5.2.2 Organizational and Policy Constraints

The organizational and policy constraints on the proposed system are:

- *The Federal government must take an active role in the development of the system* – The system must have the support of the Federal Government to provide the initial funding, governance, and structure for the system.
- *Intellectual property rights must be addressed* – The system must have defined intellectual property rights and policies that govern what types of licenses may or may not be used to protect the ownership of data sets and software. In general, data sets included in the Research Data Exchange must be made available under one or more types of open data licenses. The particular licenses which are allowed are to be determined. Notwithstanding this provision, it may be desirable to include some data sets that include proprietary data (e.g., freight movement data) and this data may have additional restrictions placed upon its access and use. Data licensing policies will need to be created and enforced.
- *Privacy must be addressed* – The travel data and information that will be provided by the Research Data Exchange may contain personally identifiable information. This data, which could range from vehicle locations to contact information from surveys, must be protected from unauthorized access or use. Privacy policies will need to be created and enforced. No personally identifiable information will be stored on any Federal portion of the Research Data Exchange.

5.3 Description of the Proposed System

The Research Data Exchange is a new system that is being developed from the ground up to support the new paradigm of "collect once, preserve, use many times" for transportation data. From a technical standpoint, the development of the Research Data Exchange can be accomplished using commonly deployed and well-understood technologies. In contrast, the institutional issues surrounding the Research Data Exchange are more complex as they deal with multiple organizations with different objectives and cultures that often do not freely share information with external users.

Figure 4 shows the five step lifecycle of data sets and Data Environments within the Research Data Exchange that demonstrates this new recycling data paradigm. Upon the completion of all five steps

within the cycle the initial data is transformed and enhanced with more data, metadata, and information for the next user to use. Walking through the process, the first step is the initial search of the Research Data Exchange to determine if it contains information of interest. Information about the types of data, the data format, and frequency of collection are among the parameters that will be available. Once the data set or Data Environment is located, it can be accessed either directly by an application or by a user of the system. As users become more familiar with the data they can begin to collaborate with other members. This collaboration could include sharing information they found in the data or asking questions about the data. If the data are incomplete or missing a data type, then those individuals or the community can produce and contribute the missing data to complete the data set or Data Environment. This process of creating/improving the data and then reposting it into the Research Data Exchange ensures that the community receives the benefits of the new data and the contributors also gain credit for their work.

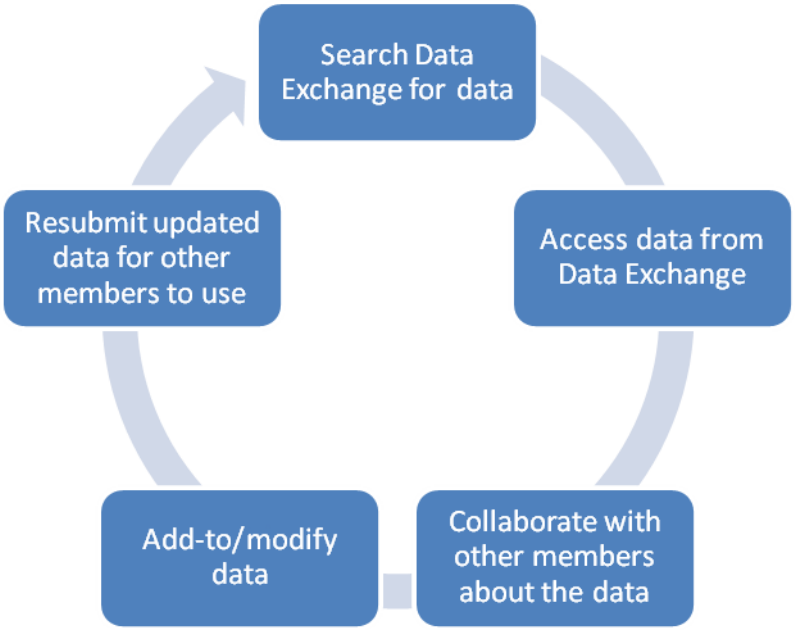

Figure 4: Data Cycle

The goal of the Research Data Exchange is to develop a transportation data sharing system that promotes multisource, multimodal, real-time, and active data acquisition and management. These new data sources and sharing capability will better support the needs of transportation researchers and developers, reducing costs while encouraging innovation.

The Research Data Exchange is a federated system of Data Management Systems containing a collection of data sets and Data Environments and real-time data feeds, plus the Data Portal (described in section 5.3.2 *Description of the Data Portal)* that serves as a gateway to the system.

5.3.2 Description of the Data Portal

The Data Portal is one single website that maintains information about and provides access to each of the included Data Management Systems. Physically, the Research Data Exchange comprises the set of linked Data Management Systems and the Data Portal. The Data Portal serves two general functions:

- The portal serves as a central location for general information about the Data Capture and Management Program as a whole.
- The portal acts as a single location for users to access all the contents of the Research Data Exchange (Data Environments, metadata, data sets, and real-time data feeds) regardless of which constituent Data Management System the information resides in.

The portal stores information about data content, projects, status, and summary information on access policies for all of the Data Management Systems associated with the Research Data Exchange. This information comes in two forms. One is the metadata detailing the data sets, Data Environments, and real-time data feeds, including data type, location, frequency, time period, and access rules. The second is in the form of news information describing changes in the Research Data Exchange, including current status of existing data sets and Data Environments, new data sets and Data Environments, and corrections to the data.

The portal is a website that provides the ability to communicate with each of the individual federated Data Management Systems. It provides mechanisms for users to link to and obtain data sets or data feeds from the associated Data Management Systems.

Another role of the portal is to register new users and allow registered users to post information about how they are using the data. This capability allows members and the administrator of the system to know who else is using the data and what they are using it for. Users register with the portal by completing a user profile located on the website.

As the gatekeeper of the connected vehicle data, the portal must allow newly registered users and returning users the ability to search and review metadata from all data sets, Data Environments, and real-time data feeds in the Research Data Exchange. The portal must have a defined control system to identify users and to allocate the proper permissions to them. Section 5.5 “User classes and other involved personnel” outlines each user role in the prototype system and what permission and access level they will have.

To the extent practical, the user interface for the Data Portal should be similar to that of the Open Source Portal, so that users familiar with one can easily navigate the other.

5.4 Modes of Operation

This section outlines the different modes of operation for a single Data Management System and the Data Portal. The portal and individual Data Management Systems all work independently of each other and so do their modes of operation. For example, one Data Management System may be put into an update mode as it makes updates to its software, but this change will only affect the single Data Management System, with all other Data Management Systems not affected. One of the key

items of metadata stored on the portal of value to the Portal Administrator responsible for the operation and well-being of the RDE is the mode information showing the status for each federated Data Management System.

The Data Portal and individual Data Management Systems have the following modes of operation:

- *Operational Mode* – This is the mode the system will be in for the majority of its operation. This mode runs when there are no changes to the system and the system is functioning properly. This mode allows all users to access the contents that they have permission to access.
- *Restore Mode* – This mode is used when the system must be restored from a previous time period because of malicious attack, system failure, or internal error within the system. During this mode the system will be unavailable to users until the restore is completed, at which time the system will be switched back to Operational Mode.
- *Update Mode* – This mode will be active when the system is going through an update to the functionality of the system. During this mode users will have limited access to the contents. This mode should be scheduled, with notices sent to members of the system announcing when this update will be conducted. Upon completion of the update the system will switch back to Operational Mode.

5.5 User Classes and Other Involved Personnel

This section identifies and describes the role of each user of the Research Data Exchange, including a description of the user, examples of the user, and the user's responsibilities. The user classes described in this section pertain to a user accessing the system through the portal, as well as administrators of any of the member Data Management Systems, and applications that use or provide data. Individual Data Management Systems may also have mechanisms for outside users to access their system directly without using the portal. These external interfaces are outside of the scope of the Research Data Exchange.

As shown in Figure 5, there are eight different roles. A user's role is defined by the roles and responsibilities the user has within the system. User levels range from an unregistered user, the lowest level, up through an administrator, the highest level. Each level provides additional privileges. These user classes are *role-based* and an individual may perform multiple roles and thus be a member of multiple user classes. For example, an individual who is a Content Manager for a specific Data Management System may also be a Registered User with Additional Access Permissions with respect to the overall Research Data Exchange.

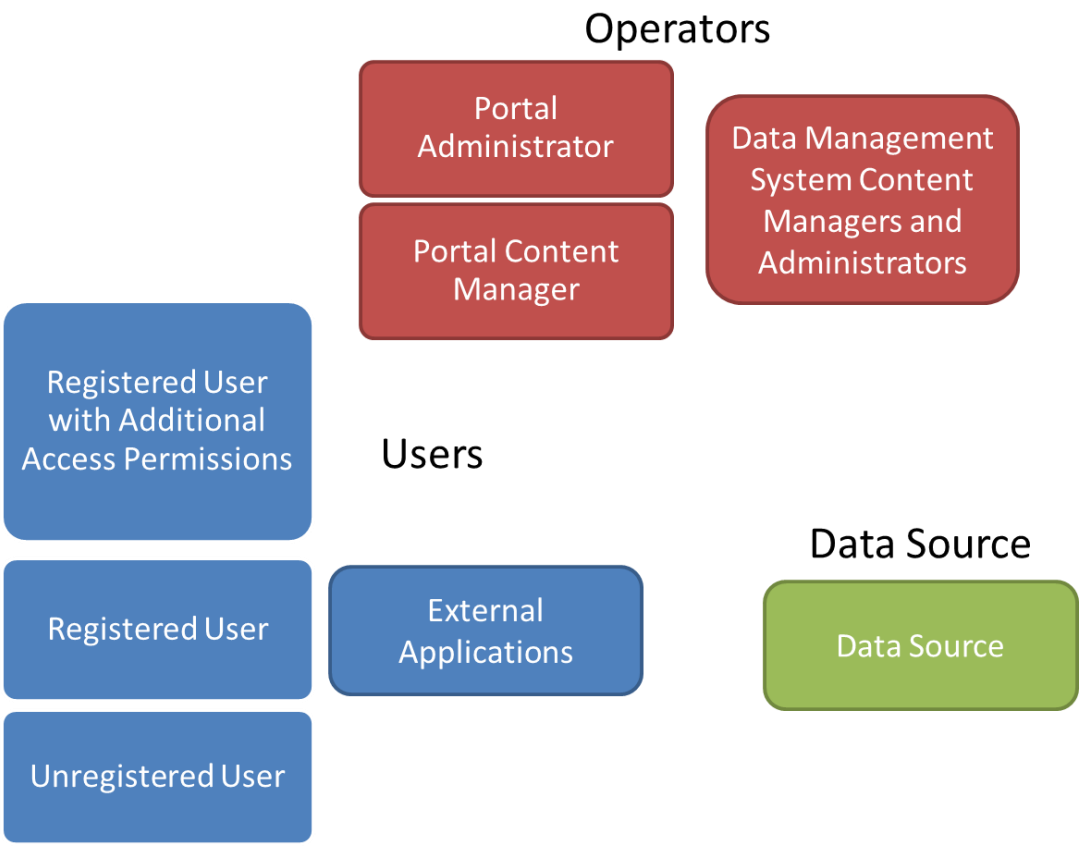

Figure 5: User Classes

5.5.1 User Class: Unregistered User

Description: Unregistered Users are defined as members of the general public who visit the system to learn about the connected vehicle project or about the Research Data Exchange. Unregistered Users are not registered with the system, and therefore cannot log in. An Unregistered User can view publically accessible information such as general information about the Reseach Data Exchange(e.g., getting started information, history/context information, terms of use, news, glossary information), as well as those data sets and Data Environments that are available to the public.

Permissions:

- Review the publicly accessible pages, including a list of data sets, Data Environments, and real-time data feeds
- Register with the Research Data Exchange
- Access and download those data sets and software utilities that do not require registration. The default access rules for data sets and software utilities are expected to permit unregistered access. However, registration requirements may be implemented for some data sets, Data Environments, or real-time data feeds.

General Examples:

- Corporations, organizations, or individuals interested in connected vehicle transportation history/context information
- Potential members looking to learn more about the data capture system before joining

Specific Examples:

- An application developer is creating software to predict travel times based on real-time transportation data, historical information, and weather conditions. The developer wants to learn if there are any already existing Data Environments that include both traffic and weather archives, for use in development and testing.
- A researcher studying traffic congestion in Washington, D.C., wants to see if there is a new data set containing D.C. traffic data.

5.5.2 User Class: Registered User

Description: To become a Registered User, a User must register through the portal. In addition to the privileges and access rights of Unregistered Users, Registered Users may have access to additional data sets and software resources. Specifically, they have access to resources that require registration for access but do not impose additional, environment-specific restrictions or approvals. Registered Users with such additional access privileges are described in the next section: Registered User with Additional Access Approvals.

Registered Users have read/write access to the Research Data Exchange community resources, which might include discussion forums, wikis, and other resources.

The registration process will include a request for contact information, information about how a User plans to use the connected vehicle data, what members are also working on the project, and any other related information that help describe their work. What information is optional versus mandatory for registration has not been determined, however in order to encourage broad access and use, mandatory information will be kept to a minimum and ease of use maximized (e.g., through the use of pick lists).

Additional Permissions:

- Search, download, and test a broader array of data sets and software
- Review and update their personal profile
- Submit a project description with information pertaining to how they plan to use or have used data from the Research Data Exchange
- Review and comment on other projects
- Update project information for their own project
- Submit data files for the system community to use
- View the public portions of other member profiles

General Examples:

- Students and researchers who are considering a transportation project and want to exchange information with those conducting similar research
- Application developers with a project idea who want to see if there are similar projects already underway

- Individuals or companies with a new modeling approach that want to leverage existing data sets for enhancing and calibrating their product.

Specific Examples:

- A university student is considering a thesis project on shockwave formations. The student uses connected vehicle metadata to learn the level of detail and types of information that may be relevant to this topic, then downloads and uses the relevant data in their research.
- A researcher is studying the benefits of using real-time data as opposed to archived transportation data when determining travel routes. The researcher registers in order to engage in discussions with other experts in this field.
- A mathematician has been hired by a company that specializes in transportation. The mathematician's new job involves analyzing data, some of which are only available on a state-run Data Management System that is part of the Research Data Exchange and has a registration requirement but no other restrictions on access. Through the Data Portal, the mathematician registers with the Research Data Exchange, views the data, metadata, and history/content information about the data sets he is interested in. This way, the mathematician has a better understanding of new technology associated with transportation data collection. As a registered User, he can access and download the data sets through the Data Portal.

5.5.3 User Class: Registered User with Additional Access Permissions

Description: A Registered User with Additional Access Permissions is a Registered User who has been granted access to one or more Data Management Systems that have rules requiring additional levels of approval. The policies for such access are controlled by each participating Data Management System. This User has met whatever additional conditions the Data Management System owner has put in place, such as a signed terms of use agreement.

Additional Permissions:

- Search, download, and test data sets and Data Environments to which they have been granted access
- Any additional permissions specific to the individual data sets, Data Environments, and real-time data feeds they have been granted access to.

General Examples:

- Researchers and application developers who need transportation data to complete their work
- Researchers and application developers studying or working with real-time data
- Agencies wanting to examine transportation data for patterns or abnormalities
- Researchers and developers working on transportation related projects who want to share their research insights and collaborate with others doing similar research

Specific Examples:

- An application developer is creating software to send pre-trip congestion alerts to cell phones via text messages. The developer uses the real-time traffic information from the Research Data Exchange to test or improve the application. Some of the data is stored in a Data Management System that requires Users to agree to a specific "terms of use" agreement.

- A university researcher is conducting research on freight movement traveler information systems. They need information about freight movements within a region. A participating Data Management System has such information, and will make it available for non-commercial research provided specific agreements on non-disclosure of the proprietary information are agreed to.

5.5.4 User Class: Portal Content Manager

Description: A person is designated a content manager when there is a need for that person to have the ability to add, update, and delete certain information in the system. Only an administrator can promote an individual to this User level. A content manager has the ability to create and modify data files and the most publicly accessible information. Unlike administrators, content managers cannot change the portal structure.

Additional Permissions:

- Monitor the sharing of data, metadata, or other information
- Remove unwanted information or messages
- Add, update, and delete data files
- Add, update, and delete getting started, terms of use, history/context, and glossary information within the Data Portal.

General Examples

- Individuals, identified by the administrator to monitor the system and ensure it is being used as intended
- connected vehicle Federal team members who update history/context information as the project moves forward
- connected vehicle test bed analysts who check data files for errors and update those files with error flags

Specific Examples

- A researcher for connected vehicle determines a way to make the data collection process used in the Michigan Test Bed test more efficient. The researcher writes an article about this new process. The content manager posts either a link to the article, or, with the copyright holder's permission, reposts the article on the Data Portal
- An administrator does not have enough time to devote to managing all aspects of the system. The administrator appoints a content manager to review submitted project descriptions and post them when approved. In addition, the content manager monitors the system and ensures that Users are not creating inappropriate or irrelevant messages.
- As the Data Capture and Management Program expands and the Research Data Exchange houses more data from a variety of sources, a Federal Data Capture and Management Program team member updates the history/context information to explain how this new data was collected.
- An analyst working for the Connected Vehicle Program reviews archived Michigan Test Bed metadata files to check for errors or inconsistencies. The analyst updates the files with error flags.

5.5.5 User Class: Portal Administrator

Description: A person is designated an administrator only when there is a need for that person to have administrator privileges. A current system administrator must promote the person to administrator level. An administrator maintains and enhances the system, manages User registrations (including the ability to approve, restrict, or delete members), and has the authority to change the portal's content and structure.

Additional Permissions:

- Monitor the sharing of data, metadata, or other information
- Remove unwanted information or messages
- Review new members to ensure the new Users are valid. [Detailed requirements have not been determined. This might be as simple as verifying that a valid email address was provided.]
- Receive Users' questions and respond appropriately
- Run system backup
- Restore system from backup after a problem occurs
- Add, update, and delete data files
- Assign access rights for data files housed in the Federal portions of the Research Data Exchange
- Test, check, and add new data sets
- Add new data sets
- Add, update, and delete history/context information within the portal environment.
- Set up tasks for an automated scheduler

5.5.6 User Class: Data Management System Content Managers and Administrators

Description: Each individual Data Management System that participates in the Research Data Exchange will have its own content managers and administrators, including USDOT Data Management Systems. The roles, responsibilities, authority, policies, and procedures will be set by each individual member Data Management System.

5.5.7 User Class: External Applications

Description: External Applications are automated processes that access the Research Data Exchange. They will have access levels and permissions that mirror those of human users. However, because of the increased potential for either intentional or unintentional harm caused by automated access, registration and approval of some sort will be required. External applications have the following characteristics:

- Directly access the data in a Data Management System through authentication
- Use some or all of the data from the system as an input into the application
- Are publically or privately owned
- Must be approved before receiving any data

Capabilities:

- Provide authentication information in order to access data
- Access and use real-time or archived connected vehicle data

General Examples:

- External application designed to take real-time data as an input
- External application designed to take archived data as an input
- External application designed to take both real-time and archived data as inputs

Specific Examples:

- An external application is designed to take real-time data inputs and determine estimated train, bus, and taxi trip times between an origin and destination. The application administrator provides authentication information in order to access real-time transportation data in the system. After receiving authentication, the same application accesses the real-time traffic data, train data, and bus data in the system and uses this information to determine estimated travel times for each method of transportation.
- An external application is designed to take in both real-time and archived transportation data and determine if a traveler should take bus, train, or subway. The decision is based on the estimated travel time and environmental impact of each method of travel. The application provides authentication information and then accesses both the archived and real-time data of interest.

5.5.8 User Class: Data Source

Description: A data source is an external source which provides real-time data to the system. These sources could include raw and processed real-time data.

Capabilities:

- Provide real-time data stream(s) to one or more participating Data Management Systems
- Provide authentication information which could include login information or security protocol information

General Examples

- A data source provides authentication information and is authenticated. Then a data stream of raw data enters the system.
- A data source provides authentication information and is authenticated. A data stream of processed data enters the system.

Specific Examples

- A wireless roadside traffic data collector provides authentication information and is authenticated by the system. Then a data stream of raw traffic data is transmitted from the wireless data collector to the system.
- A data stream is corrected by an external filter to fix errors and fill in missing information. The source sends authentication information to the system and is authenticated. Then the complete, corrected data stream is sent to the system.

5.5.8 Summary of User Classes

Table 5-1 provides a summary of the user classes and their respective permissions described in sections 5.5.1 through 5.5.8.

Table 5-1: User Class Summary Chart

Title	Description	Permissions
Unregistered User	Unregistered Users are defined as members of the general public who visit the system to learn about the connected vehicle project or about the Research Data Exchange. Unregistered Users are not registered with the system, and therefore cannot log in. An Unregistered User can view publically accessible information such as general information about the Research Data Exchange (e.g., getting started information, history/context information, terms of use, news, glossary information), as well as view and download those data sets and Data Environments that are available to the public.	• Review the publicly accessible pages, including a list of data sets, Data Environments, and real-time data feeds • Register with the Research Data Exchange • Access and download those data sets and software utilities that do not require registration. The default access rules for data sets and software utilities are expected to permit unregistered access. However, registration requirements may be implemented for some data sets, Data Environments, or real-time data feeds.
Registered User	Registered Users are Users who register with and provide information to the system. In addition to the privileges and access rights of Unregistered Users, Registered Users may have access to additional data sets and Data Environments. Specifically, they have access to resources that require registration for access but do not impose additional, environment-specific restrictions or approvals. Registered Users with such additional access privileges are described in the next section on Registered User with Additional Access Approvals. Registered Users have read/write/download access to the Research Data Exchange community resources, which might include discussion forums, wikis, and other	• All privileges of Unregistered Users • Search, download, and test a broader array of data sets and software • Review and update their personal profile • Submit a project description with information pertaining to how they plan to use or have used data from the Research Data Exchange • Review and comment on other projects • Update project information for their own project • Submit data files for the

	resources.	system community to use • View the public portions of other member profiles
Registered User with Additional Access Permissions	A Registered User with Additional Access Permissions is a Registered User who has been granted access to one or more Data Management Systems that require additional levels of registration and/or approval. The policies for such access are controlled by each participating Data Management System.	• All privileges of Registered Users • Search, download, and test data sets and source code within the Data Management Systems to which they have been granted access • Any additional permissions specific to the individual Data Management System they have been granted access to.
Portal Content Manager	A person is designated a content manager when there is a need for that person to have the ability to add, update, and delete certain information in the system. Only an administrator can promote an individual to this user level. A content manager has the ability to create and modify data files and the most publicly accessible information. Unlike administrators, content managers cannot change the portal structure.	• All privileges of Registered Users • Monitor the sharing of data, metadata, or other information • Remove unwanted information or messages • Add, update, and delete data files • Add, update, and delete getting started, terms of use, history/context, and glossary information within the Data Portal
Portal Administrator	An administrator maintains and enhances the system, manages user registrations (including the ability to approve, restrict, or delete members)), and has the authority to change the portal's content and structure.	• All privileges of Portal Content Managers • Monitor the sharing of data, metadata, or other information • Remove unwanted information or messages • Review new members to ensure the new Users are valid. [Detailed requirements have not been determined. This might be as simple as verifying that a valid email address was provided.]

		• Receive Users' questions and respond appropriately • Run system backup • Restore system from backup after a problem occurs • Add, update, and delete data files • Assign access rights for data files housed in the Federal portions of the Research Data Exchange • Test, check, and add new data sets • Add new data sets • Add, update, and delete history/context information within the portal environment. • Set up tasks for an automated scheduler
Data Management System Content Manager and Administrators	Similar to the Portal Content Manager and Portal Administrator roles, but for each individual participating Data Management System.	• Determined by the entity responsible for the specific participating Data Management System
Applications	Applications are programs which use real-time or archived data.	• Use real-time or archived connected vehicle data • Provide identification information
Data Source	A data source provides real-time data to the system.	• Provide real-time data • Provide identification information

5.6 Support Environment

The Research Data Exchange operates at the direction of the USDOT Connected Vehicle Program. As a precursor to this system a Prototype Data Exchange has been developed and is currently operated and maintained by Noblis Inc. under the direction of the USDOT Connected Vehicle Program. As the program advances the Research Data Exchange is expected to be developed and maintained by another USDOT support contractor. The data sets and Data Environments will be supported and created by contractor procurements throughout the program.

6. Operational Scenarios

The operational scenarios outlined in this section are based upon the Prototype Data Exchange that is currently being developed and maintained by Noblis at the direction of the Data Capture and Management Program. This system consists of a website and one Data Environment as of fall 2010. This prototype website serves the role of both a Data Management System and Data Portal in the context of the Research Data Exchange. The Data Capture and Management Research Data Exchange will have more capabilities than the prototype, therefore the scenarios list is expanded from those supported by the prototype in order to capture the full functionality of the Data Capture and Management Research Data Exchange.

For convenience, the scenarios are divided into four categories:

- User interaction, registration, and log-in
- Information exchange
- Information management
- System administration

6.1 User interaction, registration, and log-in

6.1.1 Operational Scenario: View publicly accessible content

Actor: Unregistered User, Registered User, Registered User with Additional Access Permissions, Portal Content Manager, or Portal Administrator (User).
Description: A User reads publically accessible content including getting started information, terms of use information, connected vehicle history/context information, news articles, and the glossary.
Preconditions: User has navigated to the system using Internet-based protocols.

Steps:

1. User navigates to the publicly accessible content of interest..
2. User reads desired information.

If the User wishes to view other publicly accessible content, continue from the beginning.

6.1.2 Operational Scenario: Search publicly accessible pages

Actor: Unregistered User, Registered User, Registered User with Additional Access Permissions, Portal Content Manager, or Portal Administrator (User)
Description: A User is looking for certain content. The User inputs search criteria into the system to find the desired information.

Preconditions: The User has navigated to the system using Internet-based protocols.

Steps:

1. The User inputs the search criteria, which is either a website link selection or a keyword.
2. User views publicly accessible information (see "View publicly accessible content").

6.1.3 Operational Scenario: Register with system

Actor: Unregistered User and Administrator
Description: An Unregistered User wants to become a Registered User and have access to additional data and/or participate in the community resources provided by the Research Data Exchange. The Unregistered User registers on the site to become a Registered User. An Administrator reviews the Unregistered User's information to make sure the new Registered User is valid.
Preconditions: The Unregistered User has navigated to the system using Internet-based protocols.
Steps:

1. The Unregistered User navigates to the registration page.
2. The Unregistered User reads the terms of use of the site.
3. The Unregistered User agrees to the terms of use of the site.
4. The Unregistered User enters all mandatory information and, at their discretion, optional requested information. [The specific mandatory and optional information has not been determined. Mandatory information is expected to include: desired username, desired password, valid email address. It may also include full name, organization, organization type, country, and interest with connected vehicle.]
5. System and administrator utilize some combination [to be determined] of automated and manual authentication and registration review procedures (e.g., CAPTCHA,[10] required response to automated email, etc.) to validate registration request.
6. If user passes validation checks, Administrator approves registration, provides email to user.
7. If user fails to authenticate (e.g., user provides an invalid email address or fails to respond to the registration email), Administrator rejects registration.

6.1.5 Operational Scenario: Unauthorized user tries to view content that is not publicly accessible.

Actor: Unregistered User
Description: An Unregistered User tries to access information that is not available to the public using non-traditional means. The Unregistered User is denied access to the content.
Preconditions: None
Steps:

1. Unregistered User attempts to view content that is not publicly accessible by going directly to the URL of the content or accessing the page through non-traditional means.
2. The Unregistered User is denied access to the content.

[10] CAPTCHA is a type of challenge-response test that attempts to ensure that the response is generated by a human rather than a computer. "CAPTCHA" a contrived acronym for "Completely Automated Public Turing test to tell Computers and Humans Apart."

6.1.6 Operational Scenario: Ask a question to the Administrator

Actor: Unregistered User, Registered User, Registered User with Additional Access Permissions, or Portal Content Manager (User) and Administrator
Description: A User has a question about information in the Research Data Exchange or about the way the Research Data Exchange is working. The User sends a message to an Administrator, who answers the question or fixes the problem if necessary.
Preconditions: The User has navigated to the system using Internet-based protocols.
Steps:

1. The User navigates to the appropriate section to send a message to the Administrator.
2. The User inputs his email address, a message subject, and his question.
3. The Administrator reads the question and sends a reply if necessary.
4. The subsequent step happens when the Administrator sends a reply:
5. The User reads the Administrator's response message.

6.1.7 Operational Scenario: Log in to the system

Actor: Registered User, Registered User with Additional Access Permissions, Portal Content Manager, or Portal Administrator (User)
Description: The User wants to access the information available at his user level in addition to the publically accessible information. The User inputs his log-in information and the system authenticates the User. The User then has access to the information available at his user level.
Preconditions: The User has navigated to the system using Internet-based protocols.
Steps:

1. The User navigates to the log-in area.
2. The User enters his log-in information.
3. The subsequent steps happen when the User enters his information correctly:
4. The User is authenticated.
5. The User is given the ability to access data, review his profile, and access the community collaboration sections of the Research Data Exchange. If the User is a Registered User with Additional Access Permissions, the User is also given access to the specific restricted access data sets and/or Data Environment(s) that he has been granted access to. If the user is a Portal Content Manager, the User is given the ability to create, update, and delete certain content information. If the User is an Portal Administrator, the user also receives the ability to modify all portal system information.
6. The System Logger records the date, time, and description of the log-in event.
7. The subsequent steps happen when the User enters his information incorrectly:
8. The system does not authenticate the log-in information and displays an error message.
9. The Registered User re-enters his log-in information. Repeat from step 3 if the information is correct or from step 5 if it is incorrect.

6.1.8 Operational Scenario: Request a new password

Actor: Registered User, Registered User with Additional Access Permissions, Portal Content Manager, or Portal Administrator (User)

Description: The User wants to log in to the system but has forgotten his password and so requests a new one. The User must provide identification information before the password is reset. The User then creates a new password and logs in.
Preconditions: The User has navigated to the system using Internet-based protocols.
Steps:

1. The User navigates to the request new password area.
2. The User provides identification information, including the email address they provided when originally registering.
3. The system sends an email to the User with a unique web URL that allows them to create a new password. It may or may not include a temporary password as part of the procedure.
4. The User goes to the provided web URL and creates a new password.
5. The User is logged into the site.

6.2 Information exchange

6.2.1 Operational Scenario: Download data (or utility software)

Actor: Unregistered User, Registered User, Registered User with Additional Access Permissions, Portal Content Manager, External Application, or Portal Administrator (User)
Description: The User finds sample data and wants to download it. If required for the information in question, the user logs in to the system and his credentials and permissions are checked against those required for accessing the requested data. If there are no restrictions on access or the user is authenticated and has the required permissions, the user downloads the file and views the data.
Preconditions: User has navigated to the system using Internet-based protocols and is logged in to the system. User has found the information to be downloaded (see "Search data")
Steps:

4. User selects the data or software to be downloaded.
5. If the user is not logged in and login and/or additional permissions are required for the data in question, the user logs in to the system.
    - The system checks the credentials and permissions of the user
    - If the user is authenticated and has the appropriate permissions, they are granted access. If not, an error message is returned to the user.
    - A limited number of attempts is allowed.
6. The User downloads the sample data.
7. The User views or stores the data or software.

6.2.2 Operational Scenario: Review and edit profile

Actor: Registered User, Registered User with Additional Access Permissions, Portal Content Manager, or Portal Administrator (User)
Description: The User wants to view and edit a user profile. The User inputs the updated profile information and the system saves the changes.
Preconditions: The User has navigated to the system using Internet-based protocols and is logged in to the system.
Steps:

1. If the User is not an Administrator or Content Manager, he navigates to his personal profile. If the User is an Administrator or Content Manager, he navigates to any user's profile.

2. User inputs updated profile information such as email address, password, name, organization, organization type, country, and interest with connected vehicle.

6.2.3 Operational Scenario: Collaborate with other Registered Users

Actors: Registered User, Registered User with Additional Access Permissions, Portal Content Manager, or Portal Administrator (User)
Description: A User wants to share information with or ask other users a question. The User sends a message or contacts other Users, who can choose to respond and collaborate with the first User.
Preconditions: Users have navigated to the system using Internet-based protocols and are logged in to the system.
Steps:

1. Registered User navigates to collaboration area.
2. Registered User inputs message information including the title and the actual message. The message could be a question, a request for help, or information that other Registered Users may find useful.
3. Another Registered User navigates to the collaboration area.
4. The other Registered User views the message.
5. If the other Registered User chooses to respond to the message, the subsequent steps take place:
6. The other Registered User submits a reply message.
7. The first Registered User views the reply. If he chooses to send a response back, continue from step 5.

6.2.4 Operational Scenario: Search for and view another Registered User's profile

Actors: Registered User, Registered User with Additional Access Permissions, Portal Content Manager, or Portal Administrator (User)
Description: A User wants to view the profile information of another registered system user. The User navigates to the profile of interest and reads the profile information.
Preconditions: User has navigated to the system using Internet-based protocols and is logged in to the system.
Steps:
This scenario begins in one of two ways. The first option is:

1. User navigates to the Registered Users list. Continue with step 3.

Alternatively:

2. User inputs all or part of another User's username. Continue with step 3.
3. User inputs his selection choice into the system.
4. User views the other Registered User's profile information

6.2.5 Operational Scenario: View projects list and project information

Actor: User, Registered User with Additional Access Permissions, Portal Content Manager, or Portal Administrator (User)

Description: The User navigates to the projects list and views all of the projects. The User selects a project and views that project's information.
Preconditions: User has navigated to the system using Internet-based protocols.
Steps:

1. User navigates to the projects list
2. User selects a project of interest
3. User views the information for that project

If the User wants to view the profile for a Registered User of that project, the subsequent steps take place:

4. If not already logged in, User logs in (Unregistered Users are not allowed to view user profiles)
5. User selects the Registered User whose profile he wishes to view.
6. User views the profile.

6.2.6 Operational Scenario: Update project information

Actor: Registered User, Registered User with Additional Access Permissions, Portal Content Manager, or Portal Administrator (called User)
Description: The User wishes to update a particular project's information. The User inputs updated project information which could include the project's title, description, leader, Registered Users, and reference information.
Preconditions: User has navigated to the system using Internet-based protocols and is logged in to the system.
Steps:

1. If the User is a Registered User, the User navigates to his project. If the User is an Administrator or Content Manager, the User navigates to any project he chooses. (See "View projects list and project information," steps 1-3).
2. User enters updated project information which may include any or all of the following: title, description, leader, Registered Users, and reference site information.
3. Registered Users can only edit information concerning their own projects.

6.2.7 Operational Scenario: Provide real-time data feed

Actor: Registered User, Registered User with Additional Access Permissions (called User) or External Application.
Description: Some of the participating Data Management Systems will be capable of providing real-time data feeds to authorized users. The authorization and approval process will be determined by each participating Data Management System. An authenticated user or external application with the appropriate permissions logs into the system and receives the real-time data that it requests and is authorized to receive.
Preconditions: User or External Application is authorized to pull real-time data
Steps:

1. User or External Application requests real-time data.
2. User or External Application provides authentication information.
3. User or External Application is authenticated.
4. User or External Application receives real-time data feed.

Note: How a participating Data Management System receives the real-time data it makes available is outside of the scope o this Concept of Operations.

6.3 Information management

6.3.1 Operational Scenario: Add data files and assign access to the files

Actor: Data Management System Content Manager, or Data Management System Administrator (User)
Description: The User needs to add new data files to a Data Environment. The User inputs the associated file information including the title, description, and categorization. If the User is an Administrator, assigning categorization includes assigning access to the files. The User inputs the actual file.
Preconditions: User has navigated to the system using Internet-based protocols and is logged in to the system.
Steps:

1. User inputs file information including: title, description, categorization (if the User is an Administrator, this includes who has access to the file), and the actual file.

If the file information input is valid, the subsequent step takes place:

2. The system stores the date, time, and description of the “add file” event as part of the event log in the data store.

If the file information input is not valid, the subsequent steps take place:

3. The User receives an invalid input error message.
4. The User reenters the file information, if he chooses. Repeat from step 2 if the input is valid or from step 3 if it is not.

6.3.2 Operational Scenario: Update data files

Actor: Data Management System Content Manager, or Data Management System Administrator (User) Description: The User needs to update a data file. The User inputs the updated file information including any or all of the following: title, description, categorization, and the actual file. The system logs event in the data store.
Preconditions: User has navigated to the system using Internet-based protocols and is logged in to the system.
Steps:

1. User searches for the appropriate data file. (See “Search sample data” or “Search non-sample data,” steps 1-4).
2. User enters updated file information which may include any or all of the following: title, description, categorization (if the User is an Administrator, this includes who has access to the file), and the actual file.

If the file information input is valid, the subsequent step takes place:

3. The system stores the date, time, and description of the “update file” event as part of the event log in the data store.

If the file information input is not valid, the subsequent steps take place:

4. The User receives an invalid input error message.
5. The User reenters the file information, if he chooses. Repeat from step 3 if the input is valid or from step 4 if it is not.

6.3.3 Operational Scenario: Delete data files

Actor: Data Management System Content Manager, or Data Management System Administrator (User) Description: The User has reason to delete a data file. The User selects the data file to be deleted and deletes the file.
Preconditions: User has navigated to the system using Internet-based protocols and is logged in to the system.
Steps:

1. User searches for the appropriate data file. (See "Search sample data" or "Search non-sample data," steps 1-4).
2. User deletes the data file.

If the deletion request input is valid, the subsequent step takes place:

3. The system stores the date, time, and description of the "delete file" event as part of the event log in the data store.

If the file information input is not valid, the subsequent steps take place:

4. The User receives an invalid request error message.
5. The User reenters the deletion information again, if he chooses. Repeat from step 3 if the input is valid or from step 4 if it is not.

6.3.4 Operational Scenario: Update portal content

Actor: Portal Administrator or Portal Content Manager.
Description: An Administrator or Content Manager has updated getting started information, terms of use information, history/context information, or an updated news article, glossary term, or website link to modify in the Data Portal. The Administrator or Content Manager inputs the updated content information.
Preconditions: Administrator has navigated to the system using Internet-based protocols and is logged in to the system
Steps:

1. Administrator or Content Manager enters updated content information including any or all of the following:
   A. news article (title, author, date, and article text)
   B. glossary term (term name and description)
   C. website link (title, URL, and description)
   D. getting started information (title and text content)
   E. history/context information (title and text content)
   F. terms of use information (title and text content)

   The Administrator can also update who has access to the content, if necessary. If the Administrator or Content Manager enters valid information, the subsequent steps take place:

2. The system stores the date, time, and description of the "update content" event as part of the event log in the data store.
3. The Administrator or Content Manager receives a "content updated" validation message.
4. If the Administrator or Content Manager enters invalid information, the subsequent steps take place:
5. The Administrator or Content Manager receives an error message.

6. The Administrator or Content Manager reenters the information. If the new information is valid, continue from step 2. If the information is not valid, continue from step 4.

6.3.5 Operational Scenario: Administrator or Content Manager deletes portal content

Actor: Portal Administrator or Portal Content Manager
Description: An Administrator or Content Manager has reason to delete getting started information, terms of use information, history/context information, or a news article, glossary term, or website link from the Data Portal. A possible reason could be the information is not longer correct or relevant to connected vehicle. The Administrator or Content Manager selects the information to be deleted and removes it from the system.
Preconditions: Administrator has navigated to the system using Internet-based protocols and is logged in to the system.
Steps:

1. Administrator or Content Manager searches for the appropriate content. (See "search publicly accessible pages")
2. Administrator or Content Manager deletes the content.
3. The system stores the date, time, and description of "delete content" event as part of the event log in the data store.

6.4 System Administration

6.4.1 Operational Scenario: Promote Registered User of the Research Data Exchange to be an Administrator

Actor: Registered User, Registered User with Additional Access Permissions, or Portal Content Manager (User) and Administrator
Description: There is a need for a User to be promoted to a system administrator. The Administrator changes the User's user level to that of an administrator.
Preconditions: Administrator has navigated to the system using Internet-based protocols and is logged in to the system. The Administrator has a reason to promote another User to be an Administrator.
Steps:

1. Administrator searches for and views the profile of the User to be promoted. (See "Search for and view another Registered User's profile.")
2. Administrator changes the User's access level to that of an administrator.

6.4.2 Operational Scenario: Notify Administrator of misuse of collaboration area so the Administrator can remove the unwanted information.

Actor: Registered User, Registered User with Additional Access Permissions (User) and Portal Administrator

Description: A User creates a message that is inappropriate or unrelated to connected vehicle. An Administrator receives notification of misuse of the collaboration area or notices the misuse himself. The Administrator removes the unwanted message.
Preconditions: User and Portal Administrator both have navigated to the system using Internet-based protocols and are logged in to the system.
Steps:

1. User posts an inappropriate or non-connected vehicle-related message in the collaboration area. (See "Collaborate with other Registered Users with Approved Projects", steps 1-4)
2. If someone informs the Administrator of the misuse of the collaboration area, the subsequent steps take place:
3. The Administrator receives notification that the collaboration area has been misused.
4. The Administrator navigates to the collaboration area and views the recent messages, including the inappropriate or non-connected vehicle-related message.
5. The Administrator deletes the message from the system.
6. The Administrator may also take any or all of the following optional actions:
    a. Notify the originator that there post has been removed, along with the reason.
    b. Warn the originator that they are at risk for removal if misuse or abuse of the system continues.
    c. Removes the User's registration. The former Registered User is now an Unregistered User.

6.4.3 Operational Scenario: Run manual portal backup

Actors: Portal Administrator
Description: An Administrator has reason to back up the portal system. The Administrator requests that backup information be sent to a specific alternative data store.
Preconditions: Administrator has navigated to the system using Internet-based protocols and is logged in to the system.
Steps:

1. Portal Administrator requests system backup.
2. Portal Administrator inputs alternative data store to which the system information should be sent.
3. The system stores the date, time, and description of the backup event as part of the event log in the data store.

6.4.4 Operational Scenario: Restore the portal from backup

Actor: Portal Administrator
Description: There is a problem with the system which requires a system restore. The Portal Administrator requests the system restore to a particular version from the alternative data store.
Preconditions: Portal Administrator has navigated to the system using Internet-based protocols and is logged in to the system.
Steps:

1. Portal Administrator requests system restoration.
2. Portal Administrator selects the version to which the system should restore.
3. The system stores the date, time, and description of the restoration event as part of the event log in the data store.

6.4.5 Operational Scenario: Add new section in the collaboration area

Actor: Portal Administrator
Description: A Portal Administrator has a reason to add a new section in the collaboration area. Possible reasons include the addition of new data sets, the need for an existing category to be broken into more specific categories, or the addition of similar projects which could benefit from a new collaboration area geared toward those projects. The Administrator creates a new section, gives it a title, and specifies who can access the section.
Preconditions: Administrator has navigated to the system using Internet-based protocols and is logged in to the system.
Steps:

1. Administrator inputs information about the new section, including a title and who can access the section.
2. The system stores the date, time, and description of the "add new section in the collaboration area" event as part of the event log in the data store.

6.4.6 Operational Scenario: Add new participating Data Management System

Actor: Portal Administrator, Data Management System Administrator, Portal Content Manager
Description: The managers of the Research Data Exchange and a non-participating Data Management System have agreed to add the Data Management System as a participating part of the Research Data Exchange. The Data Management System joins the Research Data Exchange.
Preconditions: The USDOT and the operator of a non-participating Data Management System have agreed that the Data Management System will be added to the Research Data Exchange. Administrator or both Content Manager and Administrator have navigated to the system using Internet-based protocols and is logged in to the system. The interfaces and access control policies have been agreed to by both parties. The required information about the new resource has been provided to the Portal Content Manager.

Steps:

1. The Portal Administrator and Data Management System Administrator conduct tests of the interface
2. If successful, the interface is made part of the operational system. Portal Administrator and the Data Management System Administrator set the necessary parameters and permissions.
3. The Portal Content Manager updates the portal with information on the new environment.
4. The system stores the date, time, and description of the "add new data set" event as part of the event log in the data store.

6.4.7 Operational Scenario: Add portal content and assign access to content

Actors: Portal Administrator or both Portal Content Manager and Portal Administrator
Description: An Administrator or Content Manager has new getting started information, terms of use information, history/context information, or a new news article, glossary term, or website link to add to the Research Data Exchange. The Administrator or Content Manager inputs the new content information. An Administrator assigns which user levels can access the content.

Preconditions: Administrator or both Content Manager and Administrator have navigated to the system using Internet-based protocols and is logged in to the system.
Steps:

1. Administrator or Content Manager enters new content information, which might include:
    a. news article (title, author, date, and article text)
    b. glossary term (term name and description)
    c. website link (title, URL, and description)
    d. getting started information (title and text content)
    e. history/context information (title and text content)
    f. terms of use information (title and text content)
2. The Administer inputs who can access the content.

If the Administrator or Content Manager enters valid information, the subsequent steps take place:

3. The system stores the date, time, and description of the "add content" event as part of the event log in the data store.
4. The Administrator or Content Manager receives a "content added" validation message.

If the Administrator or Content Manager enters invalid information, the subsequent steps take place:

5. The Administrator or Content Manager receives an error message.
6. The Administrator or Content Manager reenters the information. If the new information is valid, continue from step 2. If the information is not valid, continue from step 4.

6.4.8 Operational Scenario: Set up automated scheduler

Actor: Portal Administrator
Description: An Administrator needs to schedule when the system should perform backup, update the search indexing, or send data to applications. The Administrator inputs the date and time for each action and the system creates a schedule of actions to be carried out.
Preconditions: Administrator has navigated to the system using Internet-based protocols and is logged in to the system.
Steps:

1. Administrator inputs scheduling information including the date, time, and action to be completed. The action to be completed may be a system backup, a search indexing update, or a send data to applications action.
2. The Scheduler creates a schedule of actions to be carried out.
3. The Scheduler runs the schedule of actions. This includes calling for a system backup, or updating the search indexing information, all at the scheduled date and time.

6.4.9 Operational Scenario: Delete user

Actor: Portal Administrator
Description: An Administrator has a reason to delete a user. Reasons to delete a user include: the user misused the collaboration area, the user misused data, or the user no longer needs access to connected vehicle data. The Administrator removes the user from the system.
Preconditions: Administrator has navigated to the system using Internet-based protocols and is logged in to the system.
Steps:

1. Administrator navigates to any user's profile.
2. Administrator deletes the user's profile information and removes the user from the system.

7. Summary of Impacts

7.1 Operational Impacts

The Data Capture and Management Program, with the Research Data Exchange as the centerpiece, seeks to change current research data practices of "collect, use once, and neglect" to new practices based on the tenet of "collect once, preserve, use many times." In order to realize this change, the following operational impacts from the new system must be achieved:

- *Single virtual location for detailed connected vehicle data that is multi-source and multi-modal:* The system will create a single location where users will be able to gain access to consolidated real-time and archived connected vehicle transportation data from multiple sources and multiple modes.
- *Designed to reinforce long-term stewardship of data.* The system will provide a means where members will be able to store and share any improvements that they have made to existing data or new connected vehicle data. The Research Data Exchange is not intended to be a "capability in perpetuity" for maintaining and supporting all data for studies and application development. Initially data obtained will be managed under Federal aegis in order to support federal goals in application development and research. However, migration to private sector or other Data Management Systems for long-term storage of data sets beyond what is required to meet federal goals is one possible data "sunset" strategy.
- *Metadata for connected vehicle data:* The system will store metadata for all of its stored connected vehicle data. The metadata will add quality and usefulness to all of the information stored within the system.
- *Collaboration space for users:* The system will provide methods for members to collaborate with each other about how they are using the connected vehicle data. This effort will give members the ability to report errors, allow users to ask questions, and explore how other members are using the data.
- *Membership Tracking:* The system will be able to track its members' use of the system so members and managers of the system will be able to know what data and functions in the system are most popular.
- *Project Tracking*: The system will be able to support users who want to share project information. This will allow the community of researchers in a single topic area the option of sharing progress and issues with the data. This ability can help minimize duplicative work and also create a location where all members of the system can review related work.
- *Governing rules for the community:* The system will provide a means for users to agree to terms of use of the system. A governing system will protect contributors to the system and also set guidelines on how data sharing will be conducted. These governing rules will include intellectual property rights and privacy issues.
- *Lessons learned for use in deployment:* The experience gained from developing, operating, and managing the system will provide valuable insights, requirements, and lessons learned that will be utilized by implementers of connected vehicle systems, including data providers, systems integrators, and applications developers.

7.2 Organizational Impacts

No organizational changes need to be made to meet the system needs. A member of the connected vehicle Federal team (or its designee) will be the administrator. An organization or set of organizations with the appropriate expertise and experience will be required to develop, operate, and maintain the connected vehicle Research Data Exchange.

7.3 Impacts During Development

The development of the Prototype System by Noblis will have impacts on the needs and requirements for the final production-level connected vehicle Research Data Exchange. As users access, test, and interact with the prototype site, new needs will be discovered that will have to be addressed by new functionality on the site. These new needs will be recorded and included in later versions of this ConOps document.

8. Analysis of the Proposed System

8.1 Summary of Improvements

Creating a Research Data Exchange that incorporates large quantities of real-time and archived transportation data will allow researchers and application developers to easily access data and associated metadata in a single place. This saves a great deal of time, effort, and money. In addition, the Research Data Exchange allows members to see what projects others are working on and to collaborate with other project teams. This capability helps to prevent two different groups of people from unintentionally duplicating efforts, and gives members with approved projects another set of people to review and comment on their work.

As members work with and improve the data, they are encouraged to report the improvements back to the Research Data Exchange. This collaboration gives other members access to the improved data, yielding better results for all.

8.2 Disadvantages and Limitations

There are limitations to the proposed system.

- *The system cannot incorporate all the data that potential users could possibly need:* Although the system will be designed to store large amounts of various connected vehicle transportation data, it is unrealistic to believe that every bit of data can be stored in the system. There will be format, size, property rights and potentially interface issues that will prevent some transportation data from being available to the system. This limitation could limit the usefulness of the system for some users.
- *The system requires resources in both time and money to maintain and administer the system:* The system cannot survive on its own; support staff and new equipment (data storage, faster processors, software upgrades, etc.) will be needed to replace outdated and damaged equipment. Support staff will be required to ensure the system runs as it should and to monitor community activities.
- *The system cannot force members to be active:* The system must have an active community whose members are adding content and helping other members conduct their work to be successful. The system has no direct way to force members to be active on the system.
- *The system relies on USDOT to play a governance role to define, support and mandate development of data capture management rules.* While the Research Data Exchange will house federally-funded and other data sets and data feeds, the federally funded data sets and environments are critical to the success of the Research Data Exchange. Only through significant and active federal involvement will the concept of the Research Data Exchange be successful.
- *The system will not support the actual operations and management of transportation systems*; rather it is limited to supporting research, analysis, and testing.

8.3 Alternatives and Trade-Offs Considered

Two alternatives were considered. The first was the collection of stovepipe data sets on an application-by-application basis without any mechanisms to broaden access to or encourage re-use of data. Experience in other federal research programs (e.g., Next Generation Simulation (NGSIM) and the Road Weather Winter Maintenance Decision Support System) has demonstrated value in providing broad, easy to use access to data and tools that were developed using federal research funds. Typically, two thirds of the cost of a research program goes to items related to data, leaving one third available for actual analysis. By providing a convenient, easy to use repository, complete with metadata, researchers do not have to spend time and money recollecting data that may have already been collected by others. This baseline case was discarded because it would often result in data and tools being developed and used independently on each project, duplicating effort, increasing cost, and reducing the funding available to add new information and new capabilities.

The second alternative was a single, centralized data repository for all research data related to the federal effort examining mobile data communications in surface transportation to improve safety, mobility, and the environment. This alternative has the advantages of centralized control, a single set of policies and procedures, and the ability to impose uniform standards. However, it also would mean that non-USDOT entities with data sets would have to agree to these exact policies in order to provide any access to any of their data. This would almost certainly significantly reduce the number of data sets and tools that would be made available to the connected vehicle community.

In addition, it is believed that when possible, data, especially real-time data feeds, are best collected and managed locally- as close to the source as possible. The local project managers are the most knowledgeable about their data sets. In addition, many data holders might have concerns or encounter bureaucratic hurdles in transferring their Data Dets to the USDOT. This would also place a greater maintenance and cost burden on the USDOT's ITS research program. Instead, the selected approach based on the *federation* of data allows local access and governance to remain in place while simplifying access and providing a one-stop shop for data sets. In addition it allows for local quality control.

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 Document Number]
[NTL Number]

U.S. Department of Transportation

Research and Innovative Technology Administration