Data Governance Best Practices and Recommendations Report

Transportation System Management and Operations (TSMO) Vision and Regional Intelligent Transportation Systems (ITS) Architecture Update

Final — May 24, 2019

Prepared for

Prepared by

With Support from Kimley-Horn, ConSysTec, and Lumenor



relay
GREEN
MACHINE

EXPRESS LANE
Pass
ONLY
TO
Indian Trail-
Lilburn Road
TO
Lawrenceville
Suwanee Rd


Version Control

Version	Date	Description	Editor / Reviewer
0.1 DRAFT	8 April 2019	Original version	ICF
0.2 DRAFT	15 April 2019	First internal review	ICF Team
0.3 DRAFT	17 April 2019	Additional review and comments	ICF Team
0.4 DRAFT	18 April 2019	Draft Deliverable	ICF Team
1.0 Final	24 May 2019	Final Deliverable; incorporated results from FHWA Data Business Plan Workshop	ICF Team

Table of Contents

List of Figures

List of Tables

1. Introduction

1.1. Scope

This white paper on Best Practices for Data Governance (DG) explores industry recommendations on the purpose, benefits, and strategies related to applying data governance to the role of the Atlanta Regional Commission (ARC) with respect to "data". The area of data governance is extensive, with every enterprise architecture and information technology business analysis methodology promoting a framework for people, processes, and technologies to manage their data.

This paper will introduce the various topics related to data governance best practices, providing an extensive bibliography, however, the paper attempts to focus the reader on the role of ARC in fostering good data governance best practices rather than implementing regional data management processes.

1.2. Background

Effective transportation systems management and operations (TSMO) at a regional scale involves coordination among a wide array of partners -- including agencies involved in operating highways, transit services, and emergency response services, as well as the private sector – to optimize system performance.

Data sharing for enhanced situational awareness is key to implementing many TSMO strategies; moreover, sharing real-time data and predictive analytics with the public and private sectors plays an important role in influencing travel decisions. High-quality, consistent data that is managed over time and throughout its life is the foundational element of multi-agency, regional approaches to TSMO.

At the same time, public agencies operate with limited and shrinking resources, changing technological landscapes, and shifting roles and expectations. Data is becoming a major asset and investment. To manage these assets, government needs to become

- responsive with the ability to transform data into information and decisions;
- support interoperability in order to share information;
- effective and efficient data custodians to manage data discovery and access; and
- a trusted source to manage data quality and privacy.

To that end, key data should be planned and managed to support the enterprise rather than just a project, which is currently what is often done. Data governance, therefore, is becoming increasingly important for organizations and overall systems of organizations that work together. Many organizations that adopt data governance practices recognize the need to undergo a cultural transformation, changing the ways individuals and systems handle and process data. TSMO strategies also require cultural change, thus, there may be no better time for ARC and its stakeholders to adopt data governance practices then now.

1.3. Document Organization

This document describes best practices in data governance as applied by transportation organizations. The document is organized as follows:

Section 2 ARC Challenges and Data Governance Benefits. Section 2 introduces the benefits of data governance given current challenges in data access and exchange in the ARC region.

Section 3 Data Governance Overview. This section describes data governance, its definition and framework components. Many organizations claim to be the authority over the best practice. This section summarizes and identifies the commonalities of the various methodologies.

Section 4 Data Governance Framework: Business Strategies and Organization. Section 4 describes the various data governance framework components and how transportation organizations implement them. The components include goals and objectives, policies, organizational models and maturity models.

Section 5 Data Lifecycle Management. Section 5 describes an overview of the data lifecycle and categories of plans and procedures that are included to curate data over if life.

Section 6 Changing Needs in Transformative Transportation Environments. This section describes future challenges with changing and emerging transportation technologies and strategies. Detailed topics include impacts due to integrated transportation management systems, mobility on demand and accessible travel, and automated vehicles.

Section 7 Lessons Learned from Government Initiatives. Section 7 describes the approach used by transportation agencies to get started – what drives them to adopt data governance frameworks, how they get started, and steps recommended by the USDOT to set up a data governance framework.

Section 8 ARC's Role in a Regional Data Governance Framework. Section 8 provides a set of recommendations for ARC to initiate data governance for regional constituents and stakeholders (both internal and external). The section includes recommendations for ARC's role and responsibilities, as well as plans and artifacts needed to promote good data management practices for the region.

Section 9 End Notes. This section provides notes and references for citations contained in the report.

Appendices. The appendices section includes Appendix A Acronym Table and Appendix B -- the results from the data exercises conducted during Workshop #2 (2019 March 18).

Bibliography. The Bibliography, though providing the full references for the end notes, provides example documents that ARC can use to model framework components. In particular, the Data Business Plan (Hillsborough…) describes

2. ARC Challenges and Data Governance Benefits

In the ARC TSMO Vision and Intelligent Transportation System (ITS) Regional Architecture project workshop #2, participating agencies were asked about their major challenges with sharing data. From among the 53 participants, organizations identified many challenges when it comes to collecting, analyzing and sharing data, many of which can be addressed by establishing data governance policies, procedures, and standards. A robust data governance (DG) framework directly addresses these issues by defining/providing the right procedures, standards, and policies to manage data. In this sense, establishing a DG framework will increase data interoperability, quality, sharing and effectiveness as well as reduce costs.

In the workshop, when asked to describe their three major challenges when sharing data with other organizations, five common themes emerged from participants' responses. The top five issues identified by the stakeholders, along with insight on how adapting a data governance framework helps, are described below.

Challenge #1: Inconsistent access / Challenges to access (platform) / Data discovery. Agencies and stakeholders have different data sharing platforms with varying levels of access/security, which yields inconsistent access to data and even inconsistent data across stakeholders—i.e., no clear guidance on how to expose what data exists and which organization or department has it. The challenge is both knowledge-

oriented and technology-related: What data is available? From who? Where is it stored? How often is it updated? And how can it be accessed without major technology issues?

- *DG provides "rules of engagement" which describe user- and owner-roles, access procedures, and methods of exposing and describing data sets irrespective of technology or platform. The rules of engagement promulgated by a data governance framework helps improve sharing efficiencies through the adoption of: (1) data discovery services to support searching for data across multiple organizations; and (2) technology-agnostic and role-based data access methods.*

Challenge #2: Inconsistent structures, formats, and semantics. Current systems have incompatible data descriptions that makes it difficult to understand data in detail (e.g., type, meaning, scale, temporal, coverage, estimated vs. observed) and to integrate it into their systems.

- *DG framework provides rules and guidelines on how to describe, organize, and share data. This ensures that all data is collected, named, defined, and grouped consistently and according to standards across all stakeholders, including vendors. This standardization also helps with any future system/software integration, as agencies now operate using a consistent data organization and structure. Finally, this standardization facilitates aggregating data to provide key performance metrics in an efficient manner.*

Challenge #3: Unclear data responsibility. Currently, there is limited organizational structures that specify management, accountability, and audit responsibilities for data. These details include accountability for upkeep, quality, description, and dissemination to downstream users, including sharing information to other stakeholders. Data curation is often overlooked once data is collected or an application is deployed. Maintaining information about data quality, lineage, point of contact, or storage location may not be maintained because there is no one assigned to manage the data.

- *DG framework identifies the need to define the roles and responsibilities of data owners, stewards, and users of data over its lifecycle. A common theme of data governance is the relationship of people to data: who is responsible for data curation, who is responsible for ensuring the data serves enterprise needs, who is accountable for the quality and access to the data, who owns the data, how is the data used, specifically as it relates to privacy issues. As such, a DG framework designates roles such as data stewards, data custodians, data policy committees, and data champion as well as the responsibilities assigned to each.*

Challenge #4: Data restrictions. The lack of defined data ownership and rights to data also leads to unclear data distribution and use, that is, what can or cannot be shared due to contracting agreements or licensing restrictions?

- *By articulating data policies for sharing and use, a DG framework clarifies the distribution and privacy rules for requests made by internal and external stakeholders. As such, DG helps in setting clear relationships to manage shared data and information exchanges among internal and external stakeholders, addressing any policy or legal limitations for sharing data.*

Challenge #5: Limited and costly resources to manage data. There is a vast amount of data being collected and processed for static and real time consumption. Collecting, managing, and distributing this data requires resources, including human, that may exceed the financial capabilities of the stakeholders.

- *A common feature of most DG frameworks involves the development of a data business plan that*
    - *prioritizes critical data needs,*
    - *identifies redundancies in data collection, processing and storage,*
    - *develops strategies for migrating manual collection and quality control to automated processes, and frames organizational responsibilities for data stewards who role is to manage data for the enterprise*

Additional challenges were articulated by workshop participants include:

- Data needs with respect to interfaces and quality that will support my objectives and outcomes
- Privacy issues and policies
- Geographic data inconsistencies

A complete set of the challenges as well as current and future data sharing needs are included in Appendix B: Workshop #2 Data Discussion .

3. Data Governance Overview

3.1. Data Governance Defined

Data governance as a discipline has been part of enterprise architectures and information technology (IT) processes since the early 1980s and defined by many groups. Depending on the purpose, different definitions of data governance focus on specific core values that are critical to that group, for example:

MDM Institute[1] defines data governance as:

> "the formal orchestration of people, processes, and technology to enable an organization to leverage data as an enterprise asset"
>
> ***With a general focus on people, processes, and technology.***

Forrester[2] defines data governance as:

> "A strategic business program that determines and prioritizes the financial benefit data brings to organizations as well as mitigates the business risk of poor data practices and quality. At the heart of this program is ownership, accountability, processes, planning, and performance management. "
>
> ***With a focus on the fiduciary responsibility and organizational planning for managing data.***

Data Governance Institute[3] defines data governance as:

> "a system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models, which describe who can take what actions with what information, and when, under what circumstances, using what methods."
>
> ***With a focus on mid-level manager responsibilities and rules of engagement.***

NASCIO[4] defines data governance as:

> "the operating discipline for managing data and information as a key enterprise asset. This operating discipline includes organization, processes and tools for establishing and exercising decision rights regarding valuation and management of data. Key aspects of data governance include decision making authority, compliance monitoring, policies and standards, data inventories, full lifecycle management, content management, records management, preservation, data quality, data classification, data security and access, data risk management, and data valuation."
>
> NASCIO reframes the definition to cover "information or knowledge management governance". In this environment of social media, big data, and unstructured data, the renaming may be appropriate.
>
> ***With a focus on rules of engagement and operating principles for managing data.***

To more fully appreciate what data governance is, it is best to understand what it is not. According to *Oracle Best Practices in Data Governance* (2011) and Forrester, Data Governance *is not* data management or administration, data cleansing, master data management, or data storage/warehouse.

The common, recurring theme from the various DG definitions may be summarized as the

> Rules of engagement for how institutions (people and policies) manage and sustain data across the enterprise, over its lifecycle[5].

Enterprise in this context includes an organization and internal and external stakeholders. For the remainder of this document, these common themes represent data governance.

3.2. Data Governance Frameworks

A DG framework describes how all the pieces that compose data governance fits together. According to NASCIO,

> "frameworks [in general] assist in describing major concepts and their interrelationships. Frameworks assist in organizing the complexity of a subject. Frameworks facilitate communications and discussion. All of these descriptors apply as well to frameworks related to data governance. Additionally, data governance frameworks assist in demonstrating how data governance relates to other aspects of data management, data architecture, and enterprise architecture."

The Florida DOT (FDOT) Reliable, Organized, Accurate, Data Sharing (ROADS) Project Data Governance Overview presents a simplified relationship among the aspects of the framework as shown below in Figure 1. On the left side of the figure are the people (roles) associated with the framework, and the right side lists high-level responsibilities and processes.

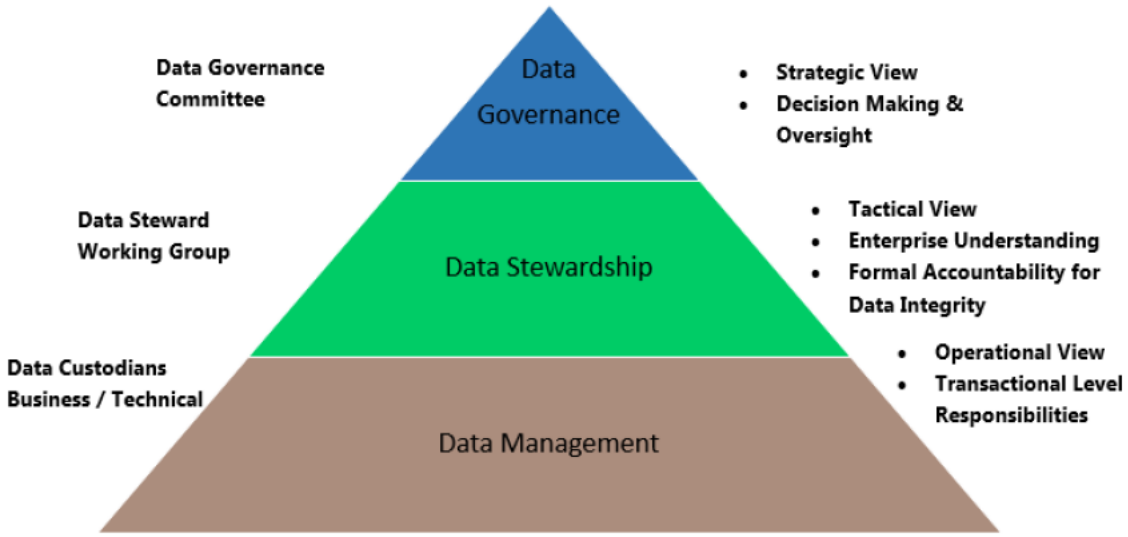

Figure 1: FDOT ROADS Project Data Governance Overview

Data Management Association International (DAMA), a formal data governance organization, published a data management body of knowledge (DMBOK2, published July 2017) that provides detailed and comprehensive context diagrams that include goals for each objective; business and technical drivers; activities and roles; and inputs and outputs. An example of one of these context diagrams is illustrated in Figure 2.

Figure 2: Data Governance and Stewardship Context Diagram from DMBOK2

The DAMA DG knowledge area cites four major objectives:

- Data Governance and Stewardship
- Business Cultural Development
- Data in the Cloud
- Data Handling Ethics

Each objective has its own context diagram much like the one shown in Figure 2.

Generally, IT Governance and enterprise architecture methodologies and tailored DG frameworks incorporate different aspects of data governance; however, there are consistent recurring themes throughout these various models, including the following characteristics:

- Accountability and leadership roles in organization
- Planning and rules for data handling – quality, integrity, access
- Strategic enterprise perspective
- Cultural change to a data-centric organization

The following sections describe critical aspects of the DG Framework.

4. Data Governance Framework: Business Strategies and Organization

4.1. Overview

Initiating a DG framework is similar to developing a strategic plan. It starts with articulating a vision as well as objectives and goals for managing, sharing, and accessing data. In the case of ARC and its stakeholders, DG includes managing and sharing information across organizations. For ARC, the "enterprise" consists of many transportation and planning organizations within the region, with each organization responsible for collecting, managing, and curating the same or similar data, allocating resources and applying their own policies and procedures to data curation activities. The data enterprise, in the ARC region, is a distributed, heterogenous environment. Most data governance frameworks assume a single organization. Transportation agencies have adapted the enterprise DG framework to extend to a multimodal, multi-jurisdictional environment, one in which ARC can play a pivotal role. The regional data governance framework comes from aligning the elements of the framework – goals, objectives, policies, procedures, organization with other regional visions such as the TSMO vision, goals and objectives. This section introduces elements of the DG framework and identifies methods to extend these elements to fit a regional DG model.

4.2. Data Governance Goals and Objectives

Data governance goals and objectives are derived from organizational vision, goals and objectives. Some emerge due to a major challenge or as an initiative to support another initiative. For example, state DOTs recognize that data governance is essential to developing and coordinating asset management systems, feature layers and linear referencing systems for their Geographic Information Systems. Initiating data governance on a project basis and using the project as a platform to expand to other domains has worked for many organizations. To that end, DG goals and objectives tend to focus on the problems encountered as well as good strategic planning practices. In ARC's TSMO Visioning Workshop Summary[6] many goals tended to focus on sharing data not only between public sector organizations but also between "public and data providers and users."[7] Goals tended to identify areas such as data integration, access and quality. These are typical goals and objectives described by DG frameworks. Examples of goals and objectives cited by transportation agencies are included below.

In its DG Primer, the Federal Highway Administration (FHWA) identifies a sample set the goals and objectives for data governance (FHWA, 2015). The goals and objectives are listed in Table 1.

Table 1: FHWA Data Governance Goals and Objectives

Goal	Objectives
Leadership – Champion data solutions to ensure accountability and increase the value of data assets.	• Promote data governance within FHWA. • Communicate data-related changes to all interested parties. • Monitor progress and ensure accountability of data governance tasks and projects.
Quality – Oversee efforts to provide acceptable quality data that is accurate.	• Establish a Data Quality Assurance Program. • Increase the accuracy and clarity of data. • Improve accessibility of data.

Prioritization – Prioritize efforts to address data gaps and needs.	• Establish clear priorities to address data gaps and needs. • Communicate priorities to FHWA business units.
Cooperation – Facilitate cross-organizational collaboration, data sharing, and integration.	• Increase opportunities for data sharing. • Eliminate data silos and other barriers. • Ensure business units know the identity of Data Stewards. • Ensure Data Stewards know the identity of Data Users.
Flexibility – Encourage creative and innovative solutions to data needs.	• Identify innovative data solutions throughout FHWA. • Communicate innovative solutions to Data Stewards and Data Users.
Utilization – Improve data utilization and ease of access.	• Promote appropriate data usage throughout FHWA. • Provide staff the means to determine the extent and availability of FHWA data.

Other organizations, such as Colorado and MnDOT address objectives such as:

- Build a culture of data cooperation by involving all organization members in data collaboration (knowledge, access, accountability, use)
- Promote knowledge of data and reduce risk
- Develop guidelines that incorporates managing information value and reducing risk
- Understand and measure benefits of data management practices
- Involve business and IT in procurement decisions that incorporate data value

Data Governance Principles

Similar to objectives, though stand-alone, are data governance principles. Principles statements are values used to guide organizations with their priorities. When unforeseen issues occur, principles provide the needs and ideals that drive decisions. To guide priorities for data governance, organizations may develop a set of principles that focus on their values. For example, as seen in Table 2, Colorado Data Organization (CDO)'s focus is on business strategies for data governance, while Minnesota DOT's (MnDOT's) principles are data centric. The principles are similar in that they address interoperability, reusability, consistency, and accountability. Florida DOT included a set of principles in the DG project charter while MnDOT and CDO included principles in their Data Management Plan.

Table 2: Data Governance Principle Focus

Business-Oriented Principles	Data-Centric Principles
The CDO is a multidiscipline function to further an end state through strategies, policies, governance, architecture and collaboration that: • Advances data awareness, discoverability, accessibility, and utilization • Fosters opportunities to integrate existing data sources with new data sources, and third-party data • Facilitates re-usable, consistent, and repeatable exchange of data between	MnDOT has adopted the following principles to better govern data. All decisions related to data should align with the principles[8]. • Data shall be managed as a state asset • Data quality fits its purpose • Data is accessible and shared as permitted • Data includes standard metadata • Data definitions are consistently used

systems via a data integrate layer and web services • Advances interoperability and data sharing, breaks down silos • Inspires innovation and create and cooperative problem solving; maximizes business insight through optimizing utilization of data • Creates a dynamic, curious, data driven environment inclusive of big data, artificial intelligence, predictive modeling, deep learning, and more	• Data management is everyone's responsibility • Data shall not be duplicated

4.3. Policies

Policies support oversight and compliance to standard operating practices (SOP). Some organizations tend to develop broad-based policies while, some are detailed and structured, such as a SOP. FDOT developed broad-based policies that incorporated their data governance objectives[9]. The FHWA Data Governance Primer (FHWA, 2015) shows a detailed set of policies that cover every aspect of the data governance organization, processes, and procedures; these policies are similar to data needs and corresponding data management requirements (as shown in Table 3).

Table 3: FHWA Data Governance Policies [source: FHWA, 2015]

Policy	Description
FHWA data are an enterprise asset.	Data, structured and unstructured, and the corresponding metadata, are business and technical resources owned in whole or in part by FHWA. FHWA data include shared data about managed entities, interests, finances, employees, resources, customers, providers, business affiliates, best practices, operating procedures, experimental results, etc. All employees must recognize that the proper management of strategic enterprise data is critical to the success of the organization.
FHWA data programs and activities must undergo IT investment process.	FHWA data programs or data related activities within IT projects require Investment Review Board (IRB) approval prior to and during an ongoing effort. This process is typically initiated, liaised, communicated to IT project managers, or executed by the Data Stewards. They are ultimately responsible for following the FHWA Information Technology Investment Process in order to gain IRB approval prior to and during all planned/ongoing data activities.
FHWA data must be consistent	All strategic FHWA data shall be modeled, named, and defined consistently, according to standards, across the organization. Efforts must be made by management to share data and not maintain redundant data without justification. Originating business stewards of data must recognize the informational needs of downstream processes and business units that may require FHWA data.
FHWA data must be of acceptable quality	Quality data are critical to ensuring FHWA mission success. Data Stewards are responsible for ensuring that FHWA data are accurate and correct for the intended purpose and use, and that data providers follow all reporting requirements regarding the collection, processing, and reporting of FHWA data, and meet all requirements of

	the Data Quality Act. Data quality standards shall be managed and applied actively to the approved reliability levels of FHWA data as defined by the business owners.
FHWA data must be interoperable with dependent systems	All enterprise data (structured and unstructured) must conform to a common set of standards and schemas across all data sharing parties. Data sharing must also be accounted for and facilitated through a designated authority.
FHWA data must be maintained at the source	All FHWA data must be maintained as close to the source as feasible, to reduce the collection and storage of redundant data.
Enterprise data must be safe and secured	FHWA data, in all electronic formats, shall be safeguarded and secured based on recorded and approved requirements and compliance guidelines. These requirements are to be determined by the OITS. Appropriate backups and disaster recovery measures shall be administered and deployed for all FHWA data. The enterprise data must adhere to the privacy rules and requests made by each respective business steward both internal and external to FHWA.
FHWA data must be accessible	FHWA data, information, and meta-data shall be readily accessible to all, except where determined to be restricted. When restrictions are made, business stewards of the data are accountable for defining specific individuals and levels of access privileges that are to be enabled. The OITS will be responsible for the implementation of proper security controls.
Meta-data will be recorded and utilized	All FHWA information system development and integration projects will utilize the defined meta-data program for data naming, data modeling, and logical and physical database design purposes. The DGAC is responsible for developing plans to capture and record specific data administration-focused meta-data consistent with the defined meta-data program.
Data stewards will be accountable by job description	Individuals designated as stewards will have specific enterprise data accountabilities incorporated into their job descriptions.
Timeliness of data	Data must be obtained, processed and be made available in a timeframe consistent with its intended use.

MnDOT, in their business plan, describes policies to improve data management through stewardship, curation, data security, database recovery, and data retention, as well as policies needed to implement data governance. The business plan identifies the executive group as responsible for developing data governance policies; the scope of their responsibility is to advance policies which includes:

- Charging a Data Stewardship Steering Committee to assess current policies relating to data to determine their efficacy
- Revising any policies that are obsolete, confusing or inaccurate
- Developing new policies that need to be implemented
- Developing an implementation plan to include a process for accountability, maintenance, communications and training

Policy needs, cited in Oregon's Public Transportation Plan (ODOT, 2018), acknowledges that ongoing changes in technology need to be met "with common system and data exchange standards", methods, and/or guidance for data collection, governance, sharing, and use. They identify several data exchange standards such as GTFS and a regional trip planner, but do not detail specific policies that emerge from these needs. ODOT is only beginning to initiate a data governance framework.

4.4. Organization

A typical Data Governance organization will be composed of

- A *Governing Board* of key stakeholders who review, promulgate, and oversee compliance with the "rules of engagement"
- *Data Stewards* who oversee the functional integrity and quality of specific data sets based on their subject matter expertise
- *Data Custodians* who perform the operational tasks of collecting, ingesting, validating, storing, and implementing tools to disseminate data sets

The MnDOT Data Governance roles (see Figure 3) include a *data set domain steward* (sometimes called data custodian) who is responsible for the operational efforts of collecting, storing, syntax, and validating the data, while the *data domain steward* (sometimes called the business data steward or data steward) is responsible for the quality, meaning, and appropriate use of data. The *data governance board* is typically in charge of advancing policies and procedures, change management, and championing the business and operational stewards. An organization may have multiple stewards responsible for different data sets.

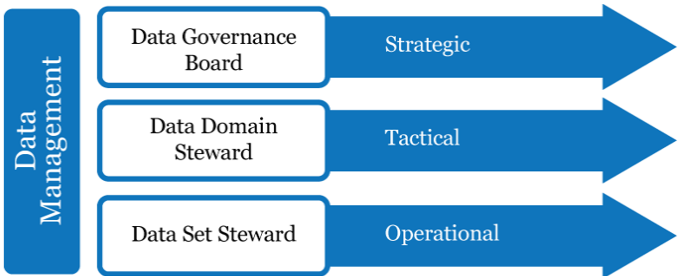

Figure 3: MnDOT Data Governance Roles

In a diverse enterprise such as ARC, additional layers may support regional coordination and data diversity. For example, FDOT adds a position which they refer to as the *enterprise data steward;* this role sits between the data governance board and the business data steward to coordinate and assess the impact of each individual agency's data usage and standardization, as illustrated in Figure 4. This model is adapted from the Florida Data Governance Policy[10].

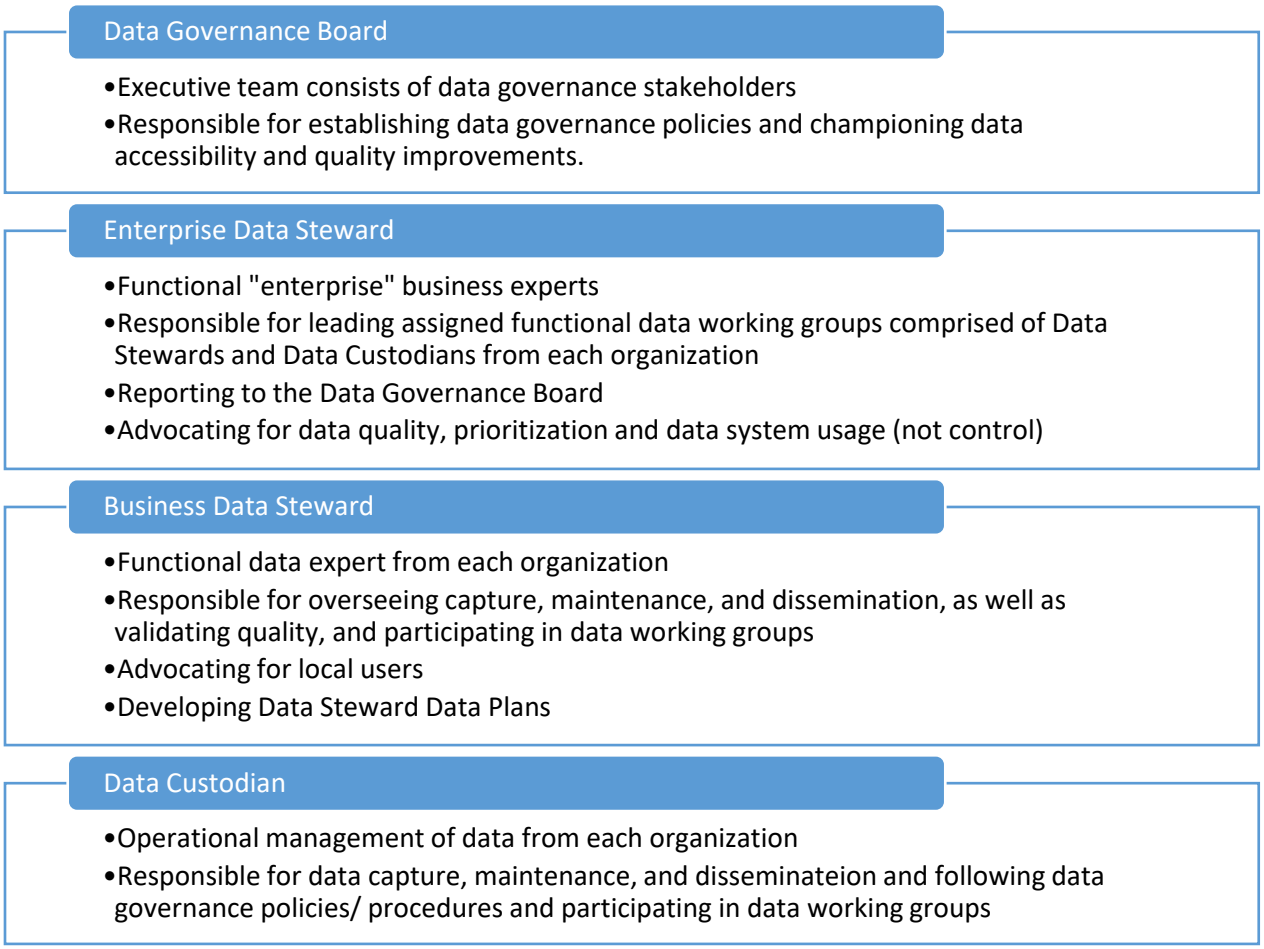

Figure 4: Extended Data Governance Roles for a Distributed Enterprise

The Enterprise Data Steward's role is to coordinate the activities of data stewards and custodians for each domain or data set. The coordination efforts include change management for data definitions and needs, interface specifications, data models and transformation, as well as developing requirements and specifications for verifying and validating existing or new data sets and interfaces. For example, there may be an enterprise data steward for transit data and another for work zone data.

The MnDOT Data Business Plan includes a comprehensive list of roles and responsibilities that can be used as reference for additional positions[11].

4.5. Performance and Maturity Models

A maturity model is the measurement of an organization's ability to continuously improve in specific areas. A Data Governance maturity model will measure the effectiveness of the DG framework such as repeatability and sustainability of organizational structures, processes, and rules of engagement, and how to improve performance. The maturity model may also measure data management performance to understand how the data governance framework impacts operational, tactical, and strategic performance.

George Firican, in a series on Data Governance[12], reviewed several enterprise information management maturity models including IBM, Stanford, Oracle, Gartner, Open Universiteit Nederland. The models are similar in that Level 1 has no processes and only ad-hoc activities while the highest level is optimized and fully

integrated with other business and technical areas; embedded in the culture and continuously assessed; and adapted to improvements and changing environments.

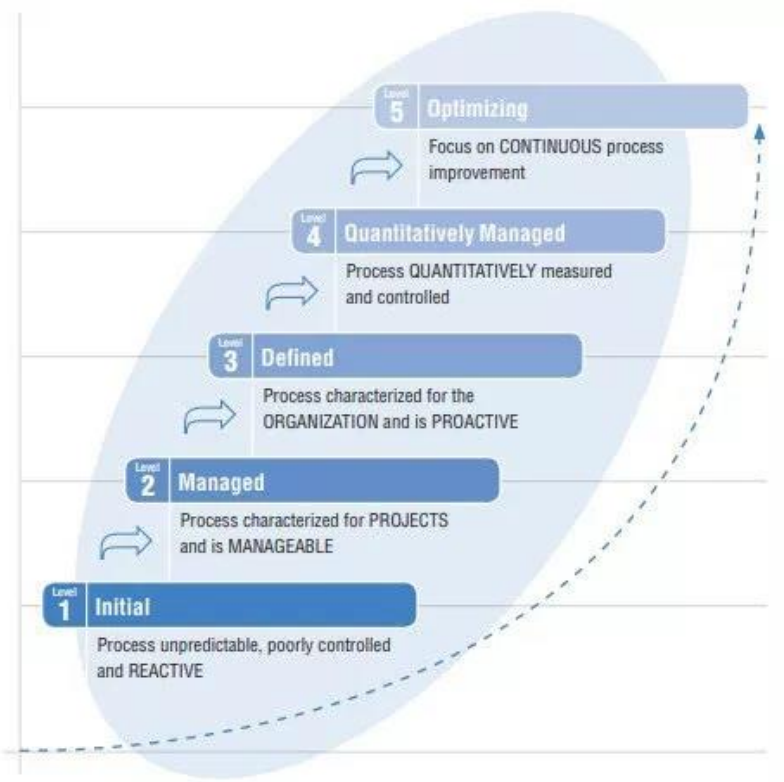

Figure 5: IBM Maturity Model

Most maturity models will generate a scorecard that measures an enterprise on several factors. The IBM model identifies four levels and 11 elements of effective Data Governance as shown in Figure 6.

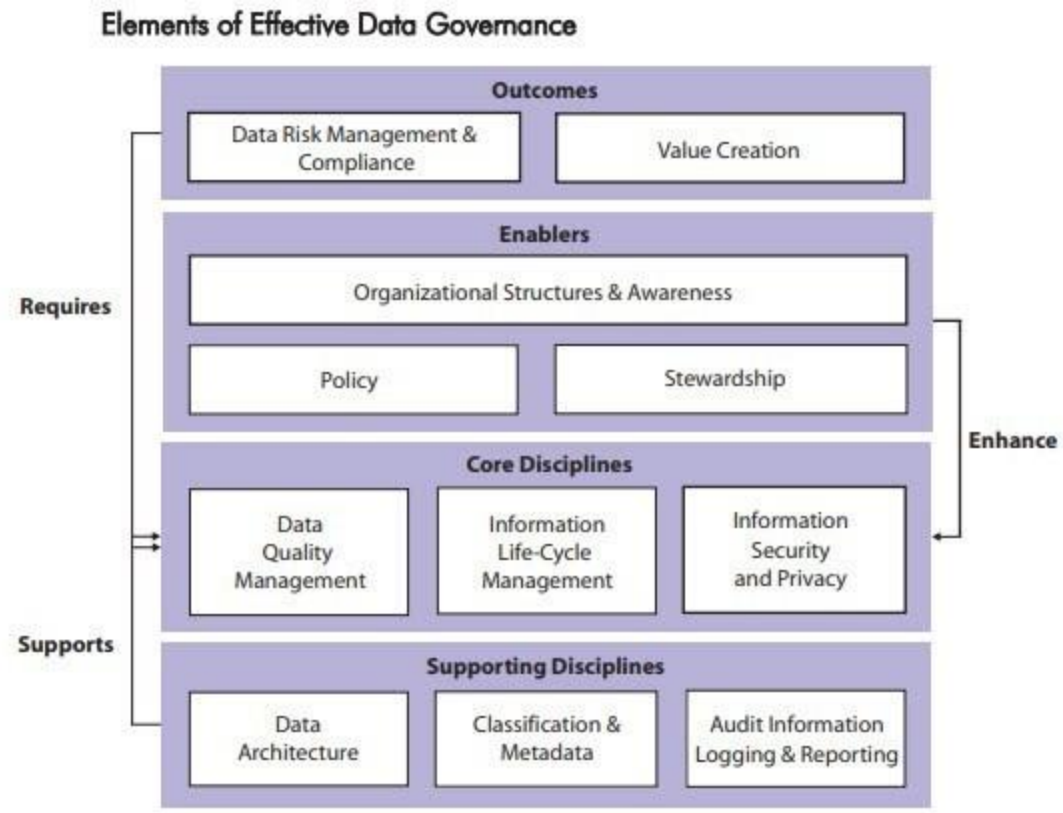

Figure 6: IBM Model to Assess Effective Data Governance

Many models create a rating system to score maturity based on a set of questions or based on measuring against goals, key performance indicators (KPI), or against a set of questions in a scorecard. An example of setting a *Value Creation* element measure will include the high-level element, objective(s), and related KPI(s) as follows:

> **Element:** measure the progress of providing discovery and access to data sets
>
> **Objective:** migrate local (workstation) data to online access
>
> **KPI:** record the number of data sets that were migrated and listed in the data catalog with respect to the total number of local data sets to identify the percent increase of accessible data

The various maturity methodologies identify different self-assessment tools, scorecards, and KPIs that may be used. Agencies can develop their own maturity models by capturing their goals and objectives from their framework and describing their objectives around the key DG components: people (roles), processes, and rules of engagement (quality, metadata, standards, etc.).

A continuously improving organization will collect and score their compliance with the objective through the KPI on a periodic basis – monthly, quarterly or annual reviews. The Data Governance Board will set the performance metrics to measure priorities and effectiveness of their policies.

5. Data Lifecycle Management

Data lifecycle management—sometimes referred to as data curation—covers the processes, rules, and responsibilities of data custodians, data stewards, and enterprise data stewards through data management processes, including access, use, and reuse. The Digital Curation Centre (DCC) curation lifecycle model shown in Figure 7, developed for all information—structured, unstructured, and semi-structured—shows that information is not static. Even so-called static, such as road networks and bus stops, encounter changes over time. The figure shows that initially, data is conceptualized through some process, whether through a Concept of Operations or project implementation. Once the need is identified, the data is created, collected or acquired. The appraisal process tends to involve cleaning or selecting appropriate data which is then ingested into a data repository where it is preserved, stored and used appropriately. This process is repeated to update the data as needed. The curation process consists of data use through selection, while the preservation process involves ingesting data usually through processing standard interfaces. Curation includes developing plans to represent, manage and preserve the data (digital objective), as depicted as the center of the lifecycle model. Collection, quality control, preservation, alignment and update processes drive the responsibilities of people who manage and use the data.

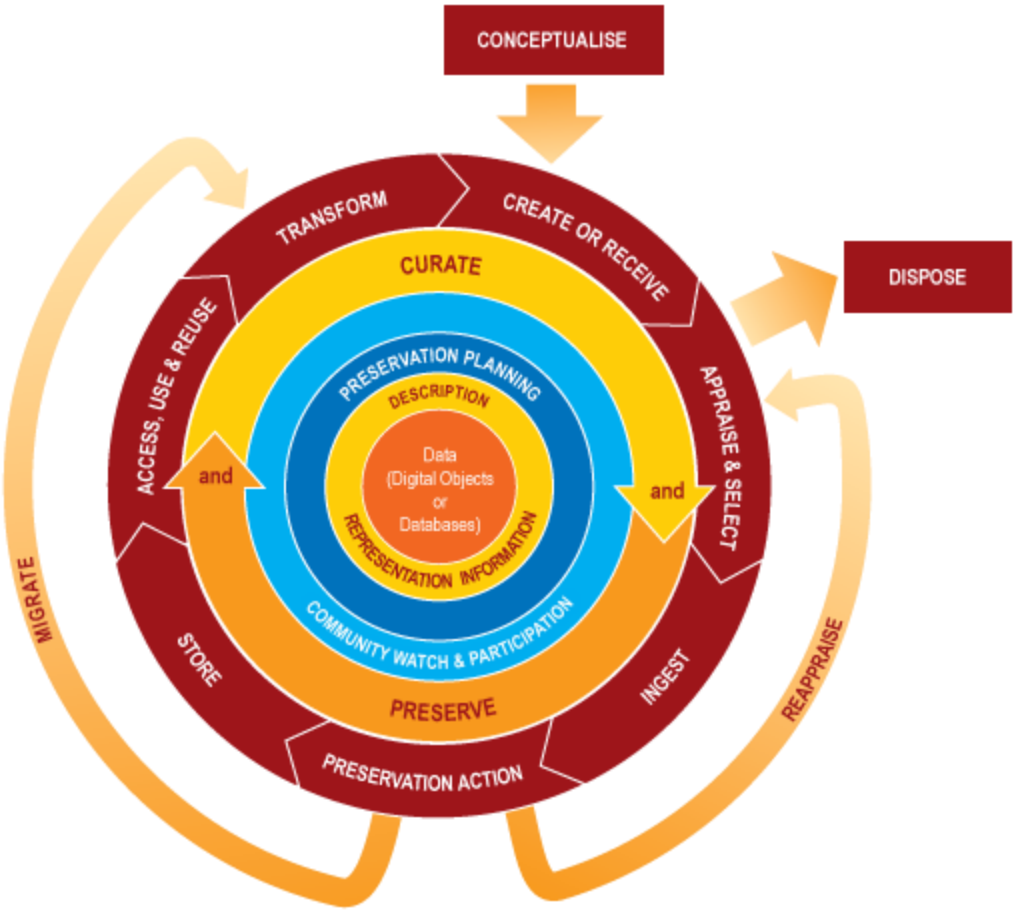

Figure 7: DCC Curation Lifecycle Model[13]

The cost of the software and hardware to store and process data is only a small fraction of the cost to collect, update, preserve and access the data. Manual collecting, transforming, updating, validating, managing and disseminating costs account for much of the investment in data; tools, particularly automated tools, to manage steps in the curation lifecycle will increase efficiency, reduce costs and enhance data quality. In particular, automated tools to collect, process and disseminate data in real-time are paid for as the value of the data becomes more critical to sustain and adapt operations of critical business strategies.

The “artifacts” that compose the lifecycle include the following types of document:

- Data catalogs
- Data dictionaries, models, and interface and feed specifications (e.g., API)
- Guidelines for applying and provisioning API and data feed specifications including procurement language and requirements to support regional interface standards
- Standard Operating Procedures for managing, ingesting, preserving, storing, and disseminating data
- Metadata requirements, including naming conventions, quality, lineage, configuration, and version control procedures
- Data improvement plans (at the data steward and enterprise steward levels)

The MnDOT Data Business Plan advocates that each data steward develop a 5-year plan to propose and prioritize their activities for the near future. A plan supports other near-term projects because the plan can leverage data or support the implementation of an activity. A typical initial priority for a data plan is to migrate spreadsheet data to a web-based tool so the data is listed in an enterprise catalog and accessible to multiple users.

The *reappraise* and *migration* processes included in the curation lifecycle are essential to ensure the data addresses changing usage; in a regional environment when downstream data users rely on regional data specifications, impacts to data meaning, formats and structure may be significant. These changes may be due to a new application or system coming on-line that produces new information or consumes data that is not currently available. The change process to agree on updates is typically assigned to an Enterprise Data Steward working group. Working group members identify impacts to upstream and downstream data systems and agree to a timeline for changes that incorporate regional concerns.

6. Changing Needs in Transformative Transportation Environments

6.1. General Implications

Transportation agencies are no stranger to technology, as they have depended on various technologies for decades to optimize operations and maximize the use of available infrastructures and resources. However, current and emerging technologies are different from previous ones in that they generate and use a significantly greater amount of data at more discrete levels, increased update rates, and at significantly faster speeds. Furthermore, advances in data storage and processing capabilities keep improving our ability to better analyze and represent real-world conditions in real time. The International Transportation Forum (ITF) discusses the implications of data in transportation (OECD/ITF, 2015).

The bullets below expand on ITF's findings to provide insight into how data can change the transportation environment within the US.

- Sensors and data storage/transmission capacity in vehicles provide new opportunities for enhanced safety.
    - There are many ongoing efforts to develop/improve the technologies (and related standards that will govern them) for data collection and vehicle connectivity (e.g., communications protocols). The projected use of these technologies is mainly targeting safety improvements through connected and automate vehicles[14]
- Multi-platform sensing technologies are now able to precisely locate and track people, vehicles, and objects.
    - Location-sensing technologies are becoming cheaper and more widely deployed. When coupled with vehicle communication advancements and widespread penetration of mobile devices (e.g., smartphone), precise and persistent tracking of people and assets/goods becomes possible in ways not previously achievable. Implementations of this capability is especially important in freight—current efforts in this field include *e-Permitting/Virtual Weigh Stations*[15] and *Universal Truck Identifier.*[16]
- Properly combined data can reveal patterns and new knowledge about transport activity and flows.
    - The fusion of purposely-sensed, crowd-sourced data generates new knowledge that was not achievable previously. Both the public and private sector are using big data to understand trends and patterns in demand, allowing the sectors to supply a better service. For instance, the Chicago Transit Authority (CTA) uses bid data to assess changing traffic and ridership patterns and re-allocate bus service where it is most needed (CTA, 2019).
- New sources of data (and analysis capabilities) can also create unique privacy risks, as location and trajectory data are inherently personal in nature and difficult to anonymize effectively.
    - The identification of patterns can have unforeseen risk, as this may open new avenues for misuse and potential manipulation of individuals and their behavior. New tracking capabilities can expose daily patterns of activity and relationships that serve as powerful quasi-identifiers. While there are many techniques to remove personally identifiable information (PII) from data,[17] doing so effectively while retaining sufficient detail for useful analysis remains a challenge.
- Data protection policies are lagging behind new modes of data collection and uses, which is especially true for location data.

- Personal information and data collection strategies were not anticipated by regulations and policy makers, and authorities have not accounted for the new knowledge that emerges from data fusion. As these continue to evolve, outdated rules will continue to govern the collection and use of personal data.

6.2. Specific Implications for Transportation Data

The following sections provide insight into how specific technologies are affecting (or expected to affect) the transportation data environment.

6.2.1. Integrated Transportation Management Systems Impacts

Interoperability of multiple modes from multiple sources is already an issue faced by transportation agencies. For instance, data is reported from field to transportation management centers (TMC) and police, identified via closed-circuit television (CCTV), and located by crowdsourced applications. In an integrated, interoperable data environment there is a need to correlate multiple incident detection/reporting channels to associate multiple reports to a single event (incident, road weather, special events, construction or maintenance work zones and lane closures) while simultaneously ensuring the integrity of the incident (versus a secondary incident) throughout the event's duration (i.e., identification, status update and closure). As such, agencies are linking incidents to public distribution channels to provide more consistent information regardless of how it is accessed—e.g., via a traveler information website, variable message sign, en-route connected vehicle app, or mobile app.

Finally, when data is integrated and shared across modes, coordinating agencies have the same situational awareness. They can develop "playbooks" and action plans wherein they coordinate strategies without significant effort. In this sense, quality programs to foment the promulgation of standards (or guidance for implementing the standards) can improve data integration. It should be noted that though some of these strategies have been partially implemented in the Atlanta region, there are information and blind spots that are shared, or the data is not of sufficient quality to be useful.

6.2.2. Mobility on Demand and Accessible Travel Impacts

Many mobile apps under development target travelers that seek multimodal and accessibility—e.g., people with disabilities or active transportation travelers (e.g., bikes, pedestrians). Emerging modes in addition to the traditional modes (car, bus or rail) have several common data needs, particularly:

- Data collection about non-vehicle pathways and facilities (infrastructure and conditions), such as:
  - Ramps, tactical paving pads
  - Sidewalk and bike path surface conditions
  - Stairs, elevator, escalator dimensions
- Data collection and distribution of vehicle facilities and availability, such as:
  - Parking for carsharing services
  - Electric charging stations
  - Parking locations for AV not currently in service
  - Ride hailing, taxi pickup / drop off locations
  - Methods to integrate payment to ensure seamless use across modes

Furthermore, agencies that regulate emerging mobility services, such as bikesharing and e-scooter services, also need to distribute and collect data that is not currently available:

- Public access zones where restrictions exist (e.g., no free bikeshare or e-scooter storage)
- Applications to audit micromobility device compliance with regulations

These emerging trends elicit questions about the role of government with respect to collecting and managing the information:

- What role does government have in collecting and updating infrastructure data, communicating condition and status information, and offering applications for travelers to locate these information sets?
- In the future, if emerging multimodal and mobility on demand services are regulated, then what role, facility, and applications will government possess to audit and ensure compliance with the regulations?

6.2.3. Automated Vehicle Impacts

Automated Vehicle (AV) technologies require dynamic acquisition and rendering of the transportation network and its conditions. Much of the data will be collected in real time and distributed by private cloud services, not the government. In the supply chain model that is anticipated by the USDOT Data Infrastructure Initiative (see Figure 8), the critical role of government, shown by the red boxes, consists of generating and distributing accurate information about road weather conditions, pending and active work zones (including lane geometry, closures, and restrictions), and other data feeds that are under development by USDOT.

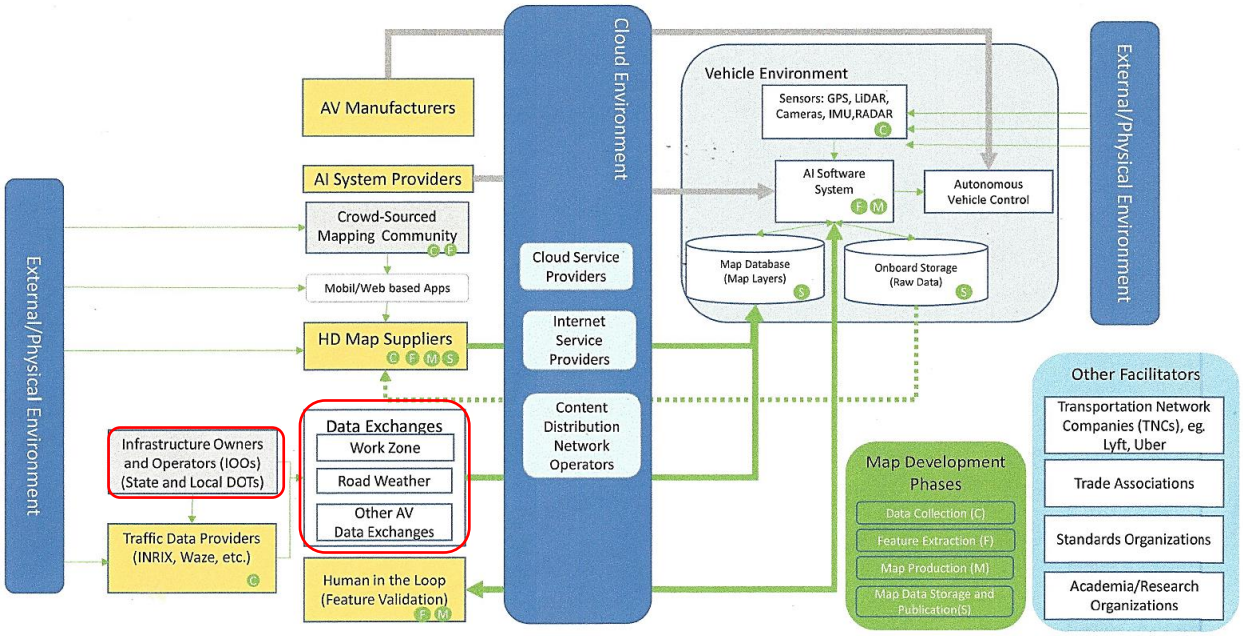

Figure 8: Digital Maps for AV Supply Chain (source: USDOT)

An oft-cited issue by map vendors and AV developers is minimal reference information issued by public agencies. These metadata products include (1) description of the roadway linear referencing system (milepost), (2) master references for road names and directionality, ramp identifiers, lane numbering, and (3) type of road work including accuracy and certainty of time and spatial values. Some of the data will be provisioned by smart work zone technologies or other IoT sensors. To that end, data sets will require frequent update and distribution. Regardless, *effective management and distribution of metadata (i.e., quality data about data) will become critical moving into the future.*

7. Getting Started with Data Governance

Several transportation agencies including FHWA and MnDOT have documented the approach they took to initiate a DG Framework. The two approaches—FHWA's approach, described in Section 7.1, and MnDOT's recommendations, described in Section 7.2 may be viewed as sequential. The initial planning, described by FHWA involves developing a Data Business Plan that identifies the vision, goals and objectives, organizational structure (roles and responsibilities for key stakeholders), reporting process, charter, data set inventory and gap analysis (maturity model). The MnDOT approach assumes that those steps have been taken, and the recommendations flow from the gap analysis.

7.1. FHWA Approach

U.S. DOT developed a Transportation Data Plan that included steps for establishing a Data Governance Framework (Vandervalk, Snyder, & Hajek, 2013). USDOT piloted this approach for local and state governments in a pilot with the Hillsborough MPO[18]. Their approach follows six steps for implementing Data Governance. The steps are referenced below:

1. **Map data programs to business objectives –** Define the relationship between the mission and business objectives of a U.S. DOT stakeholder office and how they map to the data programs managed by that office.
2. **Define stakeholder roles and responsibilities –** Using established hierarchical relationship between data management, data governance, and data stewardship, U.S. DOT stakeholder offices should define roles and responsibilities for data governance. For example:
    a. Data Governance Team – The designated individuals within FHWA Office of Operations responsible for the oversight of data programs to support the business functions of the office.
    b. FHWA Data Governance Advisory Council, Office of Operations Team Leader – Representative from the FHWA Office of Operations who will participate on the FHWA Data Governance Advisory Council.
    c. Data Business Owners – Individuals who manage the data and metadata for information systems within their area of responsibility. Data business owners are responsible for maintaining the data dictionaries for the data systems and for establishing business requirements for the use of roadway travel mobility data.
    d. Data Stewards – Individuals who ensure data is managed according to policies established by FHWA Office of Operations Data Governance Team.
    e. Community of Interest, Internal – Any persons or offices internal to U.S. DOT that collects, owns, maintains, uses or interfaces with, accesses, benefits from, or is otherwise affected by roadway travel mobility data.
    f. Community of Interest, External – Any persons or offices external to U.S. DOT that collects, owns, maintains, uses or interfaces with, accesses, benefits from, or is otherwise affected by roadway travel mobility data.
3. **Develop data governance model –** Once the mapping of data programs to agency and office mission and goals is accomplished, a data governance model diagram should be established to formalize the structure for managing the data programs. The figure below [from reference identified as Figure C-2] is a generic and high level data governance model (National Academies of Sciences, Engineering, and Medicine, 2010), a more detailed example can be found in Appendix D of U.S. DOT's Data Business Plan[19].

Figure C-2. Generic Data Governance Model (Source: NCHRP Report 666: Target Setting Methods and Data Management to Support Performance-Based Resource Allocation by Transportation Agencies, Transportation Research Board, 2010)

4. **Develop data governance charter –** The data governance charter will then set the purpose, mission, vision, goals and objectives, and data management policies for data governance within a U.S. DOT stakeholder office.
5. **Develop data catalog –** This catalog provides a centralized location for information about the data used by stakeholders involved with roadway travel mobility data programs. Involvement means the office is performing one or several of the following functions related to data: collection, analysis, reporting, dissemination, or providing guidance to other stakeholders related to those functions. The catalog should be reviewed and revised by data business owners within each U.S. DOT stakeholder office to ensure that all data systems, data standards, roles, and responsibilities, etc., are correctly identified. It should also be revised at least on an annual basis, or monthly if changes occur that require updating the information listed in the catalog.
6. **Assess data governance maturity –** Implement a data management maturity model to assess where the organization stands with respect to implementing certain data governance processes. The maturity model also can be used to benchmark for comparison or assist an agency in understanding common concepts related to an issue or process. A typical maturity model identifies levels and characteristics of those levels. The model can be used to assess an agency's status and assist in identifying next steps to achieve success toward an ultimate goal state.

7.2. MnDOT Approach

In its Data Business Plan[20], MnDOT developed a Data Governance Framework (as shown in Figure 9), assigned strategies to each of the framework components. The nine recommendations and strategies are listed in Table 4.

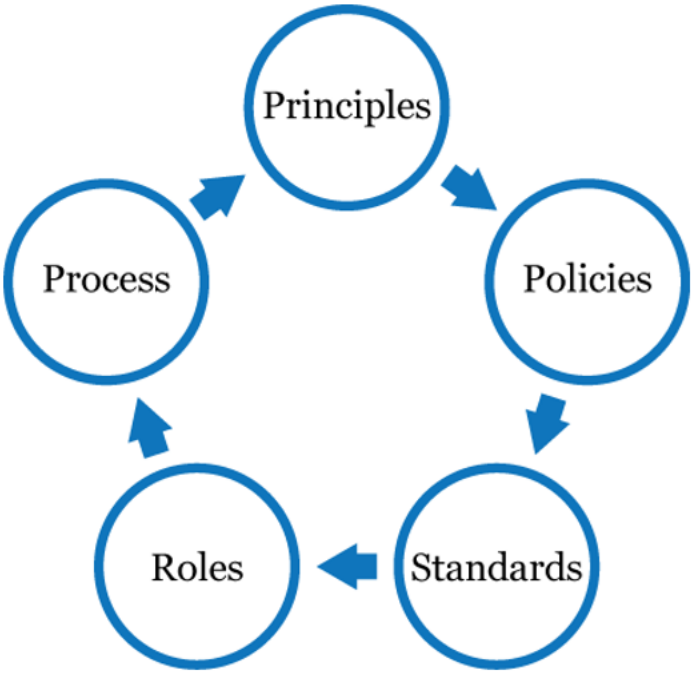

Figure 9: MnDOT Data Governance Framework

Table 4: MnDOT Recommendations and Suggested Strategies for Data Governance

Recommendation	Suggested Strategies
Recommendation 1: The Data Governance Board shall formally adopt the principles on behalf of MnDOT and incorporate them into policies, standards and processes	A. Adopt the data management principles at the initial Data Governance Board meeting B. Incorporate principles into policies, standards and processes C. Develop a communication plan to include the principles for targeted audiences such as data coordinators, data stewards and other data stakeholders D. Develop a training plan to include the principles for targeted audiences such as data coordinators and data stewards
Recommendation 2: Revise existing policies (e.g. stewardship, development, data security, database recovery, data retention) and develop additional policies needed to implement data governance at MnDOT	A. Charge a Data Stewardship Steering Committee to assess current policies relating to data to determine their efficacy B. Revise any policies that are obsolete, confusing or inaccurate. C. Develop new policies that need to be implemented D. Develop an implementation plan to include a process for accountability, maintenance, communications and training
Recommendation 3: Adopt or revise existing standards (e.g. metadata elements, naming conventions, physical data modeling) and develop additional	A. Charge a Data Stewardship Steering Committee to assess current standards relating to data to determine their efficacy B. Revise any standards that are out dated or unused. C. Develop new standards that need to be implemented

standards needed to mature data governance at MnDOT	D. Develop an implementation plan to include a process for accountability, maintenance, communications and training
Recommendation 4: Form a Data Governance Board to replace the BIC with members representing the divisions, the CIO and the Data Management Coordinator (see Figure 3)	A. Develop a staffing plan to identify positions and/or persons who take on the Data Governance Board role B. Review and adopt the Data Governance Board responsibilities as the board charter C. Develop a work plan for implementing policies, standards and processes for data governance
Recommendation 5: Create the Data Stewardship Steering Committee role as part of the larger data governance program	A. Determine the purpose or charge for each data stewardship steering committee B. Identify data domain coordinators to serve on each committee
Recommendation 6: Formalize the Data Steward role as part of the data governance program	A. Integrate the notion of data stewardship into policies, standards and processes B. Define data domains, data sets and stewards needed to represent all the data used by MnDOT's products and services C. Formally identify data stewards for core or department-wide data domains and sets
Recommendation 7: Assign the Data Management Coordinator role within MnDOT.	A. Develop a staffing plan to fill the Data Management Coordinator role.
Recommendation 8: Develop a process to integrate or create touch points between data governance and Division Directors' investment management	A. Incorporate data projects into the Division Director's IT Development Investment Plan
Recommendation 9: Initiate a project to implement a Business Data Catalog	A. Initiate an IT project to implement a Business Data Catalog using the recommendations made by the independent consultant B. Develop the catalog concurrently with the business intelligence project in order to eliminate duplication of effort during the development of the catalog iterations C. The project will implement multiple deliverables and activities, including a method to validate the data and implement a data management plan, maintenance plan and security procedures. In addition, the project will identify a tool to implement the Business Data Catalog. The data will need to be organized and cataloged based on the data domains/sets with responsibilities assigned to corresponding Data Stewards

7.3. FHWA Challenges and Lessons Learned

According to FHWA Primer, the concept for data governance "establishes the criteria and requirements for data; their quality, management, policies, business process; and risk management for handling of data. In short, it is a corporate approach to collecting and managing data."[21] Similar to the challenges faced by ARC constituent organizations, the criteria and requirements for data often highlight data issues that are common amongst transportation agencies, such as:

- Finding consistent data for business needs and partner/customer inquiries.
- Identifying real and perceived data quality issues.

- Integrating data across departments and with external organizations.
- Identifying data that are valuable and needed to drive decisions.
- Standardizing approach to address existing data and new data needs.
- Preventing redundant data collection.
- Accessing needed data.
- Keeping current with changes in computer technology and electronic data storage standards.
- Maintaining security and accessibility so that data elements cannot be lost, corrupted, or otherwise made unavailable to users.

FHWA surveyed four State DOTs (Arizona, Arkansas, Ohio and Texas) regarding their efforts in data governance and data management within the context of their Geospatial Information Systems programs.[22] The States described noticeable benefits and impacts they experienced through their efforts to promote/implement data governance strategies/programs, as well as challenges and key lessons learned. These insights are summarized in Table 5—note that GIS-specific are not included here.

Table 5. Experiences, Benefits, Challenges and Lessons Learned from Implementing Data Governance.

Experiences and Benefits	
Communicating value to executive management.	Executive management respond positively to strategies that can help the agency reduce cost or make processes more efficient with the same budget. As such, engaging the executive management's desire to save the agency time and money is a strategic method of gaining traction and support for data governance. Some agencies reported that they were able to make headway in their efforts by making it clear that the policies will save the agency resources.
Facilitating collaboration between DOT staff.	Data governance and data management policies facilitate the ability of agencies to work with, and learn from, other State agencies. Cross-compatibilities between State agencies is a huge benefit.
Organizational structures are highly effective implementation tools	Structures such as steering committees or designated oversight roles have been extremely helpful for agencies in their data governance and data management projects. These organizational structures fill a need for a feedback loop that can assess the progress of the agency and see the direction it is heading. In some instances, data governance committees have members that are from top-level management.
Positive impacts are quickly recognized and appreciated by staff.	Even at early stages, staff begin to realize change is necessary to remain relevant, do meaningful work, and spread institutional knowledge across the agency. Staff also begin to realize that data is a powerful tool and has great intrinsic value, and so are more willing to put the effort into properly maintaining it.
Challenges	
Agency culture can be difficult to overcome.	It can be difficult for some staff to view data as an asset with a monetary value. In this sense, some staff may hesitate to volunteer to do tasks that are not directed by administrators as they might not understand why change is needed. Additionally, the lack of ownership of the data process among staff and business owners means that many times parties assume that other staff will be taking care of data maintenance for them.
Bureaucracy between executives and the	The value of data governance and data management to upper management can be diluted by levels of bureaucracy. When executives are not in touch with the

agency can result in miscommunication.	day-to-day work going on within their agency, it is difficult for them to set policy that will help their staff achieve their goals.
Personnel turnover is a serious obstacle to implementation.	When key personnel leave an agency, it has a great detrimental effect on the continuity of effort, results in the loss of institutional knowledge, and hinders follow-through on existing projects. New staff or a new administration will then need to be convinced of the value, which may not be successful. All participants made note of this issue.
Internal departments and teams can have different missions.	The sub-groups within a DOT oftentimes have their own mission and goals, and the best way of achieving those goals might not align with the rest of the agency. This makes compromising and moving in a direction that benefits the agency as a whole complicated and can result in territorial disputes for resources.
Administrators' focus on engineering can prevent them from understanding the value of data governance and data management.	Administrators can be intensely focused on operations, maintenance and construction. Because of this, they can have difficulty in understanding how data governance and management fits into the agency's business operations.
Lessons Learned	
Definitions of these concepts may differ in language, but they are functionally the same.	The creation of official definitions can provide DOT staff with a point of reference when communicating with upper management, and shows it is a concept valued by FHWA.
Without a governing body, implementing data governance is very difficult.	Steering Committees and similar governing bodies are an essential part of an effective implementation process as they provide a voice of authority and vision. Implementation becomes hard to perform without this governing body.
Data governance and data management have a symbiotic relationship.	Performing data governance without the day-to-day practices of data management results in little to no progress being made towards the goals and policies set forth in a data governance policy. Likewise, data management practices that are performed without the guiding framework provided by a data governance policy will result in haphazard or unorganized activity that is subject to individual workflow preferences, and data collection and maintenance standards.

8. ARC's Role in a Regional Data Governance Framework

8.1. Recommendations for Establishing a Regional Data Governance Framework

ARC is in a pivotal position to promulgate and orchestrate a data governance framework for the region. As described in Section 7, the USDOT Data Governance Primer recommends a process for developing a Data Business Plan to initiate data governance. As a test case, FHWA consultants worked with Hillsborough MPO to pilot the approach for state and local DOTs[23]. An 8-step process was recommended in the Hillsborough MPO Pilot document [24]. However, as part of the ARC TSMO visioning process, the eight steps may be condensed to four as illustrated in Figure 10.

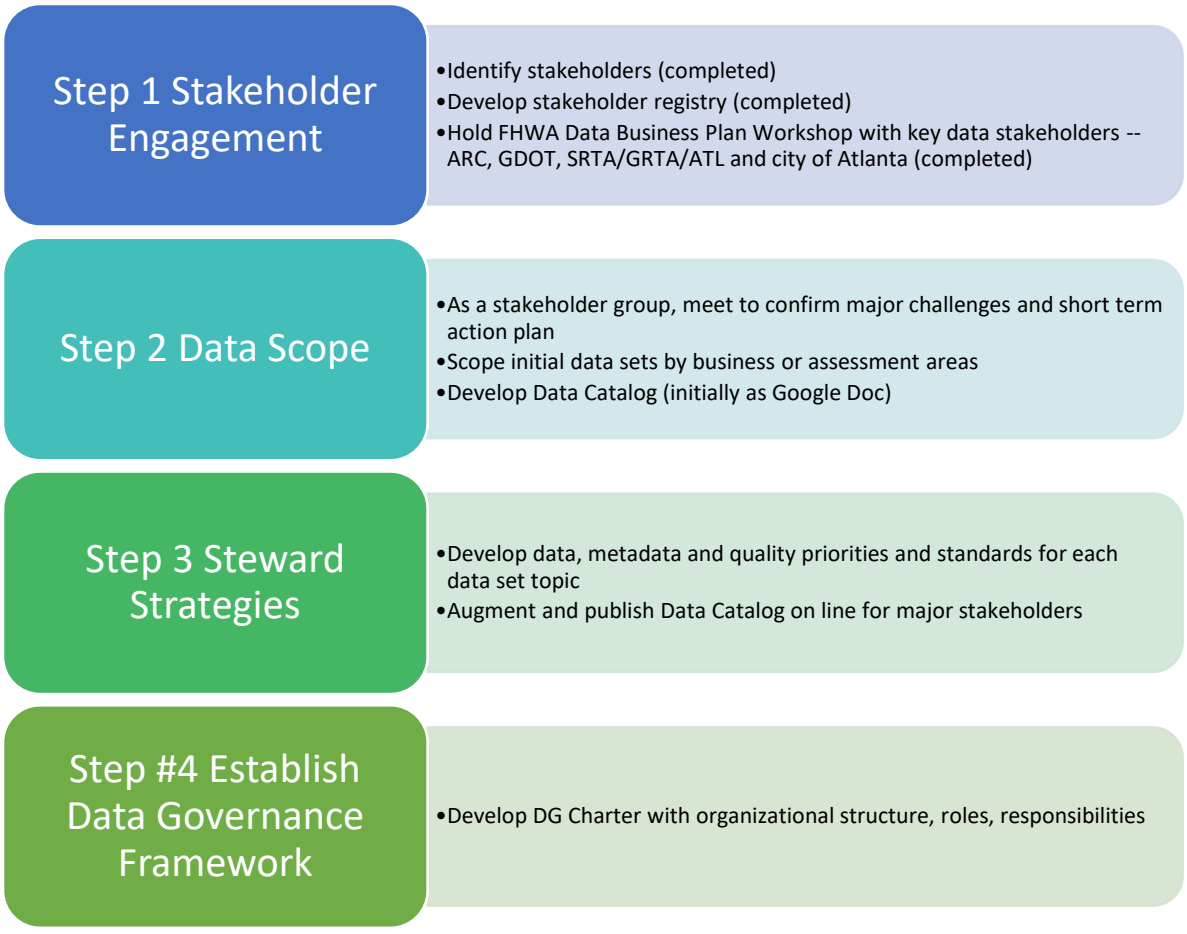

Figure 10: Data Business Plan Development Process

Governance is an iterative process which can start by applying small steps and developing into a mature framework over time. This approach, agreed to at the FHWA Data Business Plan Workshop (held May 14), uses an existing working group and an initial data inventory to document current, available data sets.

8.1.1. Data Set Catalog

A data catalog may be as simple as a web page with a set of links to the data set. It typically has information on the data set description, owner, date of publication or activation, access methods, file and data format, and data descriptions. Data catalogs tend to post data in a standardized format, like GTFS, where the data and file formats and data definitions are specified in an open, published standard. A file format may be comma separated values (CSV), JSON, XML, or Feature Services (OGIS format). A data dictionary including data formats may be a published interface document like NTCIP 1211 for Signal Control and Prioritization or GTFS. The Federal government has an on-line data catalog that is a simple list of data sets and their available formats (see Figure 11). A link is typically available for the user to download the data.

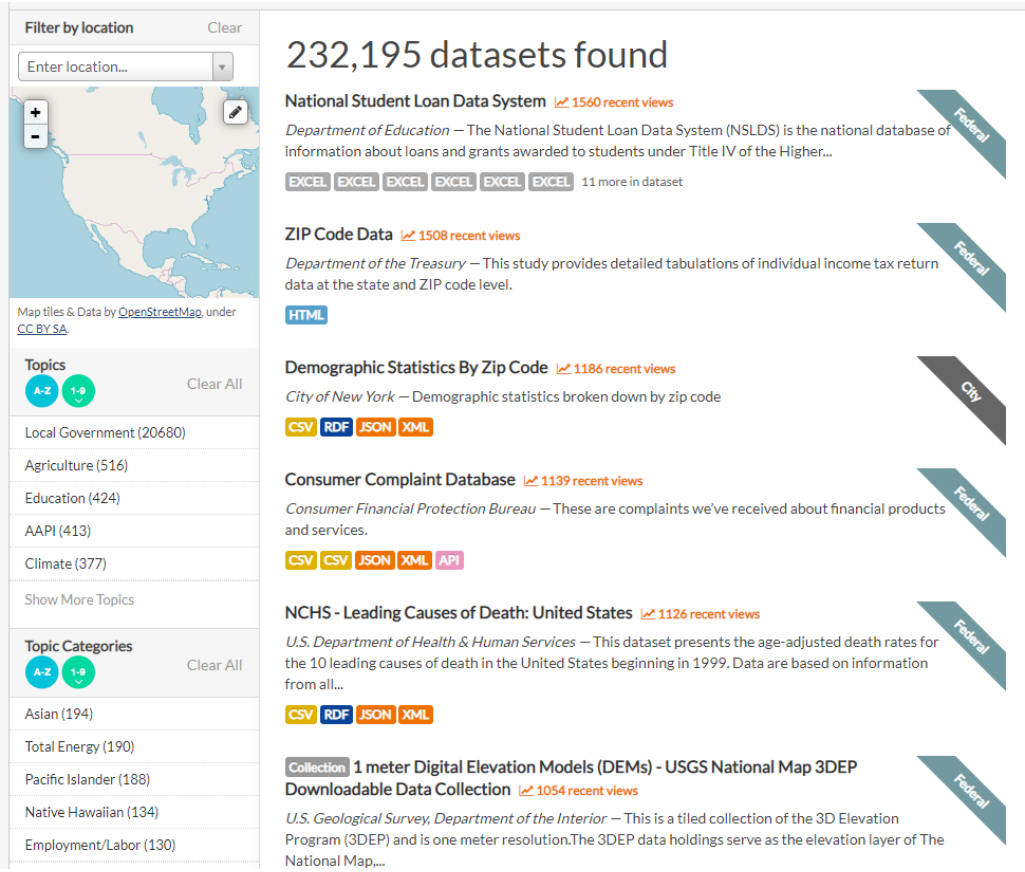

Figure 11: Federal Data Catalog (https://catalog.data.gov/dataset)

Another example, the Maryland GIS Data Catalog, shown in Figure 12, posts a short description of the data, available file formats, source data, publication date and link to the data file.

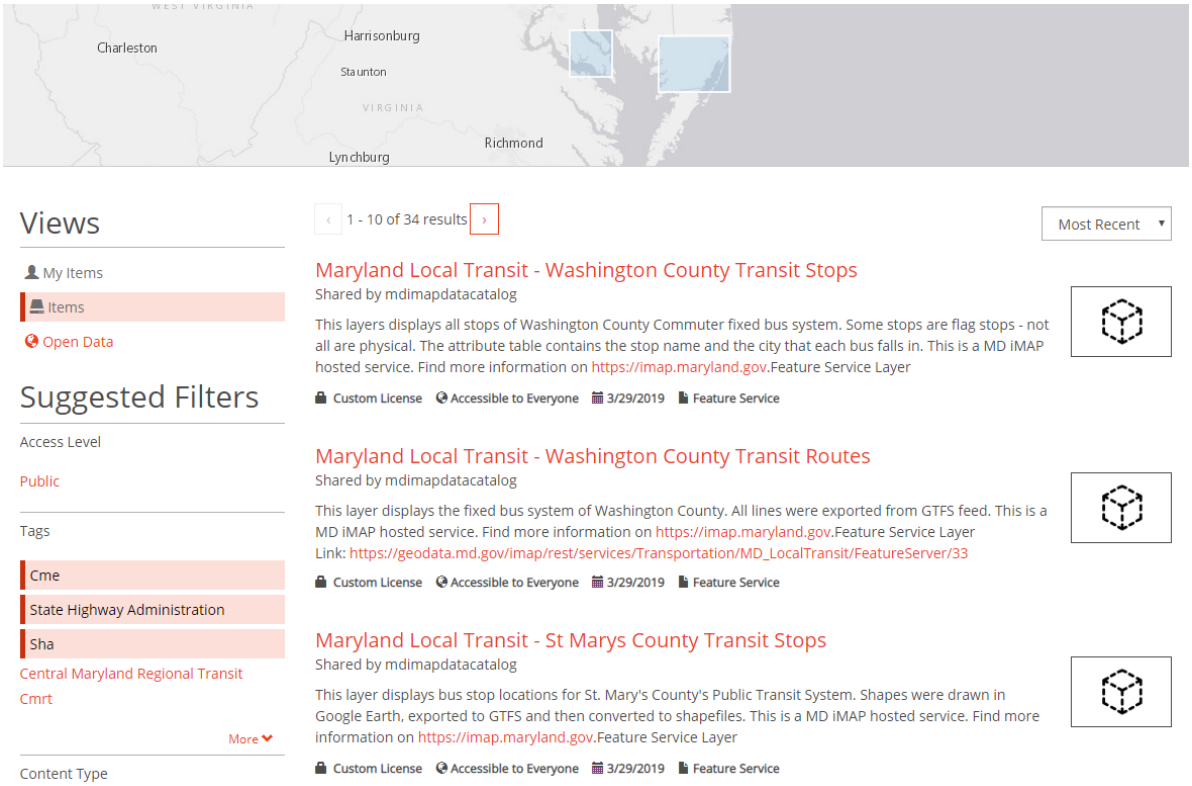

Figure 12: Snapshot of the Maryland Transit Data Catalog (https://data.imap.maryland.gov/items?page=3&tags=CME)

These on-line, public facing catalogs are simple. Both examples attach tags to the data sets which classify the type, owner, topics and sometimes geographic coverage. The details associated with the data catalog can be extended and grow over time. Table 6 shows a list of a minimum and detailed set of attributes that may be included in the catalog and additional attributes that can extend available data set information.

Table 6: Data Catalog Attribute List

Minimum Set of Attributes	Detailed Attributes
Data set name	Geographic coverage
Topic (from a selected list)	Data source and creation process
Data set description / abstract	Location referencing method(s), if geographic
Date of collection / publication	Expiration date
Owner	Update frequency of data set
File format(s)	Data steward contact information
Data dictionary (with a list of data definitions and formats) or standard that describes data.	Quality, accuracy and validation process
Link to data file by format(s)	Extended data dictionary (with examples and guidance on how data was applied to standard format)

8.1.2. Agency Roles and Responsibilities

The working group will need to make the following decisions with respect to managing the Data Catalog:

- How and where are data sets accessed? On a central site? from an agency resource?
- How to make the spreadsheet or inventory information available to authorized data users?
- Who is responsible for updating the data catalog and how often?

These roles and responsibilities will be assigned to each organization to update and augment the data catalog as additional information is accessible.

8.2. Recommendations for Establishing a Long Term Data Governance Framework

Once the initial steps for documenting data sets is established, the region can start to invest in specific areas to improve and automate data collection. This longer term effort will start from Step 2 in the Data Governance Framework Process, depicted in Figure 13.

Figure 13: Longer Term Data Governance Framework Process

The Data Governance charter establishes a formal structure to govern data quality, access and reporting. Using the recommended organizational model described in Figure 4, ARC can play a significant role to facilitate, lead or be a pivotal player in the Data Governance Board. Additionally, ARC can convene and facilitate the Enterprise Data Steward working groups as chair or secretary. Recommendations for ARC responsibilities related to the Board and EDS include:

Data Governance Board

- Lead and convene Data Governance Board
- Ensure that everyone's interest is supported and considered (creating efficiency for the regional eco-system)
- With the Board,
    - Develop charter including organization, policies, and MOUs for each organization
    - Develop quality level guidelines (including metadata requirements and access methods to share information)
    - Develop approach for data business plans, enterprise data stewards, data priorities for region by identifying priority use-cases for regional data management systems including but not limited to
        - Work zone coordination
        - Regional performance measurement
        - Bike and pedestrian information
        - Traffic volume, counts, turning movements
- Apply Board policies and procedures as a TSMO project development requirements
- Develop/support member agencies with specifications, procurement language around data.

Enterprise Data Steward

- Coordinate business data steward plans and working groups at the enterprise level
- Facilitate each EDS working group to
    - Develop regional standards and guidelines for implementing including processes, and metadata for each business area;
    - Support each business data steward in developing a 5-year plan
    - Meet periodically to report on progress of 5-year plan and add to plan based on new programs
- Generate catalog of data sets available for sharing

Business Data Stewards (BDS) and Data Custodians are stakeholder organization responsibilities. Data custodians (DC) are typically IT staff or project staff who are responsible for implementing 5-year Data Steward Plan including the standards set by the enterprise data stewards (e.g., FHWA data feed or ITS standard message specification). ARC will need to define internal BDS and DC roles to formally manage and participate within the Data Governance Board to encourage adherence to the Data Business Plan.

The typical documents that are initially required to kick off a regional Data Governance Framework includes the following:

- Data Business Plan which includes:
    - Vision, goals, objectives
    - Challenges and needs
    - Priorities
    - Action Plan
    - Change management process

- Data Governance Charter
- Catalog of Data Sets; list of available data by --
    - Business Data Area (e.g., network performance, planned projects, transit, analysis/performance, crash/safety) – see Summary from Workshop #2 for initial set of business domains and subdomains
    - Data Standards: data models, data dictionary, interface formats and methods
    - Metadata: description and files that describe the data (e.g., ownership, contact information, naming conventions, lineage, references for street names)
    - Access methods and file formats
- 5-Year Data Steward Plans (by stakeholder and business area) which include:
    - Prioritization of data sets to update and a brief description of how to manage the information (e.g., renew, transform, migrate to accessible format, etc.)
    - Data collection, transformation and translation processes
    - Update cycles
    - Metadata and quality requirements
    - Standard interfaces and methods
    - Access methods

In addition to leading the development of formal data governance processes and plans, ARC has the opportunity to encourage data governance best practices through their processes and guidance resources. ARC is currently developing Local Agency Deployment Guidelines which will support data governance best practices and resources. Once more established regional data governance plans and processes are in place, they should be integrated into other MPO processes such as Comprehensive Transportation Plans and Transportation Improvement Program (TIP) solicitations. For example, stakeholders could be asked to submit data sets during other regular cycle updates. Another example would be to ask for data management information during project development such that data governance practices are being considered at the initial stages of project inception.

9. Endnotes

[1] "What is Data Governance?" (2015-6) http://www.tcdii.com/whatIsDataGovernance.html

[2] Goetz, Michel. Data Governance and Data Management are Not Interchangeable. Forrester, Sept 11, 2015. https://go.forrester.com/blogs/15-09-11-data_governance_and_data_management_are_not_interchangeable/]

[3] DGI, Posted by Gwen Thomas – Getting Started/Defining Data Governance http://www.datagovernance.com/defining-data-governance/

[4] NASCIO, Data Governance – Managing Information As an Enterprise Asset Part 1 – An Introduction. (April 2008), p. 1. https://www.nascio.org/Portals/0/Publications/Documents/NASCIO-DataGovernance-Part1.pdf

[5] This summary excludes the MDM institute definition inclusion of "technology" and NASCIO's inclusion of critical data management methods and components. The definition focuses setting the rules for on the accountability, as well as describing planning goals and processes and policy provisions related to data governance.

[6] Atlanta Regional Transportation System Management and Operations (TSMO) Visioning Workshop Summary.

[7] Ibid. p, 12.

[8] MnDOT Data Business Plan, p, 43.

[9] FL DOT Data Governance Policy, Topic No.: 001-325-064, https://fdotwp1.dot.state.fl.us/ProceduresInformationManagementSystemInternet/FormsAndProcedures/ViewDocument?topicNum=001-325-064

[10] Ibid.

[11] Data Business Plan. Minnesota Department of Transportation, October 2007, Appendices 5 and 6.

[12] http://www.lightsondata.com/category/data-governance/

[13] DCC Data Curation Lifecycle Model, http://www.dcc.ac.uk/resources/curation-lifecycle-model

[14] see the Connected Vehicle Pilots website for information about the ongoing pilots within the US.

[15] FHWA's Office of Freight Management & Operations recently completed two initiatives that addressed electronic permitting and virtual weigh stations. An assessment of "e-Permitting" efforts underway in New York, Tennessee and Florida was completed in partnership with the I-95 Corridor Coalition. Three documents were completed documenting the "State of the Practice" in state level efforts to advance virtual weigh stations; the "Concept of Operations for Virtual Weigh Stations", and "Implementation Plan for Virtual Weigh Stations"

[16] The Office of Freight Management & Operations has initiated a project to identify the advanced technologies capable of uniquely identifying commercial vehicles subject to US Code Title 23 and 49 inspections and measurements

[17] Suggested actions to protect privacy include: Data Authentication, Data Integrity, Applications/Services Access Control, Data Confidentiality, Data Non-repudiation, and Anti-jamming (Mahmood, Zen, & Hilles, 2018).

[18] USDOT, Draft Final Report Hillsborough MPO Pilot of the Data Business Plan Guidance for State and Local DOTs: Data Business Plan. November 2016. Accessed on 4/15/2019 from http://www.planhillsborough.org/wp-content/uploads/2016/12/FHWA-Data-Business-Plan-DRAFT_DEC.pdf.

[19] US DOT Roadway Transportation Data Business Plan, Phase 1, Appendix D.

[20] Data Business Plan. Minnesota Department of Transportation, October 2007, p. 43

[21] FHWA, 2015. Data Governance and Stewardship Program Primer, https://www.fhwa.dot.gov/datagov/dgpvolume%201.pdf

[22] Green and Lucivero, 2018. Data Governance & Data Management: Case Studies of Select Transportation Agencies https://www.gis.fhwa.dot.gov/documents/GIS_Data_Governance_and_Data_Management_Case_Studies.pdf

[23] USDOT, Draft Final Report Hillsborough MPO Pilot of the Data Business Plan Guidance for State and Local DOTs: Data Business Plan. November 2016. Accessed on 4/15/2019 from http://www.planhillsborough.org/wp-content/uploads/2016/12/FHWA-Data-Business-Plan-DRAFT_DEC.pdf.

[24] Ibid., pp., 6-8.

10. Appendices

Appendix A: Acronym Table

ARC	Atlanta Regional Council
DAMA	Data Management Association International
DCC	Digital Curation Centre
DG	Data Governance
FHWA	Federal Highway Administration
IRB	Investment Review Board
IT	Information Technology
ITF	International Transportation Forum
ITS	Intelligent Transportation Systems
KPI	Key Performance Indicators
OITS	Office of Information Technology Services
PII	Personal Identifiable Information
SOP	Standard Operating Practices
TMC	Transportation Management Center
TSMO	Transportation Systems Management and Operations

Appendix B: Workshop #2 Data Discussion Summary

Summary Results from Data Governance Exercise

Workshop #2: March 18, 2019

Question #1: Challenges to Sharing

- What are **3 major challenges** to your organization sharing data with other organizations today?

Topic	Details	Count
Inconsistent Structures, Formats, Semantics	• Having an abundance of data but not synthesized in a format to share • Incompatible systems or data formats • Structure of data collected (amount of detail needed or not needed) • Not having all data consolidated • Compatible formats	11

Topic	Details	Count
	• Understanding data in detail (type, scale, temporal, coverage, estimated vs. observed, etc) • Data sharing or integration takes a long time as a result of data type inconsistency and mapping difficulties • Data formats: what data with what format and what schedule • Useable data format to integrate in their system • Software integration of system capturing same or similar data (i.e., CAD/AVL in use by everyone in region but software is different) • Inconsistent / incompatible data formats coming from multiple vendors or contractors	
Inconsistent Access / Challenges to Access (platform)	• network architecture – ports to get inside and outside secure network • Monitoring data. Maintenance and keeping service working • Data platforms that are universal • Platform or various program • The info is on the other organization's network and we don't currently have network integration • Network / firewall issues • Differing file sharing services • Signal software is not center to center • CCTV sharing between agencies • Working on different platforms	10
Data responsibility	• Unique point of contact of lack of data sharing protocol • Culture of cooperation and openness • Process of requesting • Getting permissions to share data through our legal department • Identifying the point of contact. Protocol, person knowledgeable about data source, network connection, firewall, etc. for access to a dataset • Information technology coordination • Organizational IT departments • Appropriate point of contact • Finding the right contact party • Slow response	10
Data Restrictions (license)	• Who owns the data • Restrictions on sharing due to contracting agreements / licensing • Private company data sharing restrictions • Third party data sharing • Unknown legal limitations • Policy restrictions • Data is licensed so legal agreements must be signed in order to share data • Data sharing agreements (too restrictive or nebulous) for data obtained / procured from private sector data vendors / sources	10

Topic	Details	Count
	• Legal issues • Need to record/retrieve camera footage from ratesign cams, but GDOTs cameras aren't recorded due to policy	
Access / discovery	This category is related to "Inconsistent Access / Challenges to Access (platform)" category. The Inconsistent access category is technology oriented. • Availability • Access to data • Lack of information about what data exists and which organization or department has it • Access to data for analytical purposes in an easy to use format • Not knowing which organization have what • Knowing who to share with… • Institutional bottlenecks (IT – "need to know security, proprietary, restrictions) • Internal – provisioning quick information aggregating key performance metrics	8
Too much data! Resources /costs	• Hosting large amounts of data • Resources (people) to manage and maintain data connectivity, storage, etc. • Understanding the cost / resource requirements to get clean, useable, consistent, complete operational data • Workload • Time availability of staff to create data sets to share • Cost of integration and cost of maintenance	6
Metadata / quality	• Data accuracy • Various and conflictions uses of the data • Accurate data • Differing needs / efforts across organization • Up to date data (regarding project implementation)	6
Privacy	• Privacy / privacy data public data • Privacy • Privacy issues	3
Too many choices (How do we choose the best data management approach?)	• Interface: with what interface we should go? Per security, performance, legal… • Formatting data properly for external consumption • I don't know what I don't know! Understanding the data landscape. I need on inventory, schema, and dictionary that stays up-to-data and accessible	3
Why collect	• No one knows what to actually do with data or why they even need it/want it	2

Topic	Details	Count
	• Undefined outcomes or objectives	
Inconsistent Geographic data	• Our geographic boundaries are not jurisdictional or census tracts or other standard. It makes giving and getting data sets very difficult	1

Question #2: Current Data Set Needs

- **List the 3 most important data sets that you need from other organizations today.**

Topic	Details	Count
Network performance Speed / Travel time	• Real time arterial performance • Operational data (ridership, hours, miles) • Travel speeds and volumes • INRIX speed • Travel times • Travel speeds • Private data – realtime data from 3rd parties like Motorola solutions • Speed data • Live roadside data from GDOT • Real time data • Local traffic • (location based data) • GDOT Navigator reports • Speed data	12
Planned / Project data	• Updated project implementation • Planned / upcoming transportation project data • RTP project level performance measures • TIP project level performance measures • Unit costs for materials (related to construction) • Current capital investment plans • Coordination for review of projects • Notification from GDOT for projects they are conducting in air jurisdictions (deployment horizons) • Land use plans	9
Analysis / Performance measures	• Equitable justice analysis (including environmental justice) • DASH includes all the above performance measures • Regional performance • Vehicle occupancy / transit ridership • Congestion data • ATSPM data • Origin-destination data (region-wide) including trucks and commercial vehicles • Corridor-level travel time reliability	7

Topic	Details	Count
Crash / safety data	• Live crash data • Crash / incident data • Consistent crash reporting • Accident data • Crash data • Safety numbers in real or near real time • Real-time crash data with attributes	7
Asset Management	• Device asset / maintenance data • Regional shared infrastructure resources • Asset inventory data • Maintenance data (cost for maintaining asset) • Natting of IP addresses for ATSPM	5
Traffic counts	• Traffic counts at nearby locations • Traffic data • Vehicle volume • Traffic analysis and data (level of service, counts) • Traffic counts and turn movement	5
Transit	• Transit asset • GTFS clean data for transit • Realtime GTFS • Transit info could be interesting • MARTA data	5
Geography and demographic	• Elected official district boundaries • Land use and related economic / demographic data • Tax digest information • Road centerline data updated with accurate location information • Address point data updated with location information • Socio-economic data	6
Camera	• More camera data and resulting analysis • Camera data, streaming from GDOT • Camera feeds • Historical (recorded) camera feeds of rate sign cameras	4
Traffic Signal plans and operational data	• Real time traffic signal data • High resolution signal data • Signal updates • Traffic signal timing plans • Platoon releases from adjacent jurisdiction	5
Incident	• Near real-time incident data (accidents, break downs) • Incident data • Roadway clearance time	3

Topic	Details	Count
Work Zone / closures	• TIR / construction data • Real time road closure data with accurate location information	2
Multimodal	• Mobility delays (bike, ped, bus) • Rideshare data	2
TSMO	• TSMO info	1
Lane closures	• Lane closures (planned and unplanned)	1

Question #3: Future Data Set Needs

- **Given projects under deployment, list the 3 most important data sets you expect to share in the future?**

Topic	Details	Count
CAV	• Autonomous vehicle/shuttle ridership by TOD and amount of travelers • DSRC • V2I data • Connected vehicle message format uniformity • Real time performance measure from CV • AV/CV data • Connected vehicle data • CV data as market proliferation grows • CV data • BSM data stream (basic safety message from connected vehicles) • Smart city pilot project data SPaT, smart lighting, signal priority preemption • SPaT	13
Transit	Transit Service Information • GTFS, GTFS-realtime • GTFS-realtime • Real time transit data • Transit Transit Performance • Regional transit reliability • Farebox data • Regional transit operational KPIs (calculated across data types) • On time performance / connections	4 +4

Topic	Details	Count
Performance	• Transportation performance • Manage lane performance • Travel time reliability • Future planning performance data such as visual evaluation (large scale not detail) • Comprehensive performance visualization tool (DASH TDM option planit type) • Usage / utilization (modal) • Emissions reductions / inputs • Performance metrics from smart corridor technologies (pre-emption, bus priority)	8
Traffic Data	• Traffic signal data • Traffic data • Traffic • Traffic counts by time of day/location/mode • Freight related demand by time of day and by geography • Bottlenecks • Traffic counts and turn movements	7
Project, planning and economic data	• Planned project data • Current capital investment plan • Economic impacts at major projects • economic development info (zoning, variances, reviews) • factors to help determine where to locate charging stations (e.g., population, volume, car-ownership, etc.)	5
Multimodal	• Shared bike • TNC, uber/lyft • Vanpool • Shared scooter • carpool	5
Asset Mgmt	• Connected vehicle OBU/RSU asset management data • Asset inventory using software • Roadway asset information • Asset management / vulnerable infrastructure	4
Travel Time	• BlueTOAD travel time data • Travel times • Travel time data	3
Crowd sourced data & big data sources	• Aggregated user location information for example, a device user that is using a tree map or location app in a device has agreed to share their location data. That data is compiled into an aggregated (anonymized) data set to display locations where users are present. Location data such as roads, walkways, etc. can be derived and use to detect location where map data updates. • User generated content • weather	3

Topic	Details	Count
Occupancy, volume	• Vehicle occupancy patterns in express lane corridors • Realtime data in express lanes (trips, volume)	2
Crash data	• Crash statistics • Revamp of crash reporting forms to include clean/consistent crash data	2
Fleet Management	• Vehicle trajectory information • Computer aided dispatch (connected to traffic control system)	2
Payment / Security	• Security credentialing tokens / information • Decentralized ledger information	2
Safety	• HSM predictive analytic data for data driven safety • Predictive analytics based on crash data	2
Dynamic pricing	• Dynamic pricing data	1
Pavement data	• Pavement data	1
General	• Innovative data collection	1
Video	• GDOT rate sign camera footage	1
Closure	• Right of way closure information	1
Geography / Map	• Community improvement district boundaries	1

11. Bibliography

Ahmed, M., Gopalakrishna, D., Garcia, V., Ragan, A., English, T., Zumpf, S., . . . Hsu, E. (2016). *Connected Vehicle Pilot Deployment Program Phase 1, Human Use Approval Summary.* Washington, DC: USDOT.

Ahmed, M., Gopalakrishna, D., Garcia, V., Ragan, A., English, T., Zumpf, S., . . . Hsu, E. (2016). *Connected Vehicle Pilot Deployment Program Phase 1, Participant Training and Education Plan.* Washington, DC: USDOT.

Atlanta Regional Transportation System Management and Operations (TSMO) Visioning Workshop Summary. (2019, January 23).

CTA. (2019). *Chicago Transit Authority*. Retrieved from How we manage delays: https://www.transitchicago.com/performance/wtgw/#addressingbunches

Cupoli, P., Earley, S., & Henderson, D. (2014). *DAMA- DMBOK2 Framework.* DAMA International. Retrieved March 28, 2019, from https://dama.org/sites/default/files/download/DAMA-DMBOK2-Framework-V2-20140317-FINAL.pdf

Digital Curation Center Data Curation Lifecycle Model. (n.d.). Retrieved April 4, 2019, from http://www.dcc.ac.uk/resources/curation-lifecycle-model

Earley, S. (2015). *Presentation on DAMA-DMBOK2 and CDMP.* Retrieved March 28, 2019, from Dama International: http://dama-phoenix.org/wp-content/uploads/2015/09/DAMA-Phoenix-DMBOK2.pdf

FDOT. (2017, June 14). Florida Data Governance Policy. Tallahassee, Florida.

FHWA. (2015). *Data Governance Plan, Volume 1 Data Governance Primer.* Washington, DC: USDOT.

Firican, G. (2018, October 3). *Data Governance Maturity Models.* Retrieved April 4, 2019, from LIGHTSONDATA: http://www.lightsondata.com/category/data-governance/

Gidley, S., & Rausch, N. (n.d.). Best Practices in Enterprise Data Governance - White Paper. SAS.

Goetz, M. (2015, September 15). *Data Governance and Data Management Are Not Interchangeable*. Retrieved March 28, 2019, from Forrester: https://go.forrester.com/blogs/15-09-11-data_governance_and_data_management_are_not_interchangeable/

Gopalakrishna, D., Garcia, V., Ragan, A., English, T., Zumpf, S., Young, R., . . . Hsu, E. (2016). *Connected Vehicle Pilot Deployment Program Phase 1, Outreach Plan .* Washington, DC: USDOT.

Green, M., & Lucivero, A. (July 2018). *Data Governance & Data Management: Case Studies of Select Transportation Agencies.* Washington, D.C.: US DOT. Retrieved from https://www.gis.fhwa.dot.gov/documents/GIS_Data_Governance_and_Data_Management_Case_Studies.pdf

Kitchener, F., Gopalakrishna, D., Garcia, V., Ragan, A., English, T., Zumpf, S., . . . Hsu, E. (2016). *Connected Vehicle Pilot Deployment Program Phase 1, Performance Measurement and Evaluation Support Plan.* Washington, DC: USDOT.

MnDOT. (2011). *Data Business Plan.* Minnesota Department of Transportation. Retrieved from https://www.dot.state.mn.us/tda/databusinessplan.docx

MnDOT. (2019). *MnDOT Policies - Data Data Stewardship*. Retrieved from Minnesota Department of Transportation: https://www.dot.state.mn.us/policy/dm/dm002.html

NASCIO. (April 2008, April). *Data Governance - Managing Information As An Enterprise Asset - An Introduction.* Retrieved from NASCIO Governance Series: https://www.nascio.org/Portals/0/Publications/Documents/NASCIO-DataGovernance-Part1.pdf

National Academies of Sciences, Engineering, and Medicine. (2010). *Target-Setting Methods and Data Management to Support Performance-Based Resource Allocation by Transportation Agencies - Volume I: Research Report, and Volume II: Guide for Target Setting and Data Management.* Washington, DC: The National Academies Press. doi:. https://doi.org/10.17226/14429.

ODOT. (2011). *Asset Management Strategic Plan.* Oregon Department of Transportation. Retrieved from https://www.oregon.gov/ODOT/Programs/Documents/AMI/04-amsp-10-111711_final.pdf

ODOT. (2018). *Oregon Public Transportation Plan.* Oregon Department of Transportation. Retrieved from https://www.oregon.gov/ODOT/Planning/Documents/OPTP_FINALDRAFT.pdf

OECD/ITF. (2015). *Big Data and Transport: Understanding and assessing options.* International Transport Forum at the OECD.

Thomas, G. (2019, March 28). *Defining Data Governance*. Retrieved from The Data Governance Institute: http://www.datagovernance.com/defining-data-governance/

USDOT. (Januaray 2013). *U.S. DOT Roadway Transportation Data Business Plan (Phase 1): Data Business Plan, FHWA-JPO-13-084.* Washington D.C.: US DOT. Retrieved April 4, 2019

Vandervalk, A., Snyder, D., & Hajek, J. (2013). *U.S. DOT Roadway Transportation Data Business Plan (Phase 1) - Data Business Plan.* Washington, DC: U.S. Department of Transportation.

What is Data Governance? (2015-6). Retrieved March 28, 2019, from The MDM Institute: http://www.tcdii.com/whatIsDataGovernance.html