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

Policy Analysis and Recommendations for the DCM Research Data Exchange

www.its.dot.gov/index.htm
Final Report —July 2012
FHWA-JPO-12-036

Produced by the John A. Volpe National Transportation Systems Center
U.S. Department of Transportation
Research and Innovative Technology Administration
ITS Joint Program Office

Notice

Acknowledgements

The Volpe Center team would like to acknowledge the leadership of Walter During, P.E., of the Office of Transportation Management (HOTM) within the Office of Operations, Federal Highway Administration, U.S. Department of Transportation, in providing the guidance necessary to conduct the analysis that forms the basis for this document.

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA-JPO-12-036		

4. Title and Subtitle	5. Report Date
Policy Analysis and Recommendations for the DCM Research Data Exchange	July 2012
	6. Performing Organization Code
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. Performing Organization Report
Aviva Brecher, Josh Hassol, and Suzanne Sloan	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. Work Unit No. (TRAIS)
U.S. Department of Transportation Research and Innovative Technology Administration John A. Volpe National Transportation Systems Center Cambridge, MA 02142	11. Contract or Grant No. HW4A3
12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
Federal Highway Administration (FHWA) Office of Advanced Travel Management U.S. Department of Transportation 1200 New Jersey Ave., S.E. Washington, D.C. 20590	Policy Analysis, 2011-2012
	14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

This report is a policy analysis and set of recommendations regarding open data policies and policies for new, transformative data environments that are being developed as part of the Connected Vehicle research program. It is presented in three sections:

- The first section examines the opportunities and issues associated with implementing an open data policy in support of connected vehicle transportation environments during a research phase. It defines the concept of "open data" and presents a framework for understanding the key elements of an open data policy. The report presents an analysis of the challenges to implementing an open data policy for connected transportation environments, and identifies the gaps in knowledge that need further development to present a comprehensive policy.
- The second section addresses the types of policies that result in best practices and procedures for the research data exchange. These policies are specific to the program's research, development, and demonstration of a prototype data environment and reflect policies needed to support a system.
- The third section anticipates the shift in policies instituted during the research phase as data environments are implemented and/or adopted outside of Federal government. It also identifies policies for successful transition from research to commercial use.

17. Key Words	18. Distribution Statement
Open data and open data environments, dynamic mobility applications, vehicle-to-vehicle (V2V), vehicle-to-infrastructure (V2I), connected vehicle environment, privacy, liability, security, user access controls, metadata, standards, user registration, governance, oversight, intellectual property, data capture and management, policy and institutional issues, data quality assurance	

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
None	None	69	

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

Table of Contents

Executive Summary

This report is a policy analysis and set of recommendations regarding open data policies and policies for new, transformative data environments that are being developed as part of the Connected Vehicle research program.[1] This document is presented in three sections:

- Section I examines the opportunities and issues associated with implementing an open data policy in support of connected vehicle transportation environments during the research phase. It defines the concept of "open data" and presents a framework for understanding the key elements of an open data policy for connected transportation environments. It also presents an analysis of the challenges to implementing an open data policy, and identifies the gaps in knowledge that need further development to present a comprehensive policy.
- Section II is an analysis of the types of risks and policies required for the Data Capture and Management (DCM) Program's Research Data Exchange (RDE) which is a research tool to host and provide access to data that support connected vehicle research and application development and testing. While specifically designed to support research, the RDE will also assist with the identification of effective and successful system policies and practices, including policies that will guide those entities who choose to implement in the future. As the RDE is still in concept form, this report will identify whether policies are subject to change when transitioning the technologies from research into operations, and what existing best practices exist. This section will also examine the impact of developing and operating the technologies using an open data policy as a basis. This analysis results in insights regarding the trade-offs between "open data" and its benefits versus the risks and limitations.
- Section III anticipates the shift in policies instituted during the research phase as data environments are implemented and/or adopted outside of Federal government. It also identifies policies needed in support of successful transition from research to commercial use.

In summary, the findings are:

- **Section I: Open Data Policy**

  The purpose of this section is to examine the issues associated with implementing open data and open source policies. The case for adopting an open data policy is supported by current U.S. and international examples within the public sector. Several established licensing options that would facilitate implementation while addressing important liability questions pertaining to ownership and intellectual property, are highlighted. The primary advantages gained by implementing an open data policy include: increased access to information from taxpayer-funded systems; greater information sharing across organizations; and a readily available source of high-quality real-time data that encourages innovative applications and improved operational efficiency. In order to deliver these benefits, an effective open source policy must result in data being widely accessible and cost-effective while addressing the risks concerning security, privacy,

[1] See the ITS Strategic Plan at: http://www.its.dot.gov/strategicplan/index.htm.

liability, and data quality. Efforts and next steps are defined for each of these policy areas and are summarized at the end of Section I.

- **Section II: Research Data Exchange System Policies**

  Section II identifies the policies associated with the RDE. The section defines, identifies examples of, and proposes policy recommendations for: *system and data governance and management processes, operational practices and rules of conduct, security and privacy, standards, integration, and rules for data exchange and sharing*.

  There are a number of research actions/inputs that are needed to develop full policies in each area. Efforts and next steps are defined for each of these policy areas and are summarized at the end of Section II. Briefly, these include:
  - Establish a **program-level governance** team to develop policies and assign roles and responsibilities for the RDE-level governance.
  - Establish an **RDE-level governance** to implement policies.
  - Confirm **user class definitions** with stakeholders before establishing **user access policies**
  - **Authenticate users** with information that differs based on the different levels and uses of the RDE. Look to leverage the digital security certificates being implemented for connected vehicle security.
  - Develop **policies for use of standards and certification processes** with regard to systems or equipment that federate with/connect to the RDE.
  - Ensure implementation **of security and privacy policies** that are in line with Federal government (National Institutes of Standards and Technology, or NIST) policies.
  - Review each data set to determine how **privacy, data usage and sharing, and data storage and archiving policies may need to be tailored** to ensure optimal use and accessibility of each data set.
  - Develop **data sharing agreements and licenses** that are easily accessible and available through the RDE portal site.
  - Ensure that the **rules of conduct** include **attribution for contributions**.
  - Follow industry best practices for making the web site **accessible and easy to use.**
  - Develop **system availability and recovery policies and upgrade and maintenance policies** that are in line with the parameters typical of research systems but that also meet RDE user needs.
  - Establish federation criteria to ensure that the addition of new sites or new data sets are in compliance with the key policies of the RDE.

  Finally, to best understand how policies will apply and the impact of federation, Section II recommends the development of a set of scenarios that illustrate the types of institutions/systems that are likely to connect as part of a federated system.

- **Section III: Conclusion**

  At this time, there are two overall conclusions and one set of prospective analysis worth noting:

  - **Conclusion: Implement based on an open data policy.** An open data policy is a viable option and is encouraged by the U.S. Government in general and is emerging as a trend with other governments around the Nation and around the world. The

level of "openness" is highly dependent upon some of the technical inputs – the accessibility of the RDE to public users; the critical and minimum characteristics of the data that will be captured, used, stored, and archived; and the risks/trade-offs associated with the technical definition of what it means to be open. This paper and the related other Mobility policy reports (see list on the next page) attempt to put some definition to these open questions. There is a need to have the whole set of reports and definitions vetted by the technical team and stakeholders to ensure that the basis for recommending policies is solid.

- **Conclusion: The RDE system policies can be based on proven solutions; however the federation policies require further analysis and development**. The RDE architecture and set of technologies that are proposed for use in the construction and operation of the RDE appear synonymous with other portals in use with the Federal and State governments, academia, and industry. As a result, most of the RDE system policy can draw from existing models. The key differences, though, from a policy perspective include the wide-scale federation and the monitoring and enforcement of policies through such a dispersed system. Developing a set of optional models (also referred to as "scenarios") regarding various entities that might link with the RDE and reviewing their policies and analyzing the impact to the RDE supporting policies is a useful next step to determine how the technical, policy, and institutional recommendations might align (thus supporting broader federation) or face significant impacts that might challenge federation (for instance, a conflict between privacy or data usage policies).
- **Analysis: RDE Next Steps**. Even though the RDE is being implemented for research purposes, lessons can be learned regarding future operational data environments. Further analysis on technology transfer, steps and policies to support commercialization, and the viability of sustainable marketplaces will be needed.

Relationship to other Connected Vehicle Mobility Policy Reports

This report is one in a series of six policy reports that describe and analyze the policy issues associated with connected vehicle mobility. The series includes:

- Two foundational reports that identify the critical issues and describe the best practices and lessons learned from government, industry, and academia*:*
    - *Identification of Critical Policy Issues for the Mobility Program, FHWA-JPO-12-035*
    - *State-of-the-Practice and Lessons Learned on Implementing Open Data and Open Source Policies, FHWA-JPO-12-030*
- Four reports that analyze the specific policy issues in context of the goals of the DMA and DCM programs:
    - *Policy Analysis and Recommendations for the Open Source Applications Development Portal (OSADP), FHWA-JPO-12-031*
    - *Policy Analysis and Recommendations for Development of the Dynamic Mobility Applications, FHWA-JPO-12-033*
    - *Policy Analysis and Recommendations for the DCM Research Data Exchange (this report), FHWA-JPO-12-036*
    - *Privacy and Security Analysis and Recommendations for the DCM and DMA Programs, FHWA-JPO-12-032.*

Introduction

The vision of the U.S. Department of Transportation's (U.S. DOT) Data Capture and Management (DCM) program is to research, prototype, and demonstrate new methodologies for the active acquisition and systematic provision of integrated, multi-source data to enhance current operational practices and transform future surface transportation systems management. The goals of this program are to:

- Systematically capture real-time, multi-modal data from connected vehicles, devices, and infrastructure.
- Develop data environments that enable integration of high-quality data from multiple sources for transportation management and performance measures.

The end result of this research effort is to transfer specifications and lessons learned to other entities in the commercial market to build and operate new data environments.

There are three phases to the DCM program – development of an approach that is based on meeting user needs for transformative new technologies; development and demonstration of research prototypes; and transition of research findings to commercial adoption. This report examines and analyzes the policy and institutional issues that impact and/or facilitate each phase as a means of offering recommendations. The report is structured as follows:

- Section I of this report addresses the research questions associated with offering a new and transformative approach to data capture and management. At the basis of this approach is the implementation of an open data policy –the notion that public and private sources of data will be available and accessible and allow for ubiquitous transportation information to feed real-time applications. The section presents the opportunities and issues associated with an open data policy; the section also provides policy recommendations for moving forward.
- Section II of this report identifies the policies associated with implementation of new data capture and management technologies. The section defines, identifies examples of, and proposes policy recommendations for: *system and data governance and management processes, operational practices and rules of conduct, security and privacy, standards, integration, and rules for data exchange and sharing*.
- Section III of this report envisions successful research demonstration results and anticipates the types of policies that will facilitate successful commercial adoption.

Section 1: Open Data Policy

1.1: The Context – Intent and Purpose of an Open Data Policy

1.1.1: Definition of an Open Data Policy

An open data policy defines the objective of an organization for providing open data and defines what data is "open", at what level of detail, and the principles of user access (i.e., rights for using the data, limitations on use of data, and others). An open data policy requires supporting policy measures and procedures that address the risks and challenges to implementation.

With the emergence of the internet, there is significant movement around the world to adopt open data and open source policies. Examples include:

- The U.S. Federal Government and the recent Open Government Directive.[2] This directive required all Federal departments to develop a plan to release data. The U.S. Department of Transportation (US DOT) released its inventory and plan in March 2011.[3]
- The European Union and their Open Government Data initiative which has led to development of a Public Sector Information (PSI) platform. Together, these form an overarching policy foundation that is the modified by individual countries to develop their own implementations. Examples include the United Kingdom, Spain, or Denmark.[4]
- The Canadian government and the implementation of open data portals by many large cities in Canada.[5]
- The Open Knowledge Foundation and other non-profit initiatives.[6]
- Private sector and the opportunities associated with the increasing volume and detail of information captured by enterprises, the rise of multimedia, social media, and the Internet.[7]

The motivations and objectives vary:

- From a citizen's perspective, the movement stems from a desire to have greater access to data that is collected from systems paid for through taxes, as well as to have greater transparency of government performance and greater opportunity to participate in government decisions.
- From a government's perspective, an open data policy supports data re-use through the philosophy of "collect once, re-use many times". Clear, organized data collection also allows for sharing and the ability to support innovative uses of the data.

[2] http://www.whitehouse.gov/open/documents/open-government-directive
[3] http://regs.dot.gov/docs/DOT%20Draft%20Enforcement%20and%20Compliance%20Data%20Report%20-%2005-18-2011%20-%20OCR.pdf
[4] http://ec.europa.eu/information_society/policy/psi/index_en.htm
[5] http://www.data.gc.ca/
[6] For instance, http://okfn.org/about/vision/ or http://vimeo.com/okf
[7] A recent report attempts to value this opportunity by sector at: http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation

- From a private sector perspective, open data provides new business opportunities. Some opportunities include:
    - Data can be "mashed up" with other data to form new sources of data;
    - Raw data supports new business models wherein companies refine, clean/scrub, manage, archive, distribute, or analyze; or
    - Data translate into new applications or services.
- From a transportation perspective, an open data policy is a framework for transforming the state-of-the-practice. The policy supports implementation of "open access" systems that change the current paradigm, one of individual agencies/entities collecting, using, and archiving their own data. An open data policy supports re-use of data through cooperative, dynamic sharing/exchange. In this respect, it can also be seen to reduce costs.

1.1.2: Range of Open Data Policy Models

A connected transportation environment is premised on the notion of easily accessible and available public and private sources of data that provide ubiquitous transportation information to feed real-time applications. In reviewing the policy basis for data acquisition and use across a number of models, two policy options form the bounding cases for policy choices:

- **Private-Sector, Market-Driven Policy:** Adopting a market-based policy anticipates that consumer demand and the purchase/use of data-intensive technologies will generate widespread and easily accessible data *where and when it is needed*. This model is similar to today's arrangements for access to data and anticipates:
    - The availability and quality of data sets are based on demand and have value that leads to a purchase agreement. This is similar to the data that is made available through today's market forces.
    - That the majority of the collection and distribution mechanisms remain proprietary. Data that is captured is released based on agreements only.
    - That the mechanism that creates seamless access to data sources is a set of agreements by the organizations that capture and distribute/offer access to the data. Similar to today's market, these agreements are an efficient mechanism for identifying where and when data has value, thus generating revenue streams that support ongoing investment in data capture, refinement, and distribution. Also, such agreements directly address risk and liability.
    - While the market is efficient in meeting explicit demand, it can result in fragmented access to certain types of data (assuming that the one organization does not collect it or have agreements with other organizations for access) or overlook other types of data that are not as obviously in demand.
- **A Fully Open Data Policy**: Open data generally refers to "…data [that is] free to use, reuse, and redistribute it — subject only, at most, to the requirement to attribute and share-alike."[8] It is predicated upon the existence of systems that allow for "…free and unrestricted online

[8] http://opendefinition.org/

availability, [or] open access.[9] Data is not considered open if it requires additional permission or payment for its reuse[10] or users do not have complete freedom when it comes to customization and extension of their solution and/or must rely upon one supplier for any changes.[11] Three perspectives on this model include:

- From a government perspective, open data refers to the "…principle objective that information produced or commissioned by government…should be made available for free use, re-use, and redistribution by anyone."[12] However, this policy is in conflict with many existing federal policies (both in the U.S. and other countries) that emphasize the importance of privacy protection rules and limitations on collection and reuse of data. Additionally, governments have limited budgets and a fully open data policy does not account for a way to sustainably finance operations and maintenance of systems.
- From a research perspective, it refers to "…data from a project that is released rapidly into the public domain, subject to certain conditions, including a requirement that data users not exercise their intellectual property rights in a way that would preclude other users' access to the basic data." [13] While promoting open knowledge and dynamic exchange, a fully open data policy can be a barrier to capturing and retaining intellectual property, particularly in the publication of new findings.
- From a private sector perspective, this model presents challenges for revenue generation which, as noted previously, is important to financially support the ongoing operations and maintenance of new technologies.

In seeking to gain an appropriate balance, the EU has established a set of policies that apply to a more limited and defined set of data. They term their policy public sector information (PSI), which they describe as open when appropriate[14]. Similar to the U.S., a critical goal is information reuse. Four key elements for the basis for implementing the EU policy:

- An "open by default' rule for all 'public documents' [inclusive of datasets] which will mean that they "can be re-used for any purpose, commercial or non-commercial".
- Inclusive of information from libraries, museums and archives.
- Machine readable formats and metadata[15]
- Where charges exist, they will be capped at "…marginal costs incurred for their reproduction and dissemination". In principal, the marginal cost of reproducing digital information on digital networks do tend towards zero. In practice, "…*most data will be offered for free or virtually for free, unless duly justified*".

[9] Budapest Open Access Initiative at: http://www.soros.org/openaccess/read
[10] http://www.isitopendata.org/guide/
[11] http://ckan.org/
[12] See e.g., http://opengovernmentdata.org/what/ or http://gov.opendata.at/site/history.
[13] See Robin Feldman and Kris Nelson, "Open source, open access, and open transfer: market approach to research bottlenecks" available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1127571 (May 2008).
[14] http://www.techsoupglobal.org/blog/what-do-you-think-when-you-hear-%E2%80%9Copen-data
[15] The EU referendum defines this as: "…digital documents are sufficiently structured for software applications to identify reliably individual statements of fact and their internal structure."

The application of an open data policy for the connected vehicle environment is expected to be a hybrid of these two extreme positions. A current example to draw from is the U.S.-based public-private partnership that combines publicly -sourced data from multiple agencies of weather and road conditions. This model integrates fixed (roadside) and mobile environmental sensor station (RWIS) data from 48 public agencies (State and local DOT's and Canadian provinces) and makes it available for free to the community at large. This includes private organizations, as well as any other interested parties, who capture the feed and refine the data as a means of providing services and tailored information products such as weather-based decision support systems. Known as the *Clarus* initiative, key elements of successful implementation include the definition and provision of a system for data capture; quality checking of the data; standards; a public-private model; and supporting policies and procedures.

1.1.3: Why Implement an Open Data Policy for Connected Vehicle Mobility?

The objectives associated with implementing various forms of open data policies are similar to the objectives of the connected vehicle mobility program, in particular, the direction articulated in the white paper, ***Data Capture and Management Program: Transforming the Federal Role***[16]. As one journalist wrote in the United Kingdom's Guardian newspaper, "Open data is the new gold, the fertile soil out of which a new generation of applications and services will grow. In a networked age, we all depend on data, and opening it up is the best way to realise its value, to maximise its potential."[17]

In launching its Public Sector Information policies, the European Commission's estimates "the direct PSI-related market would be around EUR 32 billion in 2010". He also estimates that each year, within the European area, "overall economic gains from opening up PSI and providing easy access for free or marginal cost of distribution could be up to EUR 40 billion".[18]

In the U.S., a recent report released by the McKinsey Global Institute notes the following:
Big data can generate value in each. For example, a retailer using big data to the full could increase its operating margin by more than 60 percent. Harnessing big data in the public sector has enormous potential, too. If US healthcare were to use big data creatively and effectively to drive efficiency and quality, the sector could create more than $300 billion in value every year. Two-thirds of that would be in the form of reducing US healthcare expenditure by about 8 percent. In the developed economies of Europe, government administrators could save more than €100 billion ($149 billion) in operational efficiency improvements alone by using big data, not including using big data to reduce fraud and errors and boost the collection of tax revenues. And users of services enabled by personal-location data could capture $600 billion in consumer surplus.[19]

The question remains, what is the appropriate level of open? At which level of openness are the goals of "accessible", "available", and, "reuse" optimally balanced against the opportunities for commercialization and market-realization of value? At what level of openness can protection of privacy and security be realized at acceptable levels?
These questions are highly dependent upon a number of factors:

[16] Located at: http://www.its.dot.gov/data_capture/datacapture_management_federalrole7.htm
[17] As described by http://www.guardian.co.uk/world/datablog/2011/dec/13/eu-open-government-data
[18] See Graham Vickery's analysis at:
http://ec.europa.eu/information_society/policy/psi/docs/pdfs/report/final_version_study_psi.docx
[19] http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation

- The risks associated with providing data in an open fashion;
- The costs of providing open data and the decisions regarding who will bear the costs/expenses; and
- Acceptance by citizen's that the data generated by them can be used for the public good.

Sections 1.2 and 1.3 will describe risks and identify options to mitigate the risks. Section 1.4 will return to these questions as a basis for providing recommendations, identifying knowledge gaps, and proposing a set of next steps.

1.2: Data Policies and Management: Risks, Opportunities, Mitigation Options, and Recommendations

Successful implementation of an open data policy will bring risks that will need to be addressed as a means of gaining acceptance. It will also offer opportunities that may need to be balanced against the risks.

Importantly, the risks and the decisions regarding the form of mitigation are typically aligned with ownership of the systems and technologies. If, for instance, the open data policy is applied to Federal data portals (as in the case of the RDE), federal policies provide the guidelines for assessing risks and choosing options; private sector assessment and decision making may follow similar paths, but may also customize the mitigation choices to maximize revenue potential.

Section 1.2 provides a description of the key risks, lists the options for risk management and mitigation, and provides policy (and, sometimes, technology) recommendations. The sections include:

1.2.1 Security and Privacy
1.2.2 Data Quality and Liability
1.2.3 Breaches of Data
1.2.4 Cooperative, Multi-Sourced and Fused Data and Liability
1.2.5 Data Ownership
1.2.6 Intellectual Property and Liability

After these descriptions, the opportunities and impacts are discussed from three owner/operator perspectives – Federal, private sector, and hybrid.

1.2.1 Security and Privacy

Addressing risks to security and privacy are critical actions that underpin successful implementation and acceptance of new technologies. A separate paper titled *Privacy and Security Analysis and Recommendations for the DCM and DMA Programs* offers a risk analysis of the RDE and the types of data that might be captured and distributed through the RDE. In summary, the risks fall into two categories:

- **Data risks** – the risk of exposing data with personally-identifiable information (PII) associated with the data:

- Preliminary analysis noted that the majority of transportation data presents a risk of geo-location information which, independently, is a low risk. The risk increases when the data can be compared against other databases that offer identifying information and allows the geo-location data to be matched to a person, a growing risk with smarter technologies.[20]
- Preliminary analysis also noted that certain public sector applications such as those associated with public safety/first responders or transit/ride-sharing may contain sensitive or confidential information (e.g., health data, incident data, or financial payment or account data).

- **System risks** – the risk of cyber-attacks and/or exposure of data (data breaches) through the lifecycle of collection, aggregation, distribution or sharing. Federation increases both of these types of risks as the links with other systems create a greater potential to insert malicious viruses or exploit technology vulnerabilities through the process. An example of the latter is a recent spate of discoveries regarding operating system holes or insertion of malware during manufacturing. [21] Another critical risk is a denial of service attack if misbehaving or malicious actors choose to shut the site down.

Mitigations options include:

- For security:
  - Careful decisions regarding operating systems including an analysis of the potential risks and vulnerabilities across varying systems. This includes consideration of how frequently patches are provided to address security vulnerabilities as they arise.
  - Scrutiny of other systems that link to the RDE and development of criteria regarding linking with other sites. Application of such criteria early in the demonstration phase will provide insights and lessons learned about the type of vulnerabilities and attacks that present the greatest risks. Further research that identifies the range of possibilities through case studies and expert brainstorming would be a useful tool for developing an auditing system that continuously monitors for a wide range of risks associated with external sites.
  - Encouragement of participants in the open data portal to look for and report anomalies.
  - Credentialing of the data and certification of the system so that users can trust the authenticity of the data and the security of the system.
  - Automated alerts regarding suspicious behavior.
  - Appropriate incident recovery plans and notifications to users whose data may have been breached.
- For privacy:
  - Anonymize the data or use other privacy enhancing technologies (PETs) that mask or de-identify the data in real-time (although these can be expensive)
  - Describe to users how privacy is protected and give users the opportunity to opt-out or opt-into the system (although the former choice may result in less data collection and, thus, potentially sub-optimal mobility applications. Studies are being performed

[20] iPhone keeps detailed log of its precise whereabouts, storing up to a year's worth of user location data at: http://www.latimes.com/business/la-fi-app-privacy-20120216,0,7863079.story
[21] http://www.latimes.com/business/technology/la-fi-tn-cyber-security-crowdstrike-20120223,0,4645028.story

on the mobility applications to determine what data and how much is needed for optimal functionality).

- Enact the Fair Information Principles Practices (FIPPs) that serve to implement privacy controls on the data.[22] FIPPs are a subset of security controls and they guide the owner/operator of a system through descriptions of the data, how it is used, and how it is protected, giving users an opportunity to fully understand how their privacy is protected. FIPPs are required for all Federal systems and followed by some private sector organizations. The Federal Trade Commission works with industry to formulate a set of privacy controls that seek to balance privacy interests with revenue opportunities.[23]

The report on privacy analysis, mentioned at the beginning of this section, provides greater detail on these risks and the options for mitigation. Notably, the ITS Legal Policy team is developing an overall privacy policy for the connected vehicle environment. This policy will inform and support the direction and development of FIPPs for the mobility elements of the connected vehicle environment. The overall privacy policy is expected by summer of 2012.

Security and Privacy Policy Recommendations: The policy regarding the creation of the RDE based on an open data policy starts with the recognition that some level of PII may be captured; policies and actions (particularly automated actions) need to be in place to ensure that such data is not posted to or made accessible within the data environments. Thus, the first policy recommendation is to develop appropriate security measures in light of the most likely risks in order to preserve privacy. The following application of policy and technology recommendations can assist in meeting these goals.

As the RDE is under Federal ownership and oversight with contracted operations, the RDE and the data are subject to Federal policies for security and privacy and for data stewardship and release. These policies include:

- **Security Requirements:** Implement the FISMA and NIST security guidelines[24]
- **Privacy Requirements:** Perform a privacy impact analysis and implement NIST privacy controls, including notice to drivers and travelers who generate the data regarding the intended uses of the data (including posting to a public website); [25] [26] additionally, ensuring that each data set is scrubbed for identifying characteristics such as origin-destination information.

[22] http://csrc.nist.gov/publications/drafts/800-53-Appdendix-J/IPDraft_800-53-privacy-appendix-J.pdf
[23] http://www.ftc.gov/bcp/bcppip.shtm
[24] FISMA is Title III of the E-Government Act (P.L. 107-347) and is located at: http://csrc.nist.gov/drivers/documents/FISMA-final.pdf.; key NIST guidelines for categorizing systems and data, selecting security controls, implementing security controls, assessing security controls and authorizing and monitoring the security state are located at: All of the following documents are located at: http://csrc.nist.gov/publications.
[25] NIST 800-122 and 800-53J at http://csrc.nist.gov/publications.
[26] As noted earlier, the level of acceptable security risk is currently being defined by the Implementation Policy and Legal Policy teams in support of defining organizational and operational model options for the connected vehicle security system. This baseline will be a significant input to the Mobility teams in understanding what level of security needed for the RDE and the applications. This analysis of the baseline will be available in Summer 2012.

A decision as recent as February 2012 resulted in the choice of cloud computing as the architectural basis of the research data exchange; with this decision, a set of recent policies from NIST will offer guidance to the RDE developers on properly analyzing and applying security and privacy for the RDE (NIST *Guidelines on Security and Privacy in Public Cloud Computing*, 800-144).[27] These guidelines present recommendations that organizations should consider when outsourcing data, applications, and infrastructure to a public cloud environment. It provides insights on threats, technology risks and safeguards related to public cloud environments. As noted on the website, the guidelines include assistance for:

- Carefully planning the security and privacy aspects of cloud computing solutions before implementing them.
- Understanding the public cloud computing environment offered by the cloud provider.
- Ensuring that a cloud computing solution—both cloud resources and cloud-based applications—satisfy organizational security and privacy requirements.
- Maintaining accountability over the privacy and security of data and applications implemented and deployed in public cloud computing environments.

And finally, the data and data release are governed by an Office of Management and Budget (OMB) memorandum (M-10-06) which instructs agencies to "…*increase accountability, promote informed participation by the public, and create economic opportunity by taking prompt steps to expand access to information; making information available online in open formats, and presuming openness to the extent permitted by law and subject to valid privacy, confidentiality, security, and other restrictions*." Application of this policy is defined by a DOT Order 1351.34[28] which applies to information that DOT generates as well as information that other parties provide to the DOT if the other parties seek to have the DOT rely upon or disseminate the information or the DOT decides to do so on its own.

The policy seeks to make DOT data available at the most detailed level possible, subject only to the limits imposed by data quality and the need for confidentiality. The policy requires that data be protected from unauthorized access, corruption, or revision as well as data must be accessible and comply with the Departmental web policy. This policy aligns well with the NIST guidelines, but further requires that data conform to the general standards of quality as established by the DOT's Bureau of Transportation Statistics (BTS) and OMB.[29]

Security and Privacy Technology Recommendations:

- If the connected vehicle system implements broad requirements for connected vehicle data to be credentialed, this technology measure can go a long way toward security and privacy. If there is a choice and mobility applications generally do not need credentials, those applications with sensitive or confidential data may consider using credentials.
- An effective technology for the data environments is the automated de-identification of data, especially if data from other sites will pass through or comingle with connected vehicle data in the Federally-sponsored data environments. The RDE team should analyze the costs versus

[27] http://www.nist.gov/itl/csd/cloud-012412.cfm

[28] http://regs.dot.gov/docs/DOT%20Draft%20Enforcement%20and%20Compliance%20Data%20Report%20-%2005-18-2011%20-%20OCR.pdf

[29] http://www.bts.gov/programs/statistical_policy_and_research/data_quality_guidelines/ Note that the OMB and BTS guidelines apply specifically to statistics and to information presentation, and are less specific regarding data quality.

effectiveness of these technologies if planning to host the data on a DOT-based cloud or servers (or contractor's cloud or servers).

- Finally, given DOT's requirement for accountability, the RDE should implement data monitoring and alert technologies that provide the operators with an ability to pull back or redact/de-identify data in real-time. It should also employ sampling methodologies to ensure that the data do not contain PII other than geo-location information.

1.2.2 Liability

Elements of liability that are critical to successful RDE implementation and must be addressed include liability due to:

- Defective data or data errors (section 1.2.2.1). Poor data quality results in applications working improperly and potentially putting the user at risk.
- Breaches of data (section 1.2.2.2).
- Use of cooperative, multi-sourced and fused data (section 1.2.2.3).
- Data Ownership (section 1.2.2.4)
- Improper handling of intellectual property (section 1.2.5).

1.2.2.1 Data Quality Liability and Mitigation Options

Data quality assurance is the process of profiling data to discover inconsistencies and other anomalies, and performing data cleansing activities (e.g. removing outliers, missing data interpolation) to improve data quality. These activities can be undertaken as part of data warehousing or as part of the Database administration of an existing piece of applications software. Data quality, however, can be difficult to define outside of a particular system or application. It is a relative term that depends upon the purposes for using the data and the system/technologies needs for a certain level of quality. It is also a multi-dimensional concept.[30]

Standard metrics that apply to data quality tend to include: completeness, accuracy, consistency (formats), and relevancy. A recent framework, illustrated in Table 1, was developed for a study by researchers associated with the Massachusetts Institute of Technology and provides an overview of some of the key dimensions, metrics, and improvement opportunities.[31] The study also reviews different data quality tools available on the market and offers insights into their capabilities.

Dimensions	Definitions	Some Metrics	Some Improvement Opportunities
Completeness	Is a concept missing? Are there missing values in a column, in a table? Are there missing values w.r.t. a reference population?	• Rate of missing values	• Crosschecking or external data acquisition • Imputation with statistical models • Statistical smoothing techniques
Accuracy	Closeness between a value *v* and a value *v'* considered as the correct representation of the reality that *v* aims to portray	• Number of incidents or malfunctions • Comparison with reality	• Analysis of consistency and likelihood of controls • Meta-data: degree of reliability

[30] From "Data Quality Assessment Methodology: A Framework" by Burns, Eugene and Purificacion O. MacDonald and Amrut Champaneri. BTS, US DOT.

[31] http://mitiq.mit.edu/iciq/pdf/an%20evaluation%20framework%20for%20data%20quality%20tools.pdf and http://www.dataqualitypro.com/?page=etl_data_quality

Consistency	Data are consistent if they respect a set of constraints	• Computation of discrimination power for controls	• Definition of a control strategy • Comparison with another apparently more reliable source
Relevancy	Is the data useful for the task at hand?	• Degree of utility	• Survey (helps to improve relevancy of planned tasks for maintaining and improving data quality

Other frameworks offer other dimensions as well including timeliness (a critical standard for some connected vehicle applications) and comparability. This latter dimension may be critical to cross-border interoperability.

Liability/Data Quality Policy Recommendations: DCM / DMA data system and data set "owners" will find the Bureau of Transportation Statistics guidelines to be a recommended foundation for the development of data quality assurance protocols. These guidelines form the basis for DOT's implementation of OMB's Information Dissemination Quality Guidelines.

Of interest in the BTS guidelines is the examination of other systems that gather and fuse data from multiple sources and the issues related to liability. For instance, the BTS guidelines use FMCSA's SafeStat to illustrate that the system receives State-generated data and makes use of it to generate important information on the safety of motor carriers. If data is deemed inaccurate, FMCSA cannot correct this dataset but must work with the State. The role for FMCSA is to monitor and report data quality issues and to work with its data-partners to raise the level of quality. Data correction guidelines have been useful in this process; so to have been automated opportunities for partners to correct and upload better quality data and/or to place a hold on disputed data. It is not clear if any such measures will be necessary with the RDE in a federated form – it depends on the use of the RDE and the type of applications it serves. However, these guidelines can help the RDE maintain quality data.

Additionally, FHWA has defined guidelines for data quality measurement and propose quality standards. These guidelines currently apply to existing ITS data from probes, signals, loop detectors, and other technologies, and are recommended as a baseline.[32]

Last, the US DOT has established guidelines for moving transportation datasets to Data.gov. The following guidelines are recommended for the RDE:

- Data Formats: XML is preferred but other formats are acceptable as long as they are structured in machine-readable form. Other formats include: RDF, CSV, TXT, KML, KMZ, XLS, XLSB, ESRI Shapefile, ATOM, RSS, or CAP.
- Data Tools: the DOT desires that when offering datasets for download, they should be classified based on delivery model:
- Data Mining/Extraction Tool – may be a database access facility, web mapping, or data visualization application. It may also be a web page that is delivered using file compression formats such as ZIP, GZIP, and TAR. Feeds should be in XML formats including ATOM, RSS, and CAP.

[32] The guidelines can be found at: http://ntl.bts.gov/lib/jpodocs/repts_te/14058_files/chap4.htm.

- Widgets/Gadgets are documented, shareable APIs and portable, standalone, embeddable data-access applets.
- Tools that explicitly restrict the data to less-than-full public use or are incompatible with Data.gov are not to be considered.

These requirements translate into an action for the mobility team to determine if and how the RDE data will be prioritized for inclusion on Data.gov. A review of the prioritization criteria suggests that the RDE will need to be compatible.[33]

Liability/Data Technology Recommendations: Other state-of-the-practice examples offer insights into different approaches to data quality assurance. The Center for AIDS Research Network of Integrated Clinical Systems[34] and the National Data Buoy Center[35] stand out as having a particularly rigorous quality processes in place that make creative use of automated techniques for real-time data quality verification. It is likely that the DCM / DMA programs will need to go well beyond even this level of data quality assurance, to include formalized protocols for data review, error documentation, and error correction.

An important tool for limiting liability is the posting of disclaimers regarding the known issues with the data and to transfer the liability to the user. This standard practice is used throughout the technology, applications, and portal worlds. Essentially, the terms of use and/or disclaimers are provided one a website (portal) and before users can gain access to the data, they must consent to the terms.
When developing appropriate terms of use or disclaimers, US DOT's legal counsel will need to be involved to ensure that the language is appropriate for Federal agency use and to ensure that, from a legal perspective, the risk is appropriately described and transferred to the user.

1.2.2.2 Breaches of Data Liability and Mitigation Options

Breaches of data were discussed earlier in the section on security. However, many States define a breach in data to result in liability only if the breach resulted in harm – physical, economic, financial, social, or other – from the breach. [36]

Policy and Technology Recommendations: Developing an ongoing threat analysis capability for the RDE is one of the more effective mechanisms for mitigating breaches. In addition, the RDE will implement security based on the FISMA and NIST guidelines. A useful, next step in research is to determine whether there are certain types of data – potentially more confidential and/or sensitive than other data types – and segment and wall-off these data from real-time exchange. Download can be accomplished through a password authentication. This is only possible if the data is not needed in real-time.

1.2.2.3 Cooperative, Multi-Source Data Liability and Mitigation Options

In today's world, an increasing number of applications are using multi-sourced, fused data. By combining data in this fashion, it is becoming increasingly more difficult to identify which data may

[33] http://regs.dot.gov/docs/DOT%20Draft%20Enforcement%20and%20Compliance%20Data%20Report%20-%2005-18-2011%20-%20OCR.pdf
[34] http://www.cnics.net/
[35] http://www.ndbc.noaa.gov/
[36] Compendium of State Privacy and Security Legislation: 2002 Overview, November 2003, NCH 20030. U.S. Department of Justice, Office of Justice Programs, Bureau of Justice Statistics.

have contributed to an incident or poor decision or, further, to assign liability based on the source or quality of the data. The connected vehicle environment will face this problem.

When such problems arise, there is a well-established torte law system in the U.S. that follows specific procedures to identify the source of the problem. Investigations center on the owner and operator of the system or provider of the source of data to question whether negligence or misbehavior played any role. If not, courts have recognized the difficulties in discerning fault due to multi-sourced data flows and are still establishing precedence in this area.

Multi-Source Data Liability Policy and Technology Recommendations: In terms of the adoption of new connected vehicle transportation technologies, the U.S. DOT's legal policy analysis team is currently reviewing these issues and will provide policy recommendations in Summer 2012. Until that time, a key recommendation is to implement the best practice of using disclaimers and Terms of Use (TOUs) on websites so that those who are employing the data sources are aware of the potential issues (i.e., data quality, origin/source, etc.) and are accepting the liability for use. While such disclaimers are used frequently throughout government and industry, it should be noted that they do not waive liability in the presence of negligence of the data provider.

1.2.2.4 Data Ownership

Another key risk for the RDE, especially with the implementation of an open data policy, is that of data ownership. The important points to note are the following:

- For data, the appropriate legal construct is the copyright. Importantly, there is, to date, no copyright on raw data.
- External databases, however, come with copyright and must be appropriately licensed for use. There are two considerations for the DCM program:
    - If/when using other databases, appropriate licenses and other paperwork must accompany the external databases to specify use and compensation (if any).
    - If/when offering the DCM databases to others for use, a license with terms of agreement will be necessary.
- The ownership issue is complicated and there is no clear precedence on how to define ownership. The textbox on the next page describes a range of potential ways that ownership of data might be claimed. Recently, a new philosophy has emerged regarding the right of the people whose actions produce data to own the data.[37] This movement is known as the "New Deal on Data" and is gaining momentum at the international level. This approach to data ownership is very different from the current private sector practice that "compensates" companies for the investment in the infrastructure and capture/ management of the data by providing them with data on the user to resell for value. This form of compensation requires that consumers consent to provide them with the data that is generated by their devices, typically in exchange for a service or use of an application. With this ownership, however, comes the responsibility of stewardship of the data to ensure its integrity and its appropriate use.[38]

[37] http://www.futuresalon.org/2010/03/3-questions-for-future-salon-speaker-mit-professor-sandy-pentland.html and http://hd.media.mit.edu/wef_globalit.pdf.
[38] http://ori.dhhs.gov/education/products/n_illinois_u/datamanagement/dotopic.html.

- Finally, much of the literature on data ownership reflects a more typical need by businesses to identify who within the firm "owns" the data and the stewardship and rights associated with each data set. Appendix C offers an example of how to create a data ownership policy within an enterprise organization. This template may be modified to fit the needs of the RDE.

Range of Options to Identify Data Ownership

- **Creator –** The party that creates or generates data.
- **Consumer –** The party that uses the data owns the data.
- **Compiler -** The entity that selects and compiles information from different information sources and owns the copyright/intellectual property associated with the database and organization of the data.
- **Enterprise -** All data that enters the enterprise or is created within the enterprise is completely owned by the enterprise.
- **Funder - T**he user that commissions the data creation claims ownership.
- **Decoder -** In environments where information is "locked" inside particular encoded formats, the party that can unlock the information becomes an owner of that information.
- **Packager -** The party that collects information for a particular use and adds value through formatting/refining the information for a particular market or set of consumers.
- **Reader as owner -** The value of any data that can be read is subsumed by the reader and, therefore, the reader gains value through adding that information to an information repository.
- **Subject as owner -** The subject of the data claims ownership of that data, mostly in reaction to another party claiming ownership of the same data.
- **Purchaser/Licenser as Owner –** The individual or organization that buys or licenses data may stake a claim to ownership.

Reference: D. Loshin, Knowledge Integrity: Data Ownership (2002 or online, June 8, 2004 at: http://www.datawarehouse.com/article/?articleid=3052)

Data Ownership Policy and Technology Recommendations: The issue of data ownership is being reviewed by the ITS Legal Policy team who will provide policy recommendations in summer of 2012. Until that time, a key recommendation is to implement the best practice of using Terms of Use (TOUs) agreements and disclaimers on the RDE website so that those who are employing the data sources are aware of the potential issues (i.e., data quality, origin/source, etc.) and are accepting the liability for use. While such disclaimers are used frequently throughout government and industry, it should be noted that they do not waive liability in the presence of negligence of the data provider or manager. In terms of developing the proper language, the U.S. DOT has many examples of such statements; ultimately, the appropriate language will be set and approved by the U.S. DOT legal counsel.

1.2.2.5 Intellectual Property

Closely related to data ownership is the critical risk of intellectual property infringement. There are two issues for attention for the DCM program, as described below. These issues include attention to database copyright and licensing; and licensing of new data capture, management, and exchange technologies.

- **Database copyright and licensing**

  Although raw data is not copyrightable, the development of databases and the intellectual property that goes into the structure and algorithms associated with databases are covered by copyright and require proper licensing to disclose attribution and, if relevant, terms of use. There are three instances when database/dataset licensing is an issue for the DCM program: (1) when the RDE provides new datasets, particularly if it provides the datasets as "open data", an associated license must accompany the data to alert the user to the agreed upon uses, the attribution of the developer's intellectual property, and any disclaimer of liability; (2) when the DCM program acquires new datasets for populating the RDE, those datasets should have clear license terms of use and attributions and the RDE will need to accommodate those terms of use; and (3) when the RDE links with external entities, the RDE design must accommodate the ability for users to recognize and access license terms associated with that entity's datasets.

  With regard to use of other datasets, the RDE data manager will need to have responsibility for ensuring that licenses associated with external datasets are recognized and properly followed.

  With regard to licensing of new DCM data from the RDE, there are a number of licenses that are considered when implementing an open data policy:

  - **Open Commons' Public Domain Dedication and License (PDDL)[39]:** The PDDL places the data(base) in the public domain (waiving all rights). It has been noted by some that there are some reference to European Data Rights in the PDDL but these have no meaning outside of the EU and not considered to be an obstacle. The PDDL may be the most open and forward looking license.
  - **Open Commons' Open Database License (ODC-ODbL)[40]:** A license that provides attribution and institutes share-alike agreements for data and databases. It allows users to copy, distribute and use a database; produce works from the database; and modify, transform, and build upon the database provided that the user attributes use of the database or works produced from the database in the manner specified by the ODbL. Users must also make clear to others the license of the database and keep intact any notices on the original database. While more comprehensive than the PDDL, this license begins to require numerous attributions. Also businesses may find the share-alike requirement to be restrictive in terms of their business models.
  - **Open Commons' Attribution License (ODC-By)[41]:** This license is similar to the ODC-ODbL except that it does not include the share-alike provision. Creative Commons considers this the most accommodating license that it offers.
  - **GeoGratis:** Is a license provided by Natural Resources Canada and allows free and open use of geo-spatial public sector information. The license grants users a non-exclusive, fully paid, royalty-free right and license to exercise all intellectual property rights in the data. This includes the right to use, incorporate, sublicense (with further right of sublicensing), modify, improve, further develop, and distribute the data; and to manufacture and/or distribute Derivative Products. The one requirement is that the

[39] http://opendatacommons.org/licenses/pddl/
[40] http://opendatacommons.org/licenses/odbl/
[41] http://opendatacommons.org/licenses/by/

Licensee shall identify the source of the Data, in the following manner, where any of the Data are redistributed, or contained within Derivative Products: "© Department of Natural Resources Canada. All rights reserved."[42]

The UK has its own version of such a license known as the **UK's Open Government License.**[43] This license governs the use and re-use of public sector information in both government and public sector use. It is based on a framework of guiding principles regarding licensing:

- Simplicity of expression - the terms should be expressed in such a way that everyone can understand them easily;
- Non-exclusivity - so that access can be provided to a range of users on fair and equal terms;
- Fairness of terms;
- Non-discrimination - terms are extended fairly to all for similar uses;
- The need for acknowledgment and attribution;
- The need for transparency by publishing standard license terms

The UK's standard approach to licensing covers:

- Free use and re-use for all purposes, both commercial and non-commercial - the **Open Government License**; and
- Free use and re-use for non-commercial purposes only - the **Non-Commercial Government License**.

An important note from a legal perspective is that these open source licenses are being tested in court cases and are being upheld. Two recent cases (as of 2009) that offer insight into how the courts see these licenses is the Jacobsen v. Katzer case (535 F.3d 1373 Fed. Cir. 2008) which upheld the conditions set by the attribution clauses in the license. Another set of cases was brought by BusyBox in US District Court against a number of redistributors who did not offer free access to the source code, as stipulated in the license. As these cases moved forward within the courts, the redistributors decided to settle out of court by providing users with access to the source code.[44]

IP Policy Recommendations: The easiest and most advantageous path is to align with several of the standardized open data licenses that already exist. The reasons include: (1) the licenses are well understood; (2) the licenses are stable (because these licenses are managed by independent authorities and many people use them, they evolve cautiously, and balance the interest of consumers and sharers of data or information); and (3) these licenses balance interests responsibly. The creators of these licenses have thought through all the issues that pertain to open data and thus provide assurance to both consumers of data and distributors of data in knowing that they have a license that will work.[45]

[42] http://geogratis.cgdi.gc.ca/geogratis/en/index.html;jsessionid=165DEA5D04EF1F09BD6F9A8319DEE702
[43] www.nationalarchives.gov.uk/information-management/government-licensing/the-framework.htm
[44] http://www.itechlaw.org/ebulletin/volume.asp?id=11&keyword=multi-source+data+liability&author_firstname=&author_lastname=#86
[45] As noted by an author of a Canadian license at: http://eaves.ca/2011/02/16/the-state-of-open-data-in-canada-the-year-of-the-license/

Additionally, the DMA program's desired approach is to require attribution; thus the program seeks licenses that ensure this particular feature.

To determine which license will work best for the RDE, the legal policy team will need to analyze the various licenses in terms of US DOT requirements.

- **Licensing of data capture, management, and exchange technologies**

  A second critical intellectual property issue focuses on the licenses and agreements associated with the technologies used with data capture, management, and exchange –both those procured for developing the RDE and those developed (or enhanced) through the prototyping of the RDE. The U.S. DOT is familiar with procuring technologies for use and has guidelines on which types of licensing agreements can be accepted. For the most part, it is expected that the procurement of existing technologies for building the RDE will be done through a contractor who will take on the responsibility for assuring proper licensing. The challenges occur with the development of new technologies or enhancements of existing technologies:

  - **Licensing of new technologies:** The DCM program has an interest in ensuring that all new technologies are released as "free and open source" so that public sector agencies, industry, and academia can all benefit. This situation is similar to the DMA program's interest in offering new software as free and open source. The U.S. DOT has this option by way of providing an appropriate license. If released as open source, the license must stipulate the terms of use, including any downstream enhancements. The license options available for releasing new technologies and applications as open source include a range of highly restrictive (no enhancements or distribution may include proprietary intellectual property) to those that are less restrictive (enhancements or modifications may be considered proprietary and available for capturing the value of the additional intellectual property). In both situations, these licenses typically require that appropriate attribution be included for the original intellectual property.

    A separate white paper that explores the open source release practices that are generally in use and will be available through the Mobility program in April of 2012. This new white paper summarizes the license options which are more fully documented in the OSADP paper along with an analysis of trade-offs among the license types.

  - **Enhancement of existing technologies:** In this situation, it is the U.S. DOT and its contractors that must receive a license from the existing IP owner to enhance or modify the technologies. Once the efforts are completed, the U.S. DOT has two options for release:

    - Transfer new intellectual property back to industry with the appropriate licensing and guidance. With existing IP, the Federal government may not be able to transfer the new enhancements as free and open source, depending on the terms of the "inbound" licenses.
    - Negotiate with the owner of the existing IP to license it to the Federal government with rights for free and open source distribution. This may involve an upfront payment in lieu of future customer purchases.

The textbox below provides an analytical framework for identifying license rights and identifies the various options for procuring RDE design and development services. It is based on a framework developed by NASCIO to guide State CIOs in working with their legal counsel.

Analysis Framework for License Rights*

In determining IP rights, states are urged to examine the particular requirements of the contract because, in many cases, that will determine the appropriate approach to IP. The following examples may assist in this analysis:

Procurement of Commercial Software and Support Services: Commercial off-the-shelf software (COTS) is virtually always subject to standardized licensing agreements. While, in certain instances, terms of the license may be negotiated, most developers/contractors do not anticipate divesting themselves of ownership of COTS software enhancements or derivative works of such software. Contractors maintain ownership over deliverables related to the maintenance, installation and configuration of COTS software.

Procurement of Standardized IT Services (such as Hosting or Disaster Recovery Services): These offerings typically do not pose difficult IP issues as appropriate use rights are stipulated through the licensing of IP embedded in the service.

Procurement of Consulting Services Involving Customized Deliverables: In this instance, the Federal government may legitimately require ownership of certain deliverables; however, the Federal government must determine the structure of licensing to entities, including whether it has any compelling need to exclude or restrict particular uses of the customized deliverable.

Procurement of Systems Integration Services: This is the most complex area. A systems integration contract may involve COTS software and ancillary services, custom deliverables produced in accordance with specific state requirements, and deliverables that combine newly created IP with pre-existing IP. Standardized IP clauses may be inadequate for this situation, and, in contracting for the design of the RDE, the Federal government will need to consider implementing a clause based on the categories of ownership described above in which particular types of IP can be designated as being licensed to the Federal government, owned by the Federal government (with, as appropriate, a license back to the contractor) or jointly owned.**

* Source: NASCIO Negotiating IP on the Way to the Win-Win: NASCIO's Intellectual Property Recommendations at: http://www.nascio.org/publications/documents/NASCIO-negotiatingIP.pdf.

** Note: Such clauses have been implemented, giving the U.S.DOT "unlimited rights" to the software, source code, and data associated with the RDE and allowing for modification and distribution without restriction.

Licensing Policy Recommendations: The primary recommendation is to review the contracts for design and development services and review/analyze the RDE technologies for existing intellectual property rights/licenses. These existing licenses will inform how free or restricted the U.S. DOT and its contractors will be in enhancing or modifying existing technologies, and the type of license and distribution arrangements that are available when transferring to the market or public sector partners. If new intellectual property is created in the design/development of the RDE, the U.S. DOT will need to have clear language with the contractor to determine who will ultimately have ownership of this IP. To do this analysis, final decisions on the RDE – its architecture and technologies/software/hardware—is needed. Also, once these issues are identified, the Legal ITS Policy team will be able to support analysis and guidance for the Program-level governance team.

1.3 Federation Policies and Access Criteria

A federated system of data environments offers an opportunity to leverage the content of other databases as well as to distribute the responsibilities and costs. According to McLeod and Heimbigner, authors who first defined the concept, *a federated database system is one which "define[s] the architecture and interconnect[s] databases that minimize central authority yet support partial sharing and coordination among database systems".*[46]

Moving from a stand-alone and/or centralized system, however, creates new or increases certain risks. With a federated system, the risks associated with privacy and security increase without the ability to control policies of the other systems. A key element in developing policy in support of federated systems is the ability to define and delineate responsibility of the parties in the event of failures and other problems. Automated auditing to look for anomalies and measure performance becomes a more critical function; and deciding who can have access and how that access is obtained is also a critical policy and function.

In an analysis of the benefits of federated versus centralized systems for the Center for Advanced Defense Studies, the authors analyzed seven metrics to understand the advantages and risks of federated systems:[47]

- **Autonomy:** According to Howard and Kanareykin,"By agreeing to participate in a federated information sharing system, each node agrees to share a certain amount of its own information, based on some form of agreement." The key to successful federation is the sharing policy and how it is enforced. Importantly, in a federated system, each node has the ability to establish their sharing policy and update it as needed. Sharing policies can be different for different partners.
- **Security:** The potential to gain improper access through a federated system is higher than with a centralized or stand-alone system, particularly an increase in risk for unauthorized data mining and (potentially) identify theft or tracking. Typically, in a federated system, each node is responsible for implementing and enforcing its own security policies, potentially at different levels of security. Also, there arises the question of security for the merged data and who is responsible.
- **Performance and scalability:** With a federated system, each query (or request) is satisfied through multiple systems providing multiple results and integrating them together. In some instances, this process may slow down performance; especially if there are large data sets from which to draw. Howard and Kanareykin note that better performance in a federated system is based on "developing cost-based optimizer[s] that is aware of the distribution and heterogeneity of the back-end servers."[48] The authors also note, however, that an advantage to federated systems is that they tend to be updated more frequently and thus function on more relevant data.
- **Usability:** The ability of the user to navigate the information sharing network is a key element of success. Federated systems run the risk of providing users with information

[46] "McLeod and Heimbigner (1985). "A Federated architecture for information management". *ACM Transactions on Information Systems, Volume 3, Issue 3*. pp. 253–278.
[47] The list of performance metrics is synopsized from Howard, N. and S. Kanareykin, Center for Advanced Defense Studies, 2007. Located at: www.c4ads.org/sites/default/files/federated vs. centralized.pdf
[48] Also, from Winter, R., 2002 "Playing Every Tune: Federated Databases." DB2 Magazine 1.

overload and/or potentially leading the user to too many other websites with different interfaces and search capabilities (which can be confusing). Federated systems are most effective when the user is presented with one interface which is capable of gathering and combining the results properly.

- **Advanced functionality:** The ability to link data and identify trends is an important function for data environments. The authors note that the "link and trend" capability may be more accurate in a federated system, given that local databases tend to be kept up to date.
- **Public perception:** The authors site examples of federated systems that were not well-accepted by the public. Most of these were in relation to law enforcement or crimes, but the underlying issue with public perception has to do with privacy. Federated systems increase the risk to privacy not only because of the (potentially) inadequate policies of one or more nodes, but also because other databases are now available to compare data trails and to discover identities or link data to specific users.
- **Operational costs:** A federated system is more likely to share costs and reuse assets and data, thereby (potentially) lowering costs. This reuse of data and assets is a key tenet in the DCM vision.

Another critical issue for federated systems is the criteria/policy for access with other entities. In many federated systems that are owned and operated by one entity (i.e., multi-national corporations or the U.S. military), the authority to set criteria is resident within one decision-making body and the ability to enforce the criteria (by granting or rescinding access) is similar. In the type of federated system envisioned by the DCM program, many nodes will have their own access policies for their sites. The DCM team will need to determine how access is best granted in the federated RDE environment. For instance, a NIST publication documents levels of identity/assurance to consider:[49]

> **Level 1:** At level 1, there is no identity proofing; names are assumed to be pseudonyms. Authentication requires that the user demonstrate that she controls the token. The sole protection of user secrets comes from the requirement that user proofing data not travel in the clear, and the only thing that the level 1 mechanisms do is provide some assurance that it is the same user who is accessing the protected data. Level of Assurance (LoA) 1 gives minimal confidence about the user's asserted identity.
>
> **Level 2:** At level 2, some identity proofing is required. (There are different requirements depending on whether the identity proofing is in person or remote; if in person, the user must show a valid current government identity document that has a picture as well as either nationality or address of record, while if remote, a financial account number is also required. Passwords and PINs are allowed for authentication, as are more secure forms of authentication (such as hardware tokens). There are system security requirements, e.g., there must be mechanisms to handle revocation of credentials, passwords must be of a certain strength. Thus LoA 2 provides some assurance regarding the asserted identity.

[49] Burr, William E., Donna F. Dodson, W. Timothy Polk, "Electronic Authentication Guideline", NIST Special Publication 800-63, Version 1.0.2, National Institute of Standards and Technology, Technology Administration, U.S. Department of Commerce, April 2006 (at: http://csrc.nist.gov/publications/nistpubs/800-63/SP800-63V1_0_2.pdf) and the American Library Association, Privacy: An Interpretation of the Library Bill of Rights, http://www.ala.org/ala/issuesadvocacy/intfreedom/librarybill/interpretations/privacy.cfm.

Level 3: At level 3, the identity documents must be verified, with a higher level of proofing on the identity than for level 2, and two-factor authentication is required. In addition, the level 3 authentication mechanisms require cryptographic-strength protection of the primary authentication token (the token can be unlocked through a key or biometric. LoA 3 gives high confidence in the identity being asserted.

Level 4: At level 4, identity proofing can only occur in person; the government ID is to be verified with the issuing agency. The assertion mechanism is "hardened," that is, only "hard" cryptographic tokens can be used, the FIPS 140-2 cryptographic module validation requirements are strengthened, and all critical data transfers are authenticated through a key bound to the authentication process. The user must prove that they control the hardware token. LoA 4 gives very high assurance in the asserted identity.

Other NIST guidelines on federated systems include:

- **A Credential Reliability and Revocation Model for Federated Identities** at: http://csrc.nist.gov/publications/drafts/nistir-7817/Draft-NISTIR-7817.pdf;
- **Privacy Management: A Positive Perspective on Privacy Standardization** at: http://csrc.nist.gov/groups/ST/IDtrust/IDtrust2012/presentations/sabo.pdf; and
- **Guidelines on Security and Privacy in Public Cloud Computing** at: http://csrc.nist.gov/publications/nistpubs/800-144/SP800-144.pdf.

In addition to these risks, other decisions that comprise a successful federation policy include:

- Who will collect the data on transactions, thereby generating additional data sets? Do these new data sets have value for anyone or any applications in the transportation realm? For monitoring RDE performance?
- Who establishes the levels of authentication and criteria for who can access and when?
- Who has responsibility for failures or incidents? Who will establish a response and recovery plan and who will implement it?

These and other considerations are discussed in a white paper done for NIST, titled: **Economic Tussles in Federated Identity Management**[50] which further analyzes the reasons for success of different identity management systems. To identify which type of technology will be most effective for the RDE, the DCM program team will need to make decisions about security, privacy and access, and determine which levels of assurance are most appropriate for the RDE.

Policy and Technology Recommendations for the Federation of the RDE: The most important aspect of policy for federation is to implement strong policies on security and privacy (as described in section 1.2). During the research phase and real-world demonstration, the RDE is owned and overseen by the Federal government and thus must apply federal policies (see the recommended NIST guidelines). Also important is to develop a set of new policies that guide data sharing and user access and establish governance and responsibilities within a federated system. To create these policies, the DCM team will need to conduct further research and provide further technical definition of the system. Recommended steps are to:

[50] Landau, S. and Tyler Moore, "Economic Tussles in Federated Identity Management", 2010. Radcliffe Institute for Advanced Study, Harvard University and the center for Research on Computation and Society, Harvard University.

- Conduct research to identify the types of organizations that are likely candidates for federation, and conduct a risk analysis on whether and how risk may increase through federation.
- Determine criteria for federation – what types of organizations should be allowed to participate in the federated system? Are there any specific practices that might exclude an organization?
- Determine appropriate levels of data quality and determine whether these standards can apply throughout the federated system (or whether the cost-burden makes it infeasible). Based on the Data.gov data format standards, determine whether these standards are enough to create the level of interoperability needed for a dynamic data exchange through the RDE.
- Determine criteria for user access – if a user gains access through one node in the system, does that or should that grant access to the entire system? Are there parts of the RDE that may need to be firewalled from everyday users due to the sensitive nature of the data? What level of assurance is needed for the RDE (based on risk)?
- Determine data sharing policies with organizations based on terms of data use that are developed with the legal policy team.

1.4: Addressing the Gaps and Next Steps

The purpose of this section was to examine the issues associated with implementing open data and open source policies that support potentially transformative connected vehicle research and operational data environments. The case for adopting an open data policy is supported by current U.S. and international examples within the public sector. There are also several established licensing options that would facilitate implementation while addressing important liability questions pertaining to ownership and intellectual property. The primary advantages realized by an open data policy include increased access to information from taxpayer-funded systems, greater information sharing across organizations, and a readily available source of high-quality real-time data that encourages innovative applications and improved operational efficiency. In order to deliver these benefits, an effective open source policy must be widely accessible and cost-effective while addressing the risks concerning security, privacy, liability, and data quality. Below are recommended efforts and next steps:

- **Security and Privacy Policy Efforts/Next Steps:**
    - Develop an approach to security using NIST guidelines.
    - Working with the ITS Legal Policy team, develop a set of privacy principles that align with the privacy principles for the broader connected vehicle program. Building from these principles, describe and enact privacy controls (NIST 800-52, Appendix J). A prior paper title *Privacy and Security Analysis and Recommendations for the DCM and DMA Programs* provides an initial baseline for developing privacy policy and controls for the Mobility programs. This paper also contains a prospective analysis of the potential PII associated with the data needed for the dynamic mobility applications.
    - Working with the CIO, determine how to ascertain whether data from other systems that is merged with RDE data can be reviewed and considered safe/secure in real-time.

- Working with the ITS Legal Policy team, determine responsibility for data that comes from other system sources and is merged with RDE data.
- Working with the technical and architecture teams, determine if there is a cost-effective way to leverage the digital certificates for security of the RDE.
- Analyze the costs and trade-offs of implementing de-identification technologies with the RDE to remove certain data characteristics as the data is streamed into/captured by the RDE. This analysis will require a prior development of user scenarios to identify what aspects of the transportation data are most valuable to researchers and applications developers. Such scenarios can be used for analysis of data sharing/use in Section II.
- Data monitoring is the final security measure to examine, in order to fulfill specified accountability requirements. Active monitoring can provide valuable feedback on security effectiveness through systematic sampling. The acceptable security risk study being conducted by the Implementation Policy and Legal Policy teams (expected in summer of 2012) will establish an important baseline for determining the most appropriate security measures and thus yield a better picture of potential costs associated with their implementation.

- **Liability Policy Efforts/Next Steps**
  - While conscientious security and data quality monitoring play important roles in limiting liability, terms of use (TOUs) policies and disclaimers can add an additional layer of legal protection. Work with the ITS Legal Policy team to develop federally compliant language and appropriate transfer of risk to the user.
  - For data quality assurance, develop a set of metrics to define quality data from an RDE perspective (for liability purposes) and a user perspective. Gather stakeholder input from mobility application developers and analysts to define the data quality requirements for DMAs and other applications, potentially in the May 2012 Mobility Workshop.
  - Further investigation is also needed on the relationship of the RDE data quality to Data.gov standards, including any compliance requirements.
  - Work with the ITS Legal Policy team to support their analysis of data ownership. Once the legal analysis is complete, apply it to the RDE and datasets to determine if there are technical, policy, or institutional impacts.
  - Work with the ITS Legal Policy team to identify the licenses that are appropriate for RDE use. Key input is needed from the technical team to describe the RDE architecture and technologies that will be used, and to analyze the components for the range of licenses that either are in existence or new intellectual property that will need to be licensed. Once decisions are made, incorporate the licenses and guidance into the portal development.
- **Federation Policy Efforts/Next Steps**
  - Develop criteria/policies for federation. Criteria/policy should be based on risks and opportunities. A more formal analysis of risks should be based on a set of scenarios, recommended for development under security and privacy and can be leveraged for analysis on data sharing and use in Section II.

Section II: Research Data Exchange (RDE) System Policies

In addition to addressing policies related to data, a successful RDE will be guided by system policies. This section builds from the policy analysis and recommendations in section I to further describe a set of policy requirements that will support the RDE ConOps and help guide development of the RDE design and prototyping. System policies are envisioned to include:

2.1 Governance needs/Governance options

2.2 Access policy options that address user access and authentication, standards and certification, and security

2.3 Data management policy options that address data use and sharing, metadata, and data storage and archiving

2.4 System policy options that address rules of conduct, system limits, monitoring and enforcement, accessibility, language, system availability and recovery, and upgrades and maintenance.

The following analysis builds from a set of technical decisions about the RDE:

- The RDE will use a cloud-based architecture approach. At this time, three alternatives are available
- The cloud-based operations of the RDE will be based on a Federal contract with an existing cloud provider. Analysis of different providers and their capabilities is under analysis along with the alternative architecture analyses. These analyses are being performed by an expert contractor, IndraSoft, Inc., a firm that will remain available to the DCM program through the design and development of the RDE and will serve as a liaison with the chosen cloud-provider.

2.1 Governance Needs/Governance Options

Governance provides a framework for decision-making and management of any enterprise in which multiple individuals and organization entities participate. A governance framework specifies the roles and responsibilities of participants and the processes by which decisions are made. It defines the structure for collective decision-making by defining end goals, allocating resources, setting priorities, monitoring progress, and determining the conditions for starting and ending programs, products, projects and processes. The governance structure for the RDE will also enable communication among participants and provide a structure for redress and conflict resolution.

Through a review of the RDE ConOps and a set of alternative architectures for the RDE[51], it is envisioned at this time that the RDE governance will be required at two levels:

- Program-Level Governance

[51] Internal review document titled, *Research Data Exchange Architecture Analysis of Alternatives (Draft, Version No. 1.0)*, February 24, 2012, prepared by IndraSoft, Inc. for the DCM Program.

- RDE-Level Governance (or system governance).

Program-level governance defines the overall roles and responsibilities of the individuals and entities that have a stake in the RDE. Once defined, this group establishes policies and guidelines for operations associated with the RDE; this group also determines and monitors the resources that will need to be committed in support of continuous operations and maintenance. At this juncture in time, these roles and responsibilities map to the current DCM Program management level. Specific roles include:

- Develop RDE Policies for:
  - Access policy options that address user access and authentication, standards and certification, and security
  - Data management policy options that address data use and sharing, metadata, data storage and archiving, and data security and authentication
  - System policy options that address rules of conduct, system limits, monitoring and enforcement, accessibility, language, system availability and recovery, and upgrades and maintenance.
- Assign a program manager who is responsible for and monitors the proper daily operations and who will carry out RDE-level governance that includes the application of privacy and security policies, and the risks. The Program Manager should have the decision-making capacity to keep operations running smoothly as well as the responsibility to lead recovery and response in the event of an incident.
- Identify a set of decision makers who; resolve conflicts; reinstate users who are accidentally or intentionally removed; commit funding; and develop criteria for federation linkages.
- Consult with legal counsel for licensing and IP issues.

RDE-level governance is concerned with customer satisfaction, system performance, and mitigation of system risks. Its governance functions are associated with implementation of content management and change control, access management, security and monitoring, and RDE operations (including troubleshooting, downtime, backups, and patches), as reflected by the list of policies that will be established by the Program-level governance body. A new role – that of the RDE and Data Operations Manager (can be the same person or entity or can be different) is needed to enact the policies set by the program-level governance body and to resolve and raise new issues appropriately.

A key issue with RDE governance is that a cloud-based architecture assigns the functions of security, privacy, and other risk mitigation to a third-party while requiring the Federal government to maintain responsibility and authority. This is a more distributed form of sharing in the oversight responsibilities and roles and responsibilities will require careful documentation.

Recommendations for Program-level governance: For Program-level governance, it is recommended to establish a decision-making body to review and vet the policy options and establish RDE-policies. The establishment of such policies will include identifying and deciding upon trade-offs. Such a group is recommended to include, at a minimum, key stakeholders such as:

- A core set of DOT decision makers
- Experts in developing data environments and managing data
- A representative of the RDE contract development team
- Representatives from the applications teams that rely upon the RDE to be available for their research
- A policy representative to identify policy issues, provide analysis, and liaison with the broader policy efforts for the connected vehicle environment
- A representative of the ITS Legal Policy team to advise on legal issues.

Recommendations for RDE-level governance: Appoint an operations manager to work with the policy analysts to test the policies and determine if changes are needed or if the policies form adequate support for the RDE.

The remaining sections of Section II define the different options for setting policies and propose recommendations.

2.2 Access Policy Options

2.2.1 User Access and Authentication

User access policies specify who can access, use, and contribute to the RDE. Access policies can range from the restrictive (access is allowed only through a full vetting and authentication of a user each time he/she would like to access the portal) to relatively open (anyone can access).

For the RDE, there are a number of considerations in analyzing the appropriate level of access:

- The RDE is being populated with anonymized data which poses little risk (if any) to exposure of personally-identifiable information.
- The DOT is working to achieve openness through an open data policy discussed in Section I and through participating in Data.Gov. The RDE potentially offers a leading example for making full data sets open to the public in a manner that has considered and addressed the issues of data ownership and intellectual property.
- The RDE, however, is a new architecture and may include security vulnerabilities that are not yet well-understood, particularly if the RDE adopts a cloud-based architecture. While NASA and DHS both offer Federal leading-examples of moving to the cloud, notably, both agencies have chosen to develop their own clouds and retain significant operational oversight. In this manner, the RDE is establishing new lessons for the Federal government.
- As it is launched, the RDE will need to define access criteria and procedures. The RDE ConOps, developed in summer of 2011, has identified up to eight separate potential user classes that desire access for different reasons. While this is a larger number than the typical data environment or portal, there are risk-based reasons to establish a finer distinction among user classes to begin, and to modify and merge some of the user classes as the RDE evolves and as risks are better understood and managed. In comparison to the levels of assurance defined by NIST 800-63 (see page 24), these

RDE user classes align with the four levels that require more knowledge in return for greater access.

User Access Policy Recommendations: The initial set of user definitions appears reasonable to use as a basis for policy, but should first be vetted with stakeholders and external users. An opportunity might include the upcoming Mobility Workshop in May of 2012. To the extent that eight categories are appropriate and decision makers are comfortable with them, the Program-level governance team can begin to define criteria for access by focusing on the following questions at each level:

- What information is needed for users to gain access? What are the potential risks associated with each user class and how will those risks be mitigated?
- Are there certain types of users that will need access to more than one area or level?
- How will the RDE store and manage the information provided by users?

Building from this access policy, the Program-level governance team will need to define procedures for when access is denied and/or how discontinued users regain access.

Closely linked with access is the issue of authentication – ensuring that the user can be trusted and, for the user, ensuring that the RDE can be trusted. There are three options available for authenticating users to consider:

- **At the level of the RDE**: Users sign up and provide their information. Upon review and approval, the user is granted access. Authentication occurs through password systems that require the user to provide a unique user name and password. Higher levels of security may require additional log-ins and passwords at each new level of the RDE.
- **At the level of the connected vehicle system:** There is an opportunity for the DCM program to consider how it might employ the digital certificates that are expected to be available through the safety portion of the connected vehicle program. From this perspective, users would authenticate themselves through the exchange of encrypted certificates—an authenticated user's certificate would open access to the RDE.
- **Through third-party cloud provider:** The DCM program can look to the cloud provider to identify what type of authentication system is existing for their cloud, and determine if it meets the needs. To do so, the cloud provider will need to be identified and reviewed to ensure that they are compliant with RDE and other Federal policies.

Authentication Policy and Technology Recommendations: The options available for authentication will need to be studied for their feasibility. It may be that the digital certificates and core system will not be ready to leverage this technology choice at the point in time when the RDE is launched. It may also be that the cloud provider has an authentication option that is cost-effective. Once these issues are better understood, a recommendation can be made on the direction for authentication policy and technology.

2.2.2 Standards and Certification Policies

Standards Policy relates to whether certain types of standards –be they data definitions, data formats, technology standards, web standards, or communications protocols, be required for:

- All data sets uploaded or downloaded or streamed into the RDE
- All technology that connects to the RDE.

There are two reasons to establish a standards policy – for interoperability and for assurances of increased data quality.

Standards Policy Recommendations: For the RDE, the first level of interoperability relates to widespread access to the RDE. As noted in the RDE ConOps, widespread access can be achieved through the use of internet standards and protocols. With regard to data standards, recommendations are offered in section 1.2.2.1 to ensure quality. Further, to ensure interoperability, the use of ITS standards to define data is an appropriate element of standards policy. Research will be needed, however, to compare the ITS data definition standards against the expected types of data for the RDE. It is likely that there may be gaps. The ITS Standards team is expected to perform this type of analysis in summer/fall of 2012 to identify gaps in standards and propose solutions for filling those gaps. Solutions include modifying an existing standard in use for other industries or developing new standards (the option that should be used as a last resort). A further decision for Standards policy is the decision on whether to harmonize standards at an international level – a decision that can only be made once standards are known. The recommendation is that once the RDE architecture is decided upon, to conduct a data and technology standards analysis.

Certification policy is associated with the decision to require equipment that connects with the RDE to be certified according to certain standards. This not only supports interoperability but also helps with security protection. As noted above, to the extent that the RDE is built using internet standards and protocols, the need for requiring certification appears minimal. For security purposes, however, if the digital certificate system is used for authentication, the certificates may undergo a certification before distribution to users.

2.2.3 Security Policies

These policies and recommendations are discussed in section 1.2.1 and result in policy recommendations to follow NIST guidelines for developing security, incident, and response plans as well as to consider use of specific types of technologies.

2.3 Data Management Policy Options

2.3.1 Data Use and Sharing Policies

Establishing policies on data use and sharing are critical elements of privacy policies (agreement on how data will be used, for what purposes and with what limitations) and of liability policies (sharing agreement help establish who is responsible for the data and who will oversee data management, quality, and proper use).

Depending on the time period of the agreement and the presence or absence of commercial interests, a variety of formats for data sharing agreements are used in industry and for research. In general, ongoing agreements are structured as memoranda of understanding or agreement. If commercial interests are present, a contract is typically used.

Appendix A lists examples of data sharing agreements and best practices which were reviewed to determine the most common elements of intergovernmental data sharing agreements. There was significant overlap between the agreements reviewed; the textbox on the next page summarizes elements common to multiple agreements.

Overview of Common Elements of Data Sharing Agreements

- Definitions
- Parties to agreement
- Period of agreement
- Purpose: Justification for access (if relevant) and Allowed uses of the data
- Data description or definition:
    - Data sets included
    - Data quality (if not a separate section)
- Data location and custodial responsibility
- Data transmission format
- Data quality:
    - Requirements for Standards
    - Liability statements / agency disclaimers
    - Metadata descriptions and requirements
- Access. Who the users are and what permissions they have.
- Dissemination to third party users. Is this allowed, what restrictions are in place?
- Derived data statement. What should happen if derived data are created?
- Source requirements. How should the data be cited? Does a disclaimer need to be included?
- Confidentiality statement for users and to guide staff in proper procedures
- Legal restrictions: Appropriate copyright, licensing, and incorporation of regulations by reference
- Disposition of data. What happens to the data at the end of the agreement period?
- Cost / cost-recovery / resources. Where agencies will incur costs, how are they to be shared?
- Agreement administration.
    - Renewal
    - Amendment process
    - Key Officials
    - Breaches to the agreement
- Signatures

Data Sharing and Use Policy Recommendations: In developing a standard agreement, the Program-level governance team will need to work closely with the ITS Legal Policy team to ensure that the correct elements and language are a part of the agreement. Before developing such an agreement, however, the Program governance team will need to establish appropriate use policies for the data.

Although the RDE is being established as a research project, use of the data may result in commercial application development, including free and open source as well as proprietary intellectual property. It is important that, before allowing for broad data uses, the ITS Legal team render an opinion regarding data ownership, appropriate licenses, and licensing procedures. It is also important to clearly describe the range of potential uses to those who share their data with the RDE and to assure that those individuals and entities are in agreement about the data uses that may go beyond research.

2.3.2 Metadata

Metadata is "…structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource."[52]

Metadata Policy Recommendation: Through the development of the RDE ConOps and with identification of best practices, the DCM team has decided to use an existing practice known as the Dublin Core Metadata Element[53] in a modified format. This best practice will provide the RDE with a unified, high-level set of standards to navigate and manage the numerous datasets within the RDE. A follow-on recommendation is to align this practice with the Transportation Research Board's (TRB) Metadata Working Group as a means of broadening acceptance throughout the transportation industry.

2.3.3 Data Storage and Archiving

Data storage and archiving policies describe the processes for storage and handling of the data, and describe the policies for length of time and location/security of archived data. According to the NIST privacy controls (FIPPs), data should only be stored for justified reasons and only for the amount of time necessary to fulfill a critical purpose. In the case of the RDE, archived data is likely to be present through the federated linkages of other agencies that collect and store data (i.e., traffic and transit agencies; planning organizations).

Data Storage and Archiving Policy Recommendations: A key reason for developing architecture alternatives based on cloud computing is the anticipation of the size of the RDE datasets, particularly when collecting the data from the Safety Pilot test, and the anticipation of the need for expansion. Given these constraints, it is recommended that data storage within the RDE be limited from a cost and management perspective.

However, the length of time for archiving is best set based on communications with the applications developers who will rely upon the RDE datasets. Notably, the DMA efforts are expected to take up to 2-3 years in development and testing; other applications and models that require longitudinal data may require storage/archiving for longer. Thus, each data set must be considered individually. The Program-level governance team and legal counsel should be engaged to set the policies for each dataset.

Also, the RDE will require policies for end-of-life processes that ensure that the data is truly deleted, are needed. A number of services exist to assist with deletion; as this is a possible

[52] NISO. Understanding Metadata, http://www.niso.org/publications/press/UnderstandingMetadata.pdf

[53] See: http://dublincore.org. The name "Dublin" is due to its origin at a 1995 invitational workshop in Dublin, Ohio. It uses the term "core" because its elements are broad and generic, usable for describing a wide range of resources. The National Transportation Library bases it digital repository on the Dublin Core.

service associated with the cloud computing provider, this task should be explored with the provider first to fully understand their processes and technologies. Options include overwriting on top of the database; wiping the hard drives; erasing/virtual "shredding" of data. Each of these processes can have a different end result in terms of the level of permanency in deletion—wiping the hard drives is probably the most permanent solution but also time consuming and expensive if other datasets need to be safety moved. From a Federal policy perspective, however, NIST, FISMA, and other guidelines direct the use of the most permanent solution once a time limit has been met.

2.4 System Policy Options

2.4.1 Rules of Conduct/System Limits

Codes of conduct clarify how users are expected to treat one another when working in the RDE, and define appropriate behavior when in the portal. These days, many portals have active community forums and blogs; many "codes of conduct" examples are directed at use of language and clarity of identity to encourage respectful discourse. The RDE is expected to have discussion areas and will need to develop a set of rules of conduct for personal behavior.

Beyond conversation, however, is a critical set of behaviors regarding:

- Relevant contributions and the limitation of materials and datasets that are not germane to the RDE's purpose
- Size and time of uploads/downloads so as not to overload communications links and interfere with other users' transactions
- Identification of rights when uploading/providing data to the RDE
- Guidelines on correction of mistakes and/or identification of mistakes and errors in others' work
- Definition of ethical behavior, particularly with regard to proper following of license agreements and other terms of use
- Definition of conflicts of interest and/or whether any data is to be considered highly restrictive and not available for download or use with other endeavors.

Appendix D provides a small set of examples of rules or codes of conduct from other portals. Most of the examples contain common elements. However, given the Federal nature of the RDE, the ITS Legal team should be engaged for development and review of the RDE codes.

2.4.2 Accessibility and Language Policies

As a Federal project, the RDE will be required to provide the RDE's datasets and supporting reference materials and policies in a manner that is compliant with the Americans with Disabilities Act (ADA). For the datasets, tagging the data with metadata helps the electronic readers used by those with sight impairments to identify and navigate the datasets (tags can be embedded within templates, item indexes, header labels, and/or style sheets for the HTML portion of the RDE).

Providing data in standardized web formats is also an important policy for meeting ADA requirements. The data formats discussed on page 17 are allowable formats and, notably, used by Data.Gov. In addition to the datasets, supporting materials should be posted on the RDE portal in MicroSoft Word or Adobe format with the appropriate alt-tags provided. As this has become common practice throughout the DOT, meeting this requirement should be straightforward for the DOT team. However, contractors may require support and/or high-level training. If this is the case, the Federal government offers an archive of guidance and tools at http://www.section508.gov for Word documents and data. For PDFs, guidance is provided by Adobe and a free PDF checker is available for checking for compliance.[54]

Recommended Accessibility Policy: It is recommended that the RDE datasets be provided to users in a standard format based on existing internet formats and that supporting documentation be translated into 508 compliance. Basic guidance for the portal development as a 508 compliant tool includes the following[55]:

- Make accessibility part of standard operating procedures: Design, develop, and test the portal for multiple browsers and versions of browsers, operating systems, connection speeds, and screen resolutions, based on an analysis of an organization's website visitors.
- Balance needs: Balance the needs of visitors who use lower-end technologies with the need to pursue more advanced technologies and the added functionality those technologies may provide.
- Use web analyzer tools or other analytic data to review visitors' technological needs at least semi-annually. Ensure that your websites continue to meet the needs of their intended audiences.
- To the maximum extent feasible, minimize page download times. In most cases, HTML pages should not exceed 100 KB.
- Don't use web design technologies (such as Flash) if the intended audience generally cannot and does not have access to those technologies.

Regarding language, American English is the standard language for American Federal websites. It is recommended that American English be the primary language in use for the RDE. Since the predominant set of data that is captured for the RDE describes American systems and transportation facilities, American English aligns with the products and technologies providing this data. The additional use of ITS Standards (formats and data definitions) and internet standards, particularly those that have been or are undergoing ISO harmonization and adoption, results in a greater, more universal understanding of the data.

Notably, the World Bank has provided grants to translate some of the ITS training, guidance, and tools into other languages where there is greater benefit to the population.

[54] at: http://www.access-for-all.ch/en/pdf-lab/pdf-accessibility-checker-pac.html
[55] Synopsized from: Providing Common Access for a Broad Range of Users at: http://www.howto.gov/web-content/accessibility/common-access.

2.4.4 System Availability and Recovery Policies

A key policy is to develop a set of system performance metrics for system availability, reliability, and redundancy. Establishing such definitions can be aided by DOT's common IT practices that identify a system as mission critical, important, or as a reference system. The chosen designation relates to a set of policies on frequency of back-up, level of redundancy required, and other metrics. A set of NIST publications is available to guide the designation of a system, the setting and testing of requirements, and guidance on monitoring, incident response plans, and recovery plans.[56]

System Availability and Recovery Policy Recommendations: According to a recent business case for RDE investment, the RDE is considered a mission critical system for testing critical processes and functions associated with the connected vehicle environment. With this designation, the program-level governance team will need to decide the parameters around operational "up-time" and the frequency of back-up systems that limit downtime. Typical parameters for mission critical systems are 24/7 with 5 minute downtimes. However, as the RDE is a research tool, these parameters do not need to be as rigorous, but do need to meet the needs of users.

A key recommendation is to map how the ways a system is likely to fail and to build in "elegant degradation" so that the entire system does not shut down immediately. Another key element of system failure is the development of communications to users regarding the un-availability of the system. The technical team needs to conduct an analysis of the potential issues that might arise from system failure while the design team builds a process whereby parts of the system shut down in a consecutive manner (or, if possible, stay running independent of each other), providing opportunities for more proactive response).

2.4.5 Monitoring and Enforcement Policies

The Program-level governance team will need to establish what system metrics or user behaviors are unacceptable. These definitions form the basis for monitoring of the system to identify anomalies and enforcement of the policies.

Key elements for monitoring and enforcement that will need to be addressed in policies have significant overlap with other policies; for instance:

- **Linkability** and the assurance that users are not attempting to link data to personal identity or otherwise obtain information through the comparison/merging of disparate databases–either RDE databases with internal databases or RDE databases with external databases. Linkability is not easy to monitor as it typically is an action that is external to the portal. To address this issue the privacy policy/FIPPs will need to describe the inappropriateness of linking databases for purposes of tracking or identifying a person associated with the data. Additionally, use of technologies [57] to further de-identify certain characteristics might help mitigate this risk; however, most

[56] NIST SP 800-60: Guide for Mapping Types of Information and Information Systems to Security Categories, located at: http://csrc.nist.gov/publications . Once the designation is decided upon, planners use FIPS 199 and FIPS 200 to set and test the minimum security requirements for the RDE. A checklist for security configuration exists in NIST SP 800-70. Security testing and assessment techniques are provided by NIST SP 800-115 which also provides guidance on developing a Plan of Action and Milestones, including incident recovery.

[57] See the privacy analysis paper titled, Privacy Analysis and Recommendations for the DCM and DMA Programs.

successful portals and systems have clear, accessible policies that detail consequences that help mitigate risky behavior on the part of users. Most critical systems provide clear user and system operators with policies that detail legal action if such a policy is violated. The ITS Legal Policy team will be instrumental in helping to set the privacy policy.

- **Hacking or introduction of malware or viruses.** Technologies are needed to help monitor the RDE for attempts at hacking or introduction of malicious software/data. However, similar to linkability, policies are an important tool to mitigate risky behavior. NIST guidelines on security provide direction on system monitoring for intentional attacks; current laws and DOT system policies are available that describe the legal ramifications for users who knowingly take such actions. The ITS Legal Policy team will need to help identify how the legal policies will apply to the RDE, particularly as the RDE will be operated by third-party operators but the responsibility and oversight remains in Federal hands.
- **System failure** is addressed in section 2.4.4 in terms of establishing system performance metrics that allow monitoring against a baseline to detect failures. In addition to those metrics, policies are needed to establish when system failure is due to issues beyond control of an operator (i.e., failure of a particular part, failure of an energy source, etc.) versus failure due to intentional negligence or malicious behavior. Typically, in such an event, a process of legal discovery is used to identify errors, negligence, and accountability. The Program-level governance team will need to work with the ITS Legal Policy team to determine consequences if the operator is negligent or intentionally malicious.

2.4.6 Policies on Upgrades and Maintenance

Regular upgrades and maintenance are a key element of data quality. Establishing policies that are supported by a commitment of funding helps to mitigate important risks and thus mitigate certain forms of liability based on negligence.

Upgrade and Maintenance Recommendations: To set a policy on upgrades and maintenance, further research is needed. The following questions about the RDE architecture/design require focus:

- What are the technologies used and how frequently do they require upgrades or maintenance? At this point, it is assumed that the RDE will make use of a third-party cloud provider but that decision is not yet final. Typically, maintenance is built into the contract; however, given the newness of the cloud concept, the Program-level governance team may desire more frequent maintenance to ensure proper operations.
- Are the technologies for the RDE evolving and, if so, is the next generation near-at-hand or a few years away? Gaining insight into the trajectory of technology evolution is helpful for establishing a baseline to determine if/when upgrades might be necessary during the operations of the RDE (and the approximate length of operations needs to be made explicit).

With these insights, a policy on upgrades and maintenance can be recommended.

2.5 Gaps and Next Steps

Section II identifies the policies associated with the RDE. The section defines, identifies examples of, and proposes policy recommendations for: *system and data governance and management processes, operational practices and rules of conduct, security and privacy, standards, integration, and rules for data exchange and sharing*.

Because the RDE architecture and technologies have not yet been chosen, there are a number of research actions/inputs that are needed to develop full policies in each area. The following provides a list of the gaps and next steps for consideration:

- **Technical Inputs Needed:**
  - Provide the policy analysts and the ITS Legal Policy team with a final choice of technologies to allow for full analysis of the policy and intellectual property/licensing issues.
- **Governance needs/Governance Efforts/Next Steps**
  - Form a Program-level governance team to develop and approve policies for the RDE. Recommended minimum participants include:
    - ITS JPO program manager
    - FHWA program manager
    - FTA and FMCSA representatives
    - ITS Policy program manager and/or policy analysts
    - ITS Legal team
  - Develop policies and procedures and guidelines for users and operators to be stored and accessible in the RDE
  - Review drafts with stakeholders before finalizing.
  - Once draft policies are developed, review against the ConOps and Architecture to assure that policies don't negatively impact the technical choices and path. If there are such challenges, analyze trade-offs and make decisions needed to finalize the policies.
- **Access Policy Efforts/Next Steps**
  - Identify how the user access controls will be implemented in the RDE architecture/design and analyze which technology choices are most viable, given the architecture. Determine if the digital certificate technology can be leveraged to authenticate users.
  - Vet the RDE ConOps user classes with stakeholders, potentially in the May 2012 Mobility workshop.
  - Work with the ITS Standards team to identify where standards are needed; based on analysis, develop a policy on requirements to implement/use standards and/or certification.
- **Data Management Policy Efforts/Next Steps**
  - Develop a set of scenarios for how the RDE data will be used by researchers and other users. Based on these scenarios, analyze whether these purposes for use are aligned with U.S. DOT and Connected Vehicle Environment policies. Determine whether there should be limitations on data use or data sharing.

- Develop a standard data sharing agreement for use with entities that can offer data to the RDE. Existing agreements can be used as a baseline.
- Work with the ITS Legal Policy team to determine how to handle copyright and licensing.
- Based on the established metadata policy for the RDE, develop an easy and accessible way for users to navigate the RDE datasets based on metadata.
- Working with the privacy policy team (a subset of the ITS Legal Policy team), identify needs for data storage and archiving and develop policies. The scenarios developed for data use can help determine reasonable ranges of time for data storage and archiving while NIST and FISMA policies can guide the development of processes for storage and archiving.

- **System Policy Efforts/Next Steps**
  - Using examples of other portals' codes of conduct, work with the technical team to draft a set of rules for the RDE and vet with the ITS Legal Policy team and stakeholders.
  - Implement the RDE using American English and according to ADA accessibility standards (508 compliance).
  - Working with the technical team, develop a baseline of system performance and availability. Use these metrics to develop policies for enforcement. Work with the Legal ITS Policy team to identify appropriate consequences for when policies are not followed by users and operators.
  - Working with IT and security experts, develop an incident response and recovery plan.
  - With the final decision on RDE architecture and technologies, have industry analysts determine the evolution of these technologies to identify reasonable metrics for maintenance and upgrades.

Section III: Conclusion

At this time, and until further technical definition of the system architecture, technologies, and datasets are provided, there are two overall conclusions and one set of prospective analysis worth noting:

- **Conclusion: Implement based on an open data policy**. An open data policy is a viable option and is encouraged by the U.S. Government in general and is emerging as a trend with other governments around the Nation and around the world. The level of "openness" is highly dependent upon some of the technical inputs – the accessibility of the RDE to public users; the critical and minimum characteristics of the data that will be captured, used, stored, and archived; and the risks/trade-offs associated with the technical definition of what it means to be open. This report, and the related other Mobility policy reports (see list on the next page), attempt to put some definition to these open questions. There is a need to have the whole set of reports and definitions vetted by the technical team and stakeholders to ensure that the basis for recommending policies is solid.
- **Conclusion: The RDE system policies can be based on proven solutions; however the federation policies require further analysis and development.** The RDE architecture and set of technologies that are proposed for use in the construction and operation of the RDE appear synonymous with other portals in use with the Federal and State governments, academia, and industry. As a result, most of the RDE system policy can draw from existing models. The key differences, though, from a policy perspective include the wide-scale federation and the monitoring and enforcement of policies through such a dispersed system. Developing a set of optional models (also referred to as "scenarios") regarding various entities that might link with the RDE and reviewing their policies and analyzing the impact to the RDE supporting policies is a useful next step to determine how the technical, policy, and institutional recommendations might align (thus supporting broader federation) or face significant impacts that might challenge federation (for instance, a conflict between privacy or data usage policies).
- **Analysis: RDE Next Steps.** Even though the RDE is being implemented for research purposes, lessons can be learned regarding future operational data environments. Further analysis on technology transfer, steps and policies to support commercialization, and the viability of sustainable marketplaces will be needed.

Envisioning An RDE Transition to Commercial Environments

Although the RDE is intended to serve as a research tool, it can provide helpful lessons learned for implementation of an operational data environment/data exchange system that is able to efficiently and reliably capture, clean, fuse, and analyze multiple real-time data streams. It is envisioned to accommodate both archived data and real-time "publish and subscribe" data in support of mobility, environmental, and safety applications.

Overall, the recommendation is to implement an open data policy with the recognition that it may evolve over time. The RDE is best modeled on a limited "public sector information" model (similar to the EU) during the research and demonstration phases.

The policies that are implemented with commercial adoption will be decided upon by the eventual owners and operators of the data environments. The following three scenarios are offered as potential options for the end-state of the RDE:

- **Federal Model:** If the Federal government finds that the RDE is most effective as a Federal tool that supports a nation-wide real-time data capture and management effort, then the Federal policies described in this paper will continue to support operations, even if those operations are contracted out to the private sector.
- **Transition to a Private Sector/Academia Model:** if the transition to commercial use is predicated upon revenue generation, it is likely that some or all of the policies in place during the research phase will change. The most likely changes are expected in the privacy policies, access policies (likely to institute fees) and in the "open data" policy as some level of exclusivity provides greater value to the data. To understand how these policies might shift, the DCM program might consider developing a set of scenarios and cost models to analyze the opportunities for revenue (or where there is value in the system) to understand how these opportunities might shift the policy basis. Such scenarios, however, require greater definition on the RDE technologies and systems and how they will work/operate in a form other than what exists now. From a policy perspective, though, if the DOT chooses to relinquish its position with funding, oversight, or governance, other Federal agencies might find themselves overseeing the data and data environments, notably the FTC with regards to privacy or the DHS with regard to cyber security and critical transportation infrastructure.
- **Hybrid:** There is a possible hybrid model that allows the DOT to retain some level of oversight and influence policies. In this scenario, the DOT would be expected to license both the new DCM technologies as well as the data in order to ensure the continuance of an open data policy. This may be challenged, however, by those industries whose devices (vehicles, RSEs, portable devices) are generating the data and may be captured and managed by those same industries. An effort by the Implementation Policy team to assess the market value of new connected vehicle data is underway and will be available in later summer of 2012. These results will help reveal the economic interest of the private sector. Additionally, to create such a hybrid, the DOT would need to commit to an ongoing financial and governance role to ensure that the data and technologies are properly operated and maintained. Analysis of this type and level of commitment is being conducted for the connected vehicle program as part of the Implementation Policy team. Results of this analysis should be available to the DCM team by summer of 2012.

Last, external forces beyond the control of the DCM program may influence the opportunities to transition to commercial operations. For instance, Congress has been deliberating on new legislation that creates new privacy protections for users. If passed, such legislation might significantly limit the economic opportunities of private sector companies. Further, if such legislation were to define geo-location data as PII and limit its collection, much of the new connected vehicle data applications may need reconsideration.

Beyond the question of what policies govern the RDE and the data during the research phase and into the operational phase, another important consideration is what type of policies might support such a transition? Or support or incentivize the market players during the transition? To understand what policies are needed, we first must understand what will be transitioned and

whether/how it has value and to whom. In addition, for the DOT to look to facilitate transition to commercial operations with the highest and most effective levels of openness, security, and privacy protection, the USDOT might consider the following:

- Ensure that the specifications and any new technologies and software are licensed for free, thereby decreasing a cost-to-entry by organizations looking to develop a data environment and increasing competition.
- Study the differences in existing data environments versus legacy data environments and provide guidance to organizations on how to migrate/enhance legacy systems to meet new data environment requirements.
- Automate as much of the policy standards as possible (i.e., security and privacy alerts, de-identification of data and/or credentialing processes, etc.)
- Document the lessons learned from building and operating the RDE based on Federal policies for security and privacy and note the benefits of building to these standards.
- Harmonize data according to international formats. Standardization supports the growth of the technology and application marketplace on a worldwide basis. Research is needed to understand whether any of the technology or application markets might benefit from harmonization (for instance, what is the variation in data formats that might prevent or facilitate the new integrated signal applications from being used anywhere in the world?)

APPENDIX A: REFERENCES

A1. University Models of ITS Open Database and Open Source Software

a) The **MIT Open Source MITSIMLab**[58] is part of the MIT Intelligent Transportation Systems research developing new computational models and applications that allow access to real time information. It offers a Linux platform for running simulation software and access to data conditional on agreeing to an Open Source License, "The MITSIMLab Version Control License". The MIT Standard Version package is provided without warranty and the license holder may modify it and post new copyrighted codebase only with a prominent notice stating how and why the file was changed. Commercial software may be developed with written permission, and reasonable fees may be charged for distribution. MITSIMLab synthesizes multiple traffic management system designs, and dynamic driver response model to real time situations, offering routing logic options. It has 3 modules: a graphic user Interface (GUI); a Traffic Management Simulator (TMS); and a Microscopic Traffic Simulator (MITSIM). *TMS might be an excellent candidate for upgrading its modeling capability with V2V and V2I input data streams, because it can run both on historical databases, and on real-time data for ATIS/ATMS incident management.*

b) **DynaMIT at http://mit.edu/its/dynamit.html** was developed with FHWA and ORNL ITS program support as a real-time computer network system to support the real time data collection and management, and the operation of Advanced Traveler Information Systems (ATIS) and Advanced Traffic Management System (ATMS) of Traffic Management Centers (TMCs). This existing data base for simulations, and the suite of dynamic transportation network models (e.g., predictive travel behavior and incident analysis and management) can serve as foundational model for the RDE, if it is expanded to accept V2V and V2I data.

c) **Several Regional and Tier 1 University Transportation Research Centers (UTCT)** have are focusing on ITS traffic management and modeling, and have developed extensive relational data bases including real-time data collection, management and dissemination with State DOT, MPO and industry partners. ITS Centers of Excellence include:

- The California Region 9 UTC led by UC Berkeley, in collaboration with UC Davis ITS program lead, UCI and other UC, Caltrans and industry partners developed- e.g., under its Advanced Transportation management and information Systems (ATMIS) testbed and PATH Intellimotion programs-capabilities for collecting, processing, analysis and display of real time data sets.
- The Texas Transportation Institute (TTI)[59] is a national leader in providing congestion and mobility information, and has a comprehensive effort on data collection, verification and analysis, with advanced query and data management capability. This archived open database is used to produce value-added data products the annual Urban Mobility Report at http://mobility.tamu.edu/ums.

[58] http://mit.edu/its/MITSIMLabOSnew.html
[59] At http://tti.tamu.edu/

- The University of Minnesota **ITS Institute**[60] received ITS America awards for collecting real-time data from smart vehicle sensors to assist MN DOT in managing congestion and improving highway safety. The **MN Traffic Observatory (MTO)**[61] researchers have polled and collected raw data from AVL, APC and ETL systems and SMART-SIGNAL data; developed traffic data pre-processing and data filtering and cleaning algorithms to remove outliers in order to improve data quality for analysis and archiving; and uses Structured Query Language (SQL) relational database for transportation data mining and fusion to support ITS systems applications.

A2. Metro Area and Regional Transportation Authorities and Traffic Management Centers (TMCs)

There are more than 75 metro area TMCs have operating deployed ITS systems. These TMCs already accept, organize and respond to real time data including weather (AWS), Ramp metering information, Fastrak Electronic Tolls Collection (ETC), 911 incident reports for EMT dispatch and Roadwork alerts through Variable Message Signs (VMS), etc. ITS/JPO could partner with the most advanced metro or Regional ITS leaders for RDE development, test and evaluations. Selected test bed examples that could also accept and process integrated RDE data as participants or users include:

- **San Diego District 11-a CA TMC** at www.dot.ca.gov/dist11/d11tmc/sdmap/tmc_main.html
- **TRIMARC**[62] system is an intelligent system operated by the Kentucky Transportation Department in the Louisville, KY and Southern Indiana urbanized area. It already includes an integrated system of distributed roadway sensors, video cameras, dynamic message signs (DMS), Auto Incident Recording System (AIRS), and the Condition Acquisition and Reporting System (CAES). It allows authorized users to access the system from any location using a web browser to report adverse conditions or incidents and to get performance information. Users can enter, view and disseminate critical road and traffic conditions via 511 phone, and Highway Advisory Radio (HAR) for motorists. It was designed and is managed by Northrup Grumann, and could be augmented with vehicle-based inputs for RDE purposes.
- **TriMet**[63] is the Tri County Metropolitan Transportation District of OR. This Public Transportation Agency and the OR Transportation Dept. (ODOT) already offer an interactive web page built around the OpenGeo Geoserver GIS mapping (OGIS), and use open source real time databases hosted in the cloud, as well as open source software (FOSS). This open data and open software policy successfully allowed new

[60] http://www.its.umn.edu
[61] at www.mto.umn.edu/Research/DMT/index.html
[62] See postings at www.trimarc.org/perl/about_trimarc.pl
[63] See at http://trimet.org/ postings and article **"***Portland Mass Transit creates geospatial maps and apps for commuters",* Government Computer News (GCN), July 29, 2010,*at* http://gcn.com/articles/2010/07/29/portland-mass-transit-uses-open-source-gis-tools.aspx?sc_lang=en; and *"Open Source Software Helps an Oregon Transportation Department for GIS, Website Development*", Government Technology, March 15, 2011 at http://www.govtech.com/e-government/Open-Source-Software-Oregon-Transportation.html. This article points out how a combination of in-house IT expertise and resources and a community of open source programmers/developers (including international contributors) collaborate online to develop new applications, such as OpenTripPlanner for ODOT and TriMet, based on public mobility data.

mobile applications to be developed collaboratively and posted for use by commuters. In-house TriMet IT expertise and resources collaborated online with a community of open source programmers/developers (including international contributors) to develop new applications, such as the OpenTripPlanner for ODOT and TriMet, using public mobility data. This type of collaborative open source software development and public real time data usage for creating diverse applications is a good example for the RDE.

A3. State Governments: Shared Data and Cloud Migration

The National Association of State Chief Information Officers- NASCIO posted resources on Best practices and policies and plans to assist a transition to IT Resource Sharing and Cloud Computing, see www.nascio.org/publications/index.cfm. Models of special relevance to RDE and DCM/DMA planning are:

- *"Capitals in the Clouds- the Case for Cloud Computing in State Government, part I: Definitions and Principles*".[64] This summary of CC options discusses the cloud deployment models, the value proposition for and advantages of adopting CC: cost efficiencies in resource pooling, rapid and elastic demand planning, and economy of scale. There are also risks for governance, security, privacy, ownership of and jurisdictional control of Cloud-stored data and information, cloud-supplier provisioning of services and lock-in, scalability and availability, vulnerability and security, etc. The paper also suggests risk management requirements to ensure that governance addresses the management of technology, institutional organization and culture, as well as supplier and portfolio of cloud services and products. A good model of a multi-state collaborative sharing IT data and services on an inter-enterprise and multi-jurisdictional basis is the Western States Contracting Alliance (WSCA) at www.aboutwsca.org The Western States GIS Collaborative is hosting all GIS data in the Cloud.
- *"Capitals in the Clouds Part II: Challenges and Opportunities to Get Your Data Right.*[65]*" Oct. 2011,* This paper stresses that Cloud Computing is only one option for shared state IT interoperable resources that avoid redundancy and reduce cost, which builds on available high speed internet connectivity, and virtualization of data warehousing. The key data issue is the requirement for **data harmonization**, using the **Extract, Transform and Load (ETL) process**, before open data can be published for shared access. The paper also addresses:
  - **Constraints on Shared Multi-jurisdictional Data and Services:** Shared Government IT framework and methodologies, such as the cloud computing option, must still address requirements for public service-oriented IT architecture, including the Information Technology Infrastructure Library (ITIL) and IT Service Management (ITSM); Control Objectives for Information and Related Technology (COBIT); project and portfolio management; agile programming capabilities for component development, etc.
  - **Data harmonization Challenges for Cloud Storage:** Data stored in the cloud for shared services must overcome challenges and meet the requirements for: data quality (cleaning, filtering, evaluation for accuracy, and reliability of contextual metadata); portability and interoperability; simplification, optimization

[64] http://www.nascio.org/publications/documents/NASCIO-Capitals_in_the_Clouds-June2011.pdf
[65] http://www.nascio.org/publications/documents/NASCIO_CloudComputing_PartII.pdf

and integration (fusion of data from disparate, diverse input streams); harmonization and consolidation of data sets for shared services. In addition governmental enforcement of Cloud data access authentication and control is needed to ensure data security for all public safety-critical applications.

- **Risk management**: There are inherent risks of data loss, security breach, or service loss for cloud data storage, which could be managed with redundant virtualization, recovery operating procedures and vendor management. This applies to any private, community or public cloud storage. The same risks and management strategies apply to IaaS, SaaS and other (hybrid) cloud platforms and services.

A4. Models for Data Sharing Agreements

- Australian National Statistics Service. A Good Practice Guide to Sharing Your Data with Others. Version 1, November 2009, at: http://www.nss.gov.au/nss/home.nsf/NSS/E6C05AE57C80D737CA25761D002FD676?opendocument
- Consequence Consortium for the European Commission: Methodologies and tools for data sharing agreement infrastructure. December 2008. http://www.consequence-project.eu/Deliverables_Y1/D2.1.pdf
- National Neighborhood Indicators Partnership: Key Elements of Data Sharing Agreements. http://www.neighborhoodindicators.org/library/guides/key-elements-data-sharing-agreements
- State of Utah: Utah Digital Spatial Data Sharing and Integration Project Memorandum of Understanding. December, 2009. http://www.fgdc.gov/grants/2009CAP/InterimFinalReports/088-09-5-UT-AppendixD-DataSharingMOU.pdf
- US Department of Health, Health Care Financing Administration; Health Resources and Services Administration; Centers for Disease Control and Prevention. Department of Health/State Medicaid Agency Inter-Agency Data-Sharing Agreement, https://www.cms.gov/smdl/downloads/SMD102298.pdf

A5. Cloud Computing Industry Leaders: Services and Best Practices

- **The IBM Federal Community Cloud and IBM Government Services**[66] offer Data Center Design with Analytics Services hosted in the cloud; the design is a modular data center with server optimization and Total Cost of Ownership (TCO) lifecycle cost modeling tools. IBM posted White papers for federal agencies:
  - "*Federal Community Cloud (FCC) for Government Organizations*" offers Infrastructure as a Service (IaaS)- secure, scalable, dedicated Federal Data Centers; Software as a Service (SaaS)- fast implementation of "development and test" environments, with Web hosting and backups, as well as customized Cloud-based applications.

[66] www.ibm.com/cloud

- "*Transforming Federal IT with Cloud Computing*"- This White paper summarizes federal guidance for agencies to consolidate data centers and migrate operations to the Cloud, offering assurance of compliance with Federal Information Management Security Act (FISMA) requirements to protect critical data security and privacy. IBM also offers consulting services to provide a seamless and efficient Cloud transition; a Smart Desktop Cloud for anytime, anywhere access to applications and data; IBM LotusLive Collaboration suite with on-line-collaboration tools; and Smart Business Development and Test Cloud to enable rapid applications development and testing.
- "*Cloud Computing Insights from 110 Implementation Projects*". The survey respondents include Government and Travel and Transportation industry Cloud Computing. Of interest to the RDE is the finding that the majority of CC projects to date focused on development and test for noncritical production workloads, as a stepping stone to operational, service-oriented architectures. Several inhibiting factors to CC were identified: Security, reliability and availability concerns; funding issues, complexity, lack of standardization in virtual machine hardware, software and OS stacks, new ways to manage data and services in the Cloud, loss of internal controls and software licensing issues, lack of skills to manage CC technologies and/or of a clear value proposition, etc.
- "*The Reservoir model and architecture for open federated cloud computing*" B. Rochwerger et al, IBM Journal of Research and Development, April 2009.

o **Microsoft**[67] offers the MS Windows SQL server Azure Cloud Computing platform and services. A relevant case study for the RDE as a community cloud solution is the regional transit data collection, fusion and analysis using cloud hosting for mobile applications is the **Public Transit Data Community** (PTDC). It was developed by EastBanc Technologies for Washington Metropolitan Transit Authority (WMATA) MetroRail and MetroBus, and is hosted in the cloud, using the MS Windows Azure platform.[68]

SQL Azure is a cloud based relational data base management platform. EastBanc developed the PTDC Application Programming Interface (API) software and data engine, which aggregate real-time data inputs from many transit vehicles and operators (vehicle locations, incidents, crowd and congestion patterns, weather, etc.), downloads it to PTDC every 20 seconds. The data is merged with the static (schedule) information, and used with an Evaluator Service and Intelligent software to calculate optimal routes and predict arrival times, or deliver delay notices. The desktop WIN based Azure platform for data storage and access has proved to be scalable, flexible and reliable. EastBanc serves as a third party resource for rapid applications development and data management, assisting WMATA to publish transit data and provide public services, and web access to transit information.

o **The CGI Group**[69] provides a how-to for federal agencies to implement Cloud First—steps for procuring cloud services using the GSA info.apps.gov storefront to ensure

[67] www.microsft.com/industry/government/developer and www.microsft.com/industry/government/federal
[68] http://www.microsoft.com/casestudies/Windows-Azure/EastBanc-Technologies/Firm-Uses-Cloud-Services-to-Unify-Data-from-Transit-Systems-Improve-Access-to-Schedules/4000009148
[69] www.cgi.com/cloud Issue Brief "*Making it Happen: Responding to federal initiatives to speed and simplify cloud adoption*", February 2011at http://www.cgi.com/files/white-papers/us-federal-cloud-computing-initiatives-white-paper.pdf

FISMA security compliance; observing Federal Risk and Authorization Management Program-FedRAMP requirements using the CIO.gov guidance; and accessing NIST resources on cloud computing.

A6. IT Companies Awarded GSA GWAC Contracts for Cloud Computing Services Applications and Infrastructure

- Alliant Technologies *Autonomic Resources*[70] ***are a small and disadvantaged service integration firm and cloud provider serving the U.S. federal government.***
- Winvale [71]is an IT Sales support firm also providing GSA schedule application and maintenance services.
- HP Enterprise Software and HP Cloud Automation offer intelligent and automated cloud service management for public and hybrid IT in the Cloud[72]; and the new HP 3PAR Converged Federated Storage Peer System Service portfolio.[73]
- Amazon Web Services (AWS)[74] offers scalable and reliable cloud infrastructure and computing platforms for government, with FISMA authorization and accreditation. "AWS scalability uses the Amazon Elastic Cloud 2 (EC2), Amazon S3, Amazon EBS, and Elastic Load Balancing." AWS has teamed with Apptis Inc. to provide AWS Services under Apptis' GSA Schedule 70 Contract, and can be procured under the Apptis GSA Infrastructure-as-a-Service (IAAS) Blanket Purchase Agreement (BPA). However, an October 2009 MIT Technology Review article[75] pointed out cyber-security vulnerability of the Amazon Elastic Cloud 2 (EC2) if attackers who know the IP address of the "virtual machines" and can map where data is physically located in the Cloud and gather intelligence. Comments point out that the MS Azure platform isolates the physical storage machine from the application tier to prevent this vulnerability in locating hardware.
- **Morph Labs** at www.morphlabs.com offers the mCloud computing virtualization services for data storage and management and dynamic IT resource provisioning for applications development, as well as integrated hardware and software cloud solutions. mCloud On-Demand is a free CC service compatible with Amazon EC2, which combines open source software with commodity hardware, and is a service supplier to Amazon. White Papers posted of interest to RDE data management include:
  - *"Controlling Cloud Resources with mCloud"*
  - "*The Revolutions of Commodity Cloud*"
  - "*The Next Maturity Step for Cloud Computing Management*"- a Forrester survey study commissioned by Morphlabs, defining Cloud Bursting flexible storage and other IaaS priorities for government respondents.
- **Accenture Technology Labs**[76] conducts R&D on emerging technologies. A White paper, "*Six Questions Every Executive in Infrastructure & Transportation Should Ask*

[70] www.allianttech.com/ and www.autonomicresources.com/
[71] www.winvale.com/
[72] http://www8.hp.com/us/en/solutions/solutions-detail.html?compURI=tcm:245-300983
[73] www.hp.com/hpinfo/newsroom/press/2011/110823xa.html
[74] http://aws.amazon.com/federal/
[75] *"Vulnerability Seen in Amazon's Cloud-Computing", Oct 2009 MIT Technology Review article at:* www.technologyreview.com/computing/23792
[76] At www.accenture.com

about Cloud Computing,"[77] offers insights and advice concerning transportation applications. It advises forming Public private partnerships (P3) to meet federal requirements, while enabling agile and innovative technology deployment, with special promise for ITS and GPS from mobile devices for fleet management, remote diagnostics. The Cloud solution that appears promising is Applications as a Service, in addition to IaaS for data storage. A case study is the use of Cloud by NJ Transit by using Sales force CRM Service to boost response capacity, instead of its legacy customer information service for buses, trains and light rail ridership. The Cloud ability to support flexible and responsive collaboration with 3rd parties is a desirable feature for multi-partner consortia. A combination of public and private cloud infrastructure can address government and partners' security and privacy concerns.

- **Google Apps for Government**[78] provides FISMA certified secure gmail, Google Sites web pages, and code-sharing, as well as GIS enabled visualization tools for federal and state agencies. Google Apps Cloud based data storage and IT services are available for purchase on the GSA schedules of several Google Apps resellers.[79] **Google Cloud Connect**[80] for MS Office enables multi-users collaboration and could be a good testbed for RDE and for ITS application development tools.
- **The Cloud Computing Interoperability Forum (CCIF)** hosted by Google Groups[81] is a group of industry stakeholders active in cloud computing interoperable platforms for application integration and stakeholder cooperation.

A7. Recent Articles on Federal cloud computing

- *"MerriTalk reports on the Status of Federal Cloud Computing"*, InfoTech article, April 25, 2011 at http://it.tmcnet.com/channels/cloud-computing/articles/ discusses the a study by VMWare: "Federal Applications Modernization Roadtrip: Express Lane, or Detour Ahead?" stated that:
  - 64% of federal CIOs expect to implement Cloud First within 2 years, to reduce costs and improve service
  - Current cost of IT federal legacy IT systems is $35.7B/year
  - $14,4B will be saved in the first year of Cloud implementation
  - Moving to Cloud will cut 30% of data center infrastructure expenses
  - 71% of CIOs and 66% IT managers see security concerns as top obstacle to cloud adoption. They believe that FedRAMP will not speed cloud computing adoption ort make Federal cloud computing more secure; agencies have to learn to use FedRAMP.

- "*The Federal Cloud Weather Report*" [82]is a survey of 167 federal CIOs and IT managers. Key findings are that few agencies are now in the Cloud:
  - 17% maintain IaaS; 15% SaaS; 13% PaaS services

[77] At http://www.accenture.com/us-en/Pages/insight-six-questions-energy-executives-cloud-computing-summary.aspx
[78] At http://www.google.com/apps/intl/en/government/index.html
[79] Posted at: www.apps.gov
[80] At http://www.google.com/apps/intl/en/business/officeconnect.html
[81] At http://groups.google.com/group/cloudforum/about
[82] By VMWare, April 18, 2011 at http://www.meritalk.com/pdfs/MeriTalk_Federal_Cloud_Weather_Report.pdf

- Of those, 64% are using private, internal cloud applications; and 18% use Hybrid-public plus private cloud
- 2 out of 5 wish to start with a single Cloud application, instead of enterprise-wide cloud strategy
- 54% of IT managers assert that mission-specific requirements hinder cloud adoption (budget constraints, security concerns, integration, staff shortage and training; culture issues
- Agencies are stuck at Go**:** 79% CIOs say they have not moved yet to Cloud First, but they are in discovery and planning stages to do so

o Recent Cloud outages including examples of loss of service, unpredictable cloud computing reliability, instances of slow diagnosis and recovery, and lessons learned for prevention and recovery are described at:

- http://searchcloudcomputing.techtarget.com/feature/Cloud-computing-outages-What-can-we-learn
- http://www.crn.com/slide-shows/cloud/231000954/the-10-biggest-cloud-outages-of-2011-so-far.htm
- http://www.crn.com/news/cloud/index/cloud-outages-cloud-services-downtime.htm
- http://money.cnn.com/2011/04/21/technology/amazon_server_outage/index.htm
- http://www.datacenterknowledge.com/archives/2011/10/07/outages-altering-cloud-perception-practice/

APPENDIX B: SUPPORTING DOT POLICIES AND GUIDANCE

RDE data management is covered by national policy and guidance concerning the handling and management of Big Data (BD), Cyber-physical security, Software productivity, and wireless spectrum usage must adhere to federal policies. The following lists of supporting policies and guidance.

B1. DOT/CIO Orders and Guidelines

- DOT CIO Policies and guidance on Open Government, eGov, Digital Transportation Exchange (DTE) and Data Centers Consolidation Initiative (DCCI) at http://www.dot.gov/open/plan/index.html
- "*USDOT Open Government Plan: April 2010-April 2012*", Version 1.2, June 25, 2010" at www.dot.gov/open/pdf/DOT_Open_Gov_Plan_V1.2_06252010.pdf
- "*2011 Data Center Consolidation Plan & Progress Report*", 9.30.2011at www.dot.gov/cio/docs/dot-fdcci-plan.pdf
- "*Digital Transportation Exchange (DTE): An Open Government Initiative*" posted at http://www.dot.gov/open/DTE
- DOT "**Information Resources Management Strategic Plan**" at **www.dot.gov/cio/docs/IRM_StrategicPlanFY2007-2012.pdf**
- *"DOT Cybersecurity Strategic Plan*", June 2010 at http://dotnet.dot.gov/technology/tech/docs/Issue-2010-JUN.pdf
- The "VisualDOT" initiative is a Cloud-based Data Visualization of geospatial transportation information system resulting from the OpenGov Directive at http://www.dot.gov/cio/visualdot.html

B2. Data Management Plans and Data Sharing Policies

Open Data policy documents and plans include:

- "*DOT Interim Identification and Prioritization Guidelines v0 1" at* www.dot.gov/open/pdf/identpriorguidelines1.0.pdf , June 2010 by the Data.gov Working Group
- "*DOT Open Government Plan*", Version 1.2 posted at http://www.dot.gov/open/plan/ addresses DOT data release and visualization policy.
- "*DOT Data Inventory Release*", Sept 30, 2010 at www.dot.gov/open/data/
- "*CIOP Chapter 34-Departamental Data Release Policy*", March 2011 posted at http://assets.sunlightfoundation.com.s3.amazonaws.com/policy/papers/DOT%20Order%201351.34.pdf

- "*Draft Preliminary Plan for Implementation of the President's Memorandum, Jan. 18, 2011- Regulatory Enforcement and Compliance Data*", May 2011 at http://regs.dot.gov/enforcementandcompliancedata.htm
- The FHWA Real Time System Management Information Program at www.ops.fhwa.dot.gov/1201/ has issued an *Interim Rule for Data Exchange and Data Formats*, and received stakeholders' comments on the proposed regulations: FHWA 23CFR Part 511, Real-Time System Management Information Program, see summary of public comments in the Federal Register, July 19, 2011 (Volume 76, Number 138, pp. 42536-42539).

B3. DOT CIO Open Data policy documents and plans

- "*DOT Interim Identification and Prioritization Guidelines v0 1" at* www.dot.gov/open/pdf/identpriorguidelines1.0.pdf , June 2010 by the Data.gov Working Group
- "*DOT Data Inventory Release*", Sept 30, 2010 at www.dot.gov/open/data/
- "*CIOP Chapter 34-Departamental Data Release Policy*", March 2011 posted at http://assets.sunlightfoundation.com.s3.amazonaws.com/policy/papers/DOT%20Order%201351.34.pdf
- "*Draft Preliminary Plan for Implementation of the President's Memorandum, Jan. 18, 2011- Regulatory Enforcement and Compliance Data*", May 2011 at http://regs.dot.gov/enforcementandcompliancedata.htm

B4. USDOT Reports on Connected Vehicle Data Environments (DCM and DMA)

- "State-of-the-Practice Policies and Lessons Learned on Open Data and Open Source" (March 2012 draft)
- "Concept of Operations: Data Capture and Management Research Data Exchange" (August 2011 draft)
- "Policy and Institutional Issues Analysis for the Dynamic Mobility Applications (DMS) Open Source Applications Development Portal (OSADP)" (Oct. 2011 draft)
- "Real-Time Data Capture and Management Program Vision: Objectives, Core Concepts and Projected Outcomes."
- "Identification of Critical Policy Issues for the Data Capture and Management (DCM) and Dynamic Mobility Applications (DMA) Programs" (March 2012 draft)
- "Connected Vehicle Environment: Governance Roundtable Proceedings from June 20, 2011", FHWA-JPO-11-129 (Aug 2011)
- "Real-time Data Capture and Management State of the Practice Assessment and Innovations Scan- Guidelines for Selecting a Cloud Provider" by SAIC, Delcan and UVA (Nov 18, 2011); and Overview presentation by Mohammed Yousuf, FHWA Office of Operations R&D, Sept 26, 2011
- "*Metadata Guidelines for the Research Data Exchange*", Noblis draft report for the ITS Mobility Program (Nov 16, 2011)

Appendix C: Managing a Data Ownership Policy

David Loshin is frequently quoted with regard to data ownership policies and data governance. His advice on managing a data ownership policy comes from his research, as described in *Enterprise Knowledge Management: The Data Quality Approach, by David Loshin, from Morgan Kaufmann, a division of Elsevier (Copyright 2001).* Loshins' advice is to describe the following features.[83]

1. The senior level managers support for the enforcement of the policies
2. All data sets covered under the policy
3. The ownership model (in other words, how is ownership allocated or assigned within the enterprise) for each data set
4. The roles associated with data ownership (and the associated reporting structure)
5. The responsibilities of each role
6. Dispute resolution processes
7. Signatures of those senior level managers

Loshin offers a template for describing the ownership policy for a specific data set:

Data Set Name				
Primary Owner				
Data Set Location				
	Owner	Responsible party	Reports to	Notes
Data definition				
Access/Definition				
User support				
Data packaging				
Data delivery				
Maintenance				
Data quality				

[83] http://searchdatamanagement.techtarget.com/feature/Data-governance-Information-ownership-policies-and-roles-explained

Business Rules				
Metadata				
Standards management				
Supplier management				

And offers the following steps for defining a data ownership policy:

1. Identify the interested parties or stakeholders associated with the enterprise data. This includes identifying the senior level managers that will support the enforcement of the policy.
2. Catalog the data sets that are covered under the policy.
3. Determine the ownership models in place and whether these are to continue or will be replaced.
4. Determine the roles that are and are not in place. Assign the responsibilities to each role, and assign the roles to interested parties.
5. Maintain a registry that keeps track of policies, data ownership, roles, responsibilities, and other relevant information.

Appendix D: Examples – Rules of Conduct

The following sites offer variations on establishing rules of conduct:

- European Union Smart Cities Code of Conduct: http://eu-smartcities.eu/code_of_conduct. This site includes examples of setting expectations for following policies on data ownership and licensing.
- The Oncology Portal: https://www.theoncologyportal.com/code-of-conduct. This site addresses issues of being anonymous with the portal and identifying and resolving conflicts of interest.
- Cork Institute of Technology: http://its.cit.ie/index.cfm/page/codeofconduct. This site provides a one-stop policy arena where users can find all policies that both guide their behavior as well as set up expectations and limitations from a systems perspective.
- Ubuntu: http://www.ubuntu.com/project/about-ubuntu/conduct. This site generically sets the boundaries on behavior.

Bibliography

Badger, Lee, Tim Grance, Robert Patt-Corner, and Jeff Voas. ***DRAFT Cloud Computing Synopsis and Recommendations***. National Institute of Standards and Technology, Special Publication 800-146, May 2011. Located at: http://csrc.nist.gov/publications/drafts/800-146/Draft-NIST-SP800-146.pdf.

Ball, Alex. ***How to License Research Data***. Digital Curation Centre in association with Joint Information Systems Committee, United Kingdom. Located at: www.dcc.ac.uk/resources/how-guides/license-research-data.

Ballard, Chuck, Nigel Davies, Marcelo Gavazzi, Martin Lurie, and Jochen Stephani. ***IBM Informix: Integration Through Data Federation: Managing Information and Protecting Development Assets, Integrating Heterogeneous Sources of Data, and Data Federation and Join Optimization***. IBM.com/Redbooks, August 2003. Located at: http://www.redbooks.ibm.com/redbooks/pdfs/sg247032.pdf.

Bureau of Transportation Statistics. ***DOT Report for Implementing OMB's Information Dissemination Quality Guidelines***. Report, August 2002. U.S. Department of Transportation. Located at: http://www.bts.gov/programs/statistical_policy_and_research/data_quality_guidelines/

Burns, Eugene M., Purificacion O. MacDonald and Amrut Champaneri. ***Data Quality Assessment Methodology: A Framework***. Bureau of Transportation Statistics, Joint Statistical Meeting – Section on Government Statistics., p. 334-337. 2000. Located at: http://www.amstat.org/sections/srms/proceedings/y2002/files/jsm2002-000347.pdf.

Cambridge Systematics Inc., Texas Transportation Institute, and Battelle Memorial Institute. ***Traffic Data Quality Measurement: Chapter 4 – Guidelines for Data Quality Measurement***. Report prepared for the U.S. DoT, September 15, 2004. Located at: http://ntl.bts.gov/lib/jpodocs/repts_te/14058_Files/chap4.htm.

Cambridge Systematics Inc. ***Assessing the Value of the ADOT&PF Data Programs***. White paper prepared for the Alaska Department of Transportation and Public Facilities. September 2009.

Cambridge Systematics Inc. ***Data Action Plan: Alaska Data Business Plan.*** White paper prepared for the Alaska Department of Transportation and Public Facilities. September 2009.

Cambridge Systematics Inc. ***Data Governance, Standards, and Knowledge Management***. White paper prepared for the Alaska Department of Transportation and Public Facilities. September 2009.

City of Vancouver. ***Open Data License: Catalogue Beta v2 – Terms of Use***. January 26, 2011. Located at: http://data.vancouver.ca/termsOfUse.htm.

Danish Agency for Culture. ***Interoperability of Digital Repositories***. The Knowledge Exchange, 2007. Located at: http://www.knowledge-exchange.info/Default.aspx?ID=290.

Data.gov Working Group. ***Data.gov Interim Identification & Prioritization Process and Guidelines v1.0***. U.S. Department of Transportation, June 2010. Located at: http://www.dot.gov/open/pdf/identpriorguidelines1.0.pdf.

Delcan, Science Applications International Corporation, and the University of Virginia. ***Real-Time Data Capture and Management State-of-the-Practice Assessment and Innovations Scan: Lessons from Scan of Current Practices***. Report for the U.S. Department of Transportation, March 7, 2011.

Department of Homeland Security. ***Privacy Policy Recommendations for Federated Information-Sharing System***. Data Privacy and Integrity Advisory Committee, DHS. 2011.

Department of Homeland Security and Department of Justice. ***Implementation Guidelines for the National Information Exchange Model (NIEM)***. Located at: http://it.ojp.gov/default.aspx?area=nationalInitiatives$page=1072.

Dowers, Robert C. and Laurence F. Pulgram. ***Effects of Recent Rulings on the Enforceability of Open Source Licenses***. Fenwick & West LLP (Mountain View, CA) as published in eBulletin, International Technology Law Association, Volume 3, Issue 3, June 2009. Located at: http://www.itechlaw.org/ebulletin/volume.asp?id=11.

Dunn, Peter (Purdue University), Todd Guttman (Ohio State University), Gunta Liders (University of Rochester, and Carol Blum (Council on Government Relations). ***Access to and Retention of Research Data: Rights and Responsibilities***. Council on Government Relations, March 2006. Located at:www.cogr.edu/viewDoc.cfm?DocID=151536.

Dutta, Soumitra and Irene Mia. ***The Global Information Technology Report 2008-2009: Mobility in a Networked World***. Copyright © 2009 by the World Economic Forum and INSEAD. Printed and bound in Switzerland by SRO-Kundig, Geneva.

Eaves, David. ***The State of Open Data in Canada: The Year of the License***. Eaves.CA blog discussion of licenses, February 16, 2011. Located at: http://eaves.ca/2011/02/16/the-state-of-open-data-in-canada-the-year-of-the-license/.

Federal Cloud Blog. ***DoD to Issue Commercial Cloud Policy Directive***. August 24, 2011. Located at: http://fedcloud.wordpress.com/tag/disa.

Forrester Consulting. ***The Next Maturity Step for Cloud Computing Management***. Technology Adoption Profile for Morph Labs Inc. Forrester Research Inc., August 2011.

French, Paul. *"Public or Private Cloud? – Balancing Security, Cost Savings and Efficiencies for Government Agencies".* Web article for Government Security News (GSN) Magazine, August 23, 2011. Located at: http://www.gsnmagazine.com/article/24302/public_or_private_cloud_balancing_security_cost_sa.

Gitter, Donna. ***The Challenges of Achieving Open-Source Sharing of Biobank Data***. Biotechnology Law Report, December 1, 2010. Pg. 623(13) Vol. 29 No. 6. Located at: http://www1.cuny.edu/mu/scholarship/2011/02/01/the-challenges-of-achieving-open-source-sharing-of-biobank-data/.

Govtech.com. *"California Federated Data Center Bringing Results, Officials Say."* Web article in Government Technology, May 2, 2011. Located at: http://www.govtech.com/policy-management/California-Federated-Data-Center-Bringing-Results.html.

Graux, Hans. ***Open Government Data: Reconciling PSI Re-Use Rights and Privacy Concerns***. European Public Sector Information Platform, Topic Report No. 2011/3, October 2011. Located at: http://epsiplatform.eu/sites/default/files/Topic_Report_Privacy.pdf.

Green, Ann, Stuart Macdonald, and Robin Rice. ***Policy-Making for Research Data in Repositories: A Guide***. Joint Information Systems Committee, United Kingdom. May 2009, Version 1.2. Located at: http://www.disc-uk.org/docs/guide.pdf.

Hewlett-Packard. ***Five Myths of Cloud Computing***. Business white paper, October 2009. Located at: http://www.hp.com/hpinfo/newsroom/press_kits/2011/HPDiscover2011/DISCOVER_5_Myths_of_Cloud_Computing.pdf.

Howard, Newton and Sergey Kanareykin. ***Analysis of Federated and Centralized Information Sharing Architectures***. White paper for the Center for Advanced Defense Studies, 2007. Located at: http://www.c4ads.org/sites/default/files/Federated%20vs%20Centralized.pdf.

ICT Pulse. Open source, open data: challenges and opportunities. Web proceedings of the Caribbean Open Data Conference, January 26, 2011. Located at: http://www.ict-pulse.com/2012/01/open-source-open-data-challenges-and-opportunities/.

IndraSoft, Inc. ***Research Data Exchange Architecture Analysis of Alternatives (Draft)***. White paper prepared for the U.S. Department of Transportation, February 24, 2012, Version 1.0.

Internet Governance Forum – Sixth Annual Meeting. ***SOP 110/123 Open Data: Challenges and Solutions/Public Sector Information Online: Towards a Global Policy Framework***. United Nations, September 27-30, 2011. Located at: http://www.intgovforum.org/cms/component/content/article/71-transcripts-/898-sop-110123-open-data-challenges-and-solutions--public-sector-information-online-towards-a-global-policy-framework-centre-science-development-a-media-studies-and-retired-sole-.

Jackson, Jana M. ***DoD Journey to the Cloud – DISA R.A.C.E. Private Cloud***. ViON Corporation, presentation on June 14, 2011. Located at: http://media.govtech.net/GOVTECH_WEBSITE/EVENTS/PRESENTATION_DOCS/2011/GTC_Southwest_2011/CloudComputingInGovernment_ViON.pdf

Jansen, Wayne and Timothy Grance. ***Guidelines on Security and Privacy in Public Cloud Computing***. National Institute of Standards and Technology, Draft Special Publication 800-144, January 2011. Located at: http://www.nist.gov/customcf/get_pdf.cfm?pub_id=909494.

Kash, Wyatt. *Open Development Delivers Unexpected Benefits".* Government Computer News, web article, November 10, 2010. Located at: http://gcn.com/articles/2010/11/15/interview-rob-vietmeyer-forge-mil.aspx?sc_lang=en.

Kundra, Vivek. ***Federal Cloud Computing Strategy***. Presentation on February 14, 2011 at the Cloud Security Alliance. Located at: http://www.cio.gov/documents/Vivek-Kundra-Federal-Cloud-Computing-Strategy-02142011.pdf.

Kundra, Vivek. ***Federal Cloud Computing Strategy***. Strategic report, February 8, 2011. White House. Located at: http://www.cio.gov/documents/federal-cloud-computing-strategy.pdf.

Lieberman, Danny. *" Best Ways for Businesses to Prevent Data Breaches".* Web article for Infosec Island, February 1, 2012. Located at: http://www.infosecisland.com/blogview/19383-Best-Ways-for-Businesses-to-Prevent-Data-Breaches.html.

Martinelli, Fabio, Marinella Petrocchi, Ilaria Matteucci, Alexey Orlov, Dmitry Starostin, Marco Luca Sbodio, Alvaro Arenas, Benjamin Aziz, Shirley Crompton, and Michael Wilson. ***Methodologies and Tools for Data Sharing Agreements Infrastructure***. Report for the Consequence Consortium, 2008.

McGurrin, Mike, Anna Corrigan, James Larkin, and Karl Wunderlich. ***Concept of Operations: Data Capture and Management Research Data Exchange***. Report for the U.S. Department of Transportation, August, 2011.

Morph Labs. ***Controlling Cloud Resources with mCloud***. White paper, Morph Labs, Inc. April 2011.

Morph Labs. ***The Revolution of Commodity Cloud: How Commodity Hardware and Software Impacts IT Procurements***. White paper, Morph Labs, Inc. October 2011.

Nance, Richard E. ***Distributed Simulation with Federated Models: Expectations, Realizations, and Limitations***. Proceedings of the 1999 Winter Simulation Conference pages 1026-1031. Located at: http://www.informs-sim.org/wsc99papers/149.PDF.

The National Archives, *"Open Government License for Public Sector Information".* Web information located at: http://www.nationalarchives.gov.uk/open-government-license/

National Association of State Chief Information Officers (NASCIO). ***Capitals in the Clouds: The Case for Cloud Computing in State Government Part I: Definitions and Principles and Part III Recommendations for Mitigating Risks: Jurisdictional, Contracting, and Service Levels***. NASCIO Governance Series, June 2011. Located at: http://www.nascio.org/publications/documents/NASCIO-Capitals_in_the_Clouds-June2011.pdf and at http://www.nascio.org/publications/documents/nascio_cloudcomputing_partiii.pdf.

National Association of State Chief Information Officers (NASCIO). ***Data Governance Part III: Frameworks-Structure for Organizing Complexity***. NASCIO Governance Series, May 2009. Located at: http://www.nascio.org/publications/documents/NASCIO-DataGovernancePTIII.pdf.

National Association of State Chief Information Officers (NASCIO). ***Negotiating IP on the Way to the Win-Win: NASCIO's Intellectual Property Recommendations.*** NASCIO Governance Series, 2005 at: http://www.nascio.org/publications/documents/NASCIO-negotiatingIP.pdf.

National Statistical Service. ***A Good Practice Guide to Sharing Your Data with Others***. Version 1, November 2009. Located at: http://nss.gov.au/nss/home.nsf/NSS/E6C05AE57C80D737CA25761D002FD676?opendocument.

Nolle, Tom. *"Cloud Computing and Mobile Behavior: A New Services Opportunity."* Web article for searchCloudProvider.com, April 2, 2012. Located at: http://searchcloudprovider.techtarget.com/tip/Cloud-computing-and-mobile-behavior-A-new-services-opportunity.

Parsons, Mark. ***Expert Report on Data Policy – Open Access***. Global Research Data Infrastructures, University of Colorado. October 2011. Located at: http://www.grdi2020.eu/Repository/FileScaricati/e31a1aab-b01e-4e7e-9b10-0fd93d4b710f.pdf.

Perens, Bruce. *"Open Standards, Principles and Practice".* Web article, October 13, 2010. Located at: http://perens.com/OpenStandards/Definition.html.

Repositories Support Project. ***Setting Up a Repository: Planning Checklist, Resourcing for Sustainability, and Policies and Legal Issues***. Joint Information Systems Committee, United Kingdom. Located at: http://www.rsp.ac.uk/.

Research Data Strategy Working Group. ***The 2011 Canadian Research Data Summit: Mapping the Data Landscape.*** Proceedings, Ottawa Convention Centre, September 14-15, 2011. Located at: http://rds-sdr.cisti-icist.nrc-cnrc.gc.ca/docs/Summit_Backgrounder.pdf.

Riegle, Robert. ***Privacy Impact Assessment for the Department of Homeland Security State, Local, and Regional Fusion Center Initiative***. Department of Homeland Security, December 11, 2008. Located at: http://www.dhs.gov/xlibrary/assets/privacy/privacy_pia_ia_slrfci.pdf.

Riley, George F., Mostafa H. Ammar, Richard M. Fujimoto, Alfred Park, Kaylan Perumalla, and Donghua Xu. ***A Federated Approach to Distributed Network Simulation***. Mendeley, Volume 14, Issue 2, Pages 116-148.

Savitz, Eric. *"How Cloud Computing Can Boost Development Nations."* Web article in Forbes, November 3, 2011. Located at: http://www.forbes.com/sites/ciocentral/2011/11/03/how-cloud-computing-can-boost-developing-nations/.

Science Applications International Corporation (SAIC), Delcan Corporation, and the University of Virigina. ***Real-Time Data Capture and Management State-of-the-Practice Assessment and Innovations Scan: Guidelines for Selecting a Cloud Provider***. White paper developed for the U.S. DOT, November 18, 2011.

Science Applications International Corporation (SAIC), Delcan Corporation, and the University of Virigina. ***Real-Time Data Capture and Management State-of-the-Practice Assessment and Innovations Scan: Recommendations Plan for Data Environment Development and Management***. White paper developed for the U.S. DOT, September 14, 2011.

Shackelford, Dave. *" Network Virtualization in the Cloud: Managing Virtualization Security Risks"* and "*Virtualization 101: Best Practices for Securing Virtual Machines"*. Web articles for SearchSecurity.com. January 13, 2012 and April 12, 2012. Located at: http://searchcloudsecurity.techtarget.com/tutorial/Network-virtualization-in-the-cloud-Preventing-virtualization-security-risks and http://searchsecurity.techtarget.com/magazineContent/Virtualization-security.

Shook, William A. ***GAO Decision: Matter of Technosource Information Systems, LLC; TrueTandem LLC***. Government Accountability Office, October 17, 2011. Located at: http://www.gao.gov/decisions/bidpro/405296.htm.

Smith, Michael A., Andrew J. Schain, Kendall Grant Clark, Arlen "Ken" Griffey, and Vladimir Kolovski. ***Mother, May I? OWL-Based Policy Management at NASA***. Penn State University. Located at: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.85.5881.

Talend Open Data Solutions. ***Open Source Data Management in the Public Sector***. White Paper, 2010. http://www.talend.com.

U.S. Department of Transportation. ***DOT Order 1351.34: Departmental Data Release Policy, Chapter 34***. Located at: http://regs.dot.gov/docs/DOT%20Draft%20Enforcement%20and%20Compliance%20Data%20Report%20-%2005-18-2011%20-%20OCR.pdf.

U.S. Department of Transportation. ***Data Capture and Management Program: Transforming the Federal Role***. White paper, May 2010. Located at: http://www.its.dot.gov/data_capture/datacapture_management_federalrole7.htm.

U.S. Department of Transportation. ***Data Center Consolidation Plan***. Report, Version 7, September 30, 2011. Located at: http://www.dot.gov/cio/docs/dot-fdcci-plan.pdf.

U.S. Department of Transportation. ***Open Government Plan, April 2010-April 2012***. Report, Version 1.2, June 25, 2010. Located at: http://www.dot.gov/open/pdf/DOT_Open_Gov_Plan_V1.2_06252010.pdf.

U.S. Department of Transportation. ***Real-Time Data Capture and Management Program Vision: Objectives, Core Concepts, and Projected Outcomes***. White paper, April 2010. Located at: http://www.its.dot.gov/data_capture/datacapture_management_vision1.htm.

Vandervalk, Anita/Cambridge Systematics Inc. ***Real-Time Data Capture and Management Evaluation and Performance Measures – Evaluation Framework***. Report for U.S. Department of Transportation, September 1, 2011. FHWA-JPO-11-136.

Vandervalk, Anita/Cambridge Systematics Inc. ***Data Capture and Management: Needs and Gaps in the Operation and Coordination of the U.S. DOT Data Capture and Management Programs***. White paper prepared for the U.S. Department of Transportation, November 2010. FHWA-HOP-11-004.

Wells, Garth. *"Using DTS to Automate a Data Import Process"*. Web article for SQL Team.com. August 11, 2002. Located at: http://www.sqlteam.com/article/using-dts-to-automate-a-data-import-process.

Yau, Nathan. *"Data.gov in Crisis: The Open Data Movement is Bigger than Just one Site"*. Electronic article in The Guardian, April 5, 2011. Located at: http://www.guardian.co.uk/news/datablog/2011/apr/05/data-gov-crisis-obama.

Yanosky, Ronald. ***Institutional Data Management in Higher Education***. Educause Center for Applied Research, Research Study 8, 2009. Located at: http://net.educause.edu/ir/library/pdf/ers0908/rs/ers0908w.pdf.

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

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

FHWA-JPO-12-031

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