NCHRP 08-36, Task 100

Transportation Data Self Assessment Guide

Requested by:

American Association of State Highway and
Transportation Officials (AASHTO)
Standing Committee on Planning

Prepared by:

Craig Secrest and Kyle Schneweis
High Street Consulting
and
Greg Yarbrough
Wilbur Smith Associates

August 2011

AASHTO
THE VOICE OF TRANSPORTATION

Disclaimer

Table of Contents

1   Executive Summary

Over the last several decades, the demands on transportation data programs to adequately support decision-making have grown and evolved. During the Interstate era, data needs were reasonably straightforward and focused, but data programs must now be far more diverse and robust to support current decision making processes and provide the level of accountability the public demands. Other factors that are causing changes to data programs and the demands made upon them include technological change, increased data collaboration and integration, financial considerations, federal requirements, data applications, and changing agency roles and relationships.

At the same time that the data world is rapidly evolving, the resources available to make needed investments in agency data programs has been shrinking and data program management processes in DOTs and MPOs are often weak or non-existent. In fact, few agencies have formal comprehensive data management programs in place or, where they do exist, they are ad hoc, inconsistent, and fail to achieve good strategic alignment. As a result, transportation officials are increasingly concerned and uninformed about the adequacy, direction, and management of their data programs. Senior leaders are often frustrated that data programs do not adequately support strategic-level decision-making, while data managers are frustrated that senior management cannot articulate what data is needed, data applications and uses are poorly understood, and there is often limited leadership support for improving data management processes and increasing data program resources.

> Note: for purposes of the discussion in this report, we selected "data management" to serve as catchall term for polices, processes, structures, standards, practices, etc. associated with the effective administration of agency data programs.

1.1   Study Purpose and Focus

The purpose of NCHRP Project 08-36, Task 100 is to propose a framework and conceptual design to serve as the preliminary thinking for the future development of a tool and/or resource to help transportation agencies assess the adequacy, direction, and management of their data programs. As such, the study seeks to address three key objectives:

1. Identify and define current practices that are applicable to development of a framework for a data program self assessment tool;

2. Understand the agency needs and challenges associated with data program management; and

3. Provide a conceptual design and road map for future development of a self assessment tool.

The focus of the proposed conceptual design is intended to go beyond creating a tool to help agencies evaluate "nuts and bolts" considerations about data quality. As illustrated in Table 1.1, the research effort strived to define a framework that also will help agencies assess the alignment of their data programs with strategic agency needs, and evaluate the adequacy and effectiveness of existing data management processes. The research also sought to describe how the proposed tool should fit into broader agency data management practices and considerations (e.g., gaining leadership buy-in, evaluating the current "as is" situation, establishing program goals and objectives, defining roles and relationships, developing and implementing data plans, and maintaining a regular program re-evaluation cycle).

Table 1.1: Key Assessment Categories and Topic Areas

Strategic Alignment	Data Quality
- Alignment with strategic goals - Clear and appropriate organizational roles - Alignment with user needs - Identification of data sources, uses, and users - Data utilization and visualization	- Accuracy - Consistency - Reliability - Timeliness - Completeness - Currency - Integrity - Confidentiality
Data Program Management Processes - Clear definitions - Ability to segregate, aggregate, and analyze - Time and resources for conducting analysis and visualization - Regular audits and validation procedures - Consideration for program trade-offs, costs and life-cycles - Mechanisms for security, privacy, and ethical considerations - Data collaboration - Management continuity	

1.2 The Need for Data Program Self Assessment

The need for transportation agencies to do a better job planning and managing their data programs is universally recognized and a self assessment tool is widely accepted as an appropriate means to help agencies do so. Opinions differ on why a tool is needed. Data staff often focus on considerations associated with data quality and want a tool that will help them improve their existing data resources and capabilities, data program managers want a tool that will help them assess and improve management processes, while senior management want a tool that will help their agencies align data programs and data spending with their strategic decision-making needs and associated priorities. That said, the different perspectives also share opinions about how a self assessment tool could help their agencies, including:

- **Provides a credible process for communicating with leadership and data program needs**
- **Better informed decisions**
- **Better understanding of deeds**
- **Greater collaboration/integration**

- **Increased cost effectiveness**
- **Added institutional capacity**
- **Better data inventory and quality**
- **Improved risk management**
- **Stronger credibility and transparency**

In addition to these benefits, the act of performing the assessment can help close the gap between the different perspectives. Specifically, assessments can engage senior management in a discussion about how the data programs are currently serving their needs, provide data staff with a means to inform discussion about trade-offs between new data collection and data management options, and create a credible assessment process that can justify the development of improved data management processes.

1.3 The Proposed Framework

The self assessment framework proposed in this report is intended to recommend a methodology that is straightforward, yet can cover the diverse range of transportation data programs and can be adapted to fit the unique circumstances and assessment needs of individual agencies. The framework is based on a

"maturity-model" approach, which is essentially technical jargon for comparing practices and/or performance against established benchmarks or scales of proficiency. It includes a three-step process that leads an agency through preparing for an assessment, conducting it, and then developing/implementing a plan for making needed program refinements and improvements (See Figure 1.2).

Figure 1.2: Framework Overview

1 PREPARE	**2** ASSESS	**3** IMPROVE
1. Identify project team	1. Set maturity targets	1. Analyze maturity results
2. Establish assessment goals	2. Assess current status	2. Develop management plan
3. Inventory data programs	3. Determine maturity gaps	3. Implement plan
4. Prioritize assessment needs		
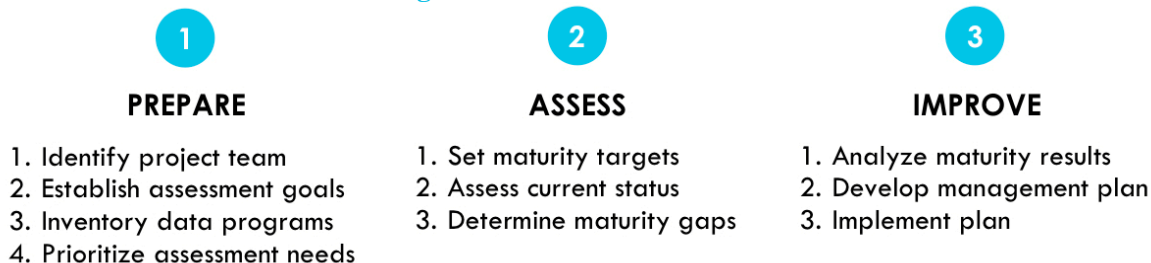
1.3.1 Assessment Framework Step 1: Preparation

The first step of the assessment will guide agencies in planning and preparing for a self assessment. This will include four specific activities:

1. **Creating a Project Team** – The planning and execution of the assessment should include representatives from different levels and disciplines throughout the organization. Having a strong project leader and support from leadership will be critical to the success of the assessment.

2. **Establishing Assessment Goals** – A clear direction should be set forth at the beginning of a self assessment initiative to ensure it produces the desired results. The goals should be influenced by user/stakeholder needs and outlined as clearly and specifically as possible to ensure that the goals are achievable and realistic.

3. **Data Program Inventory** – As part of the self-assessment preparation step, an agency will need to take stock of what data programs exist and how they are aligned within the organization. The main purpose of this activity is to establish a "map" of where data is gathered, analyzed, managed, housed, and used within the organization.

4. **Data Assessment Prioritization** – To keep the assessment focused and appropriately scoped, it is recommended that the project team prioritize among the inventoried data programs based on the established assessment goals to identify what data program elements are applicable to these goals and thus need to be assessed, and what about these elements need to be considered.

1.3.2 Assessment Framework Part 2: Gap Analysis

The second step in the framework focuses on evaluation of an agency's actual data programs through the application of "maturity scales," which will use transportation industry data and other relevant information and standards to establish "maturity intervals" that define different levels of performance for discrete data programs. The purpose of this component is three-fold: 1) it identifies the process through which an agency could assess the current status and performance of its data programs; 2) it provides guidance and information to help an agency determine where it wants its data programs to be; and 3) it guides an agency in analyzing its program shortcomings and identifying options to address them, and establishes a means for an agency to clearly understand and articulate gaps in its data program capabilities, characteristics, and management. These objectives are addressed through the three sub steps:

1. **Current Status Assessment** – For each relevant data element and assessment category, an agency will need to apply the applicable maturity scale and associated guidance/tools tools to assess the current status of their data programs. Performing this task will require a time and resource-intensive effort on the part of an agency to identify, collect, and assemble the information needed to support the assessment.

2. **Target Setting –** For each data element that is evaluated, an agency will need to determine the optimum maturity interval it should strive to achieve for each applicable assessment category.

3. **Gap Determination and Analysis** – By providing agencies with a means to determine both where their data programs currently are and where they need to be, the self assessment framework will facilitate identification of data program gaps and help agency officials translate these gaps into required investments, institutional capacity building needs, and new or revised process and policy requirements. It is anticipated that the primary means to conduct this analysis will be through a simple database tool, such as the example provided in Figure 1.3, which helps to visualize where the most severe gaps exist, identifies trends in those gaps, and facilitates analysis of which gaps present the greatest threat to the success of the organization.

Figure 1.3– Example Database Tool
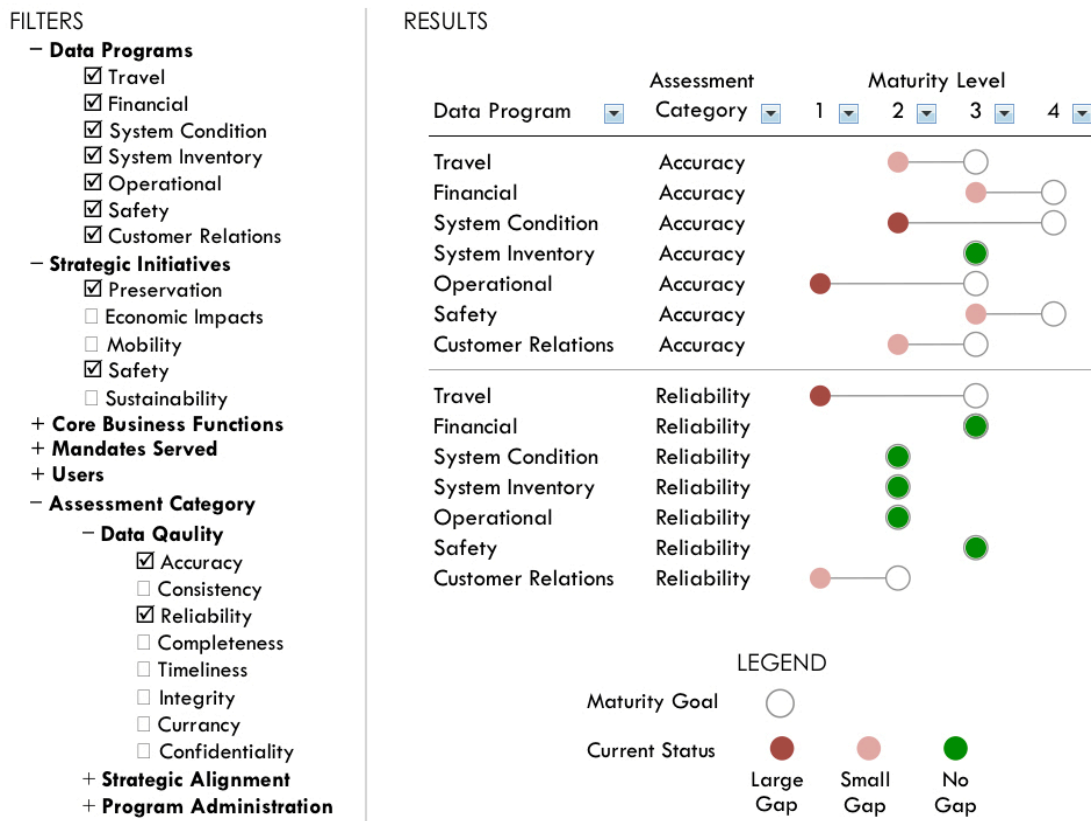
1.3.3 Plan Development and Implementation

Once the first two steps of the assessment have been completed, an agency should have a clear understanding of both how its various data program elements are aligned within the organization, and where gaps in program management processes and data quality exist. The final framework step is

designed to assist agencies in prioritizing needed data program improvements and provide guidance for the development of new or revised processes and polices that constitute an effective long-term data management plan. As part of this step, it will be important to help agencies understand that improving data management is a major undertaking that will take significant time, resources, leadership and cultural change. Components of this step consist of the following two activities:

1. **Plan Development** – Based on the results of the gap analysis, agencies should develop a plan to improve the ways data programs serve their organization. The plan should be consistent with the original goals of the assessment and consider the time, staff, and budgetary resources that are available, and any existing agency data management processes. The details and focus of the plan should address topics such as policies and procedures, management processes, organizational roles, definitions, monitoring, data availability, and needs prioritization.

2. **Plan Implementation** – The final task in the assessment framework is to implement the data management plan. The implementation process, like the plan itself, will vary from organization to organization and will heavily rely on the goals of the assessment itself, but should generally include communications, human resource, and enforcement components.

1.4 Potential Next Steps

The scope of developing a comprehensive data program self assessment tool for transportation agencies will be large and complex. The following is a proposed three-phased approach to framework refinement and implementation:

- **Phase I: Framework Review and Refinement** – An effort should be made to vet the proposed framework with a large cross-section of senior agency managers, data practitioners, and other interested parties and make any needed changes to the conceptual design.

- **Phase II Prototype Development/Proof of Concept** – A logical second step would be to develop and implement a prototype or targeted version of the proposed framework that could be used as a proof of concept and a learning mechanism to inform development of a comprehensive tool and resource. This effort also could include development of a high-level assessment mechanism that could be used to assess strategic alignment considerations and general data management processes for any data element.

- **Full Model Development** – Once a prototype has been developed, tested, and proven, the initiative could then be expanded to accommodate a broad range of transportation data program elements. It is also anticipated that the tool would essentially need to be a "living resource" and evolve as technology, the state of the practice, and agency needs change. Development of the full model would thus likely need to have either an ongoing maintenance component or require occasional updates.

2   Introduction

As transportation practitioners, when we make decisions we usually look for data to inform them. The right data helps us make the right decisions; transportation data programs and transportation decision-making are thus interwoven issues that must be considered together.

The demands on transportation data programs to adequately support decision-making clearly have evolved. Until the late twentieth century, transportation decision-making in the United States focused largely on the single mission of designing and building a world-class Interstate Highway System. As a consequence, transportation agencies' data needs for supporting decision-making, if not simple, were focused. Good data helped engineers build a road network that has anchored American prosperity for more than fifty years. Over the last several decades, however, transportation agencies have seen their missions grow more complex. As the Interstate System was completed, agencies shifted their focus to helping more people travel more efficiently on more congested roads, keeping them safe, offering choice in travel modes, protecting the environment, supporting economic growth and freight movement, and maintaining aging infrastructure.

At the same time, funding mechanisms set up to pay for building the Interstates have failed to deliver the funds that agencies need to meet their expanded roles. In this environment, decision-making has become much more multi-dimensional; agencies must now figure out how to allocate resources among competing needs and how to do more with less. They also are under greater pressure from the public to be accountable for their performance. The data needed to support decisions and provide accountability in this new environment is thus very different and much more robust from that of the Interstate era.

There are a range of forces that are influencing agency data programs themselves. In addition to the changing data needs associated with modern day decision-making, expanding federal requirements have created additional data demands, while shrinking agency budgets often limit the resources available for investment in (and maintenance of) data programs. Additionally, both technology and practices associated with data program management are evolving quickly, as are the way agencies must coordinate with partners and stakeholders on the collection and management of data.

The overall result is that transportation agency data programs have failed to keep pace for some time. In fact, a TRB Special Report from the early 1990s titled *Data for Decisions: Requirements for National Transportation Policy Making* called attention to the inability of transportation data programs to support modern day decision-making needs. Two decades later, despite some fixes at the national level, state and local transportation leaders are increasingly concerned and uninformed about the adequacy of and direction of their data programs.

2.1   Project Objective and Scope

The purpose of NCHRP Project 08-36, Task 100 is to propose a framework and conceptual design for creating a tool to help transportation agencies assess their data programs. Before reading this report, it is important to first understand what is meant by a "framework" – the intent of this relatively small research effort was not to develop an actual data program self assessment tool for transportation agencies; that will be a different, broader and much more resource-intensive initiative. Instead, the focus of this study can be thought of as preliminary thinking for the later development of a tool. As such, the study seeks to address three key objectives:

1. Identify and define current practices that are applicable to development of a framework for a self assessment tool;
2. Understand the agency needs and challenges associated with data program management; and
3. Provide a conceptual design and road map for future development of a self assessment tool.

The focus of the proposed conceptual design is intended to go beyond creating a tool to help agencies evaluate "nuts and bolts" considerations about data quality. The research effort strived to define a framework that also will help agencies assess the alignment of their data programs with strategic agency needs, and evaluate the adequacy and effectiveness of existing data management processes. The research also sought to describe how the proposed tool should fit into broader agency data management practices and considerations (e.g., gaining leadership buy-in, evaluating the current "as is" situation, establishing program goals and objectives, defining roles and relationships, developing and implementing data plans, and maintaining a regular program re-evaluation cycle).

2.2 Study Process

2.2.1 Literature Review

The study process for this project focused on building from the large bodies of knowledge about generic data program management and transportation decision-making, and applying it to develop a framework and conceptual design for a data program self assessment tool. Key steps included: 1) a literature review; 2) research on agency data programs, practices and self assessment needs; and 3) development of a proposed framework.

The focus of the literature review covered two areas. First, it investigated a range of strategic-level materials that, combined, provide a big picture perspective on the major trends and issues that are influencing transportation data program needs. Second, it explored a range of transportation-specific and general materials (e.g., reports, whitepapers, articles, and websites) that discuss data management and provide models, approaches, and guidance for assessing data programs.

The literature review produced a wealth of information that supported development of the proposed framework for conducting transportation data program assessments. Much has recently been written about the need for transportation decision-making to become more performance-driven and thus more data based. The literature documented the importance and potential benefits of good data management and data program assessment practices for both public and private sector enterprises. In addition, the research identified several different (albeit similar) models and structures for instituting data management and conducting data assessment. Finally, the body of materials reviewed provides insight into the challenges organizations may encounter and the factors that will contribute to success as they develop and implement data management and program assessment initiatives. Findings from the literature review are incorporated throughout the body of this report; complete documentation of the literature review is provided in Appendix A.

2.2.2 Agency Research

The project research team conducted a three-part targeted research initiative designed to inform development of a typology of agency data programs, to help identify appropriate data program assessment

parameters, and to support development of the self assessment framework. This effort included the following elements:

- **Review of Transportation Data Plans** – Limited examples where comprehensive data business or management plans have recently been put in place appear to exist in the transportation industry. The research team reviewed plans by the Alaska and Virginia Departments of Transportation (DOTs) and the Maricopa Association of Governments in Phoenix, Arizona, and several related documents from data offices within the Oregon, Michigan, Minnesota, and Washington DOTs. Review of these documents provided insight into how transportation agencies organize and manage their data programs and the criteria by which they measure success within their programs.

- **Identification of Existing Transportation Data Programs** – The research team sought to compile a comprehensive list of the data programs that are currently in use at transportation agencies. In addition to reviewing the sources cited above, the team conducted a wide-ranging search of existing research and publications, including Transportation Research Board Circulars on "Enhancing the Value of Data Programs: A Peer Exchange," and "Integrating Roadway, Traffic, and Crash Data: A Peer Exchange," and an AASHTO Standing Committee on Planning study on "Where Data Groups are Organizationally Located within State DOTs." The team also reviewed several U.S. DOT resources related to safety data, operations data, and the use of data for performance measurement. Several of these resources contained smaller case studies or references to data programs that allowed for an overview of existing data programs without a lengthy investigation of each.

- **Interviews with Transportation and Data Practitioners** – The project team conducted a series of phone interviews with members of the transportation and data sector, including agency executives who rely on data to inform their decisions and data practitioners experienced in collection, management, and analysis of data. The purpose of the interviews was to gather information from knowledgeable experts in the field about the challenges to managing data programs, the motivation for and benefits of better program assessments, the forces for change for DOT data programs, and how a self assessment tool might be constructed and used. A list of those interviewed for this project is included in Appendix B.

2.2.3   Framework Development

The proposed framework for the self assessment tool was developed through an iterative process. Based on the findings of the project background research, the consultant team established the three-step framework described in this report, and then worked with the project panel through the project deliverable review process to make refinements as needed.

2.2.4   Document Description

The remainder of this report is intended to document the overall findings and recommendations from the project.

- Chapter 3 summarizes the findings of the broad-based research on current data program management practices and associated considerations;

- Chapter 4 relates the findings from chapter 3 to the unique data program management practices, issues, and needs of the surface transportation industry, particularly DOTs and metropolitan planning organizations (MPOs);

- Chapter 5 presents the recommended self assessment framework; and

- Chapter 6 describes the work that will be needed to implement the framework and identifies potential next steps for developing an actual data program self assessment tool.

It is also worth pointing out a caveat about some of the terminology used in this report. As some of the literature outwardly acknowledges, the terminology associated with the management of data programs is undisciplined, overlapping, and often confusing. In particular, terms such as *data governance*, *data management*, *and data stewardship*, are often used interchangeably by data professionals and data customers alike, yet the terms also appear to mean different things to different people. For purposes of the discussion in this paper, we selected "data management" to serve as catchall term for polices, processes, structures, standards, practices, etc. associated with the effective administration of agency data programs.

3 Managing Data Programs

Development of a meaningful and useful data program self assessment tool should be informed by a strong understanding of practices from both the public and private sectors. The following section presents study findings, largely from the literature review and discussions with data practitioners, on experiences, information, and considerations that are relevant to transportation data program assessment and management.

3.1 What Are Good Data Management Practices?

As noted in a 2007 paper developed by the IBM Data Governance Council, "*many companies are just learning to examine their data governance practices, and searching for industry benchmarks and common frameworks to ground their approach.[1]*" Firms that specialize in data management and governance issues and public sector entities alike are finding they need to address:

- A lack of cross-organizational data governance structures, policy-making, and risk assessment causing a disconnect between business goals and data programs;

- Governance policies that are not linked to data requirements, data collection, or analysis needs and capabilities;

- Risks are not addressed from a lifecycle perspective with common data repositories, policies, standards, and calculation processes;

- Metadata and business glossaries are not used to track data quality, bridge semantic differences, and demonstrate the business value of data; and

- Controls, compliance, and architecture are deployed before long-term consequences are properly evaluated.

Clearly, there has been no shortage of thinking about how to respond to these issues – many articles and papers present ways to organize, explain, and conduct data management initiatives. The overall body of literature reviewed for this project implies good data management must be structured as an evolutionary and cyclical process. Consolidating information from the various flow-charts, tables, diagrams, and lists that accompany many of the articles on data management, the following appear to be the key steps in a meaningful data management structure (although the lines between these steps are often blurred):

- **Leadership Support and Direction** – An organization's leaders must recognize and embrace the need for data management. This means providing adequate resources, granting needed authority, providing input on direction and priorities, and supporting necessary policy and management changes.

- **Customer Outreach** – The relevance and adequacy of an agency's data programs should never be assumed, rather a process should exist for data practitioners to regularly interface with end users to better understand strategic data needs and inform both "valuation and risk assessment" and "goal setting."

---

[1] IBM Governance Council, Maturity Model: Building a Roadmap to Effective Data Governance, http://www-935.ibm.com/services/uk/cio/pdf/leverage_wp_data_gov_council_maturity_model.pdf (2007).

- **Valuation and Risk** – An organization should assess the value of data resources to an organization and evaluate the risk different data resources pose to the success and health of an entity.

- **Assessment** – Effective data management requires a comprehensive and honest assessment of the current "as is" situation with respect to strategic alignment, management processes, and data quality.

- **Goal Setting** – An organization must define where it wants to be with the various components of data management. Preferably, goals and objectives are established based on fiscal and other practical realities, and reflect both strong input from data users and stakeholders, and buy-in from senior leadership.

- **Defining Roles and Relationships** – A key part of management centers on defining and assigning roles and establishing clear lines of responsibility relative to data resources.

- **Developing Data Plans** – The results of the prior steps should be incorporated into a well-defined data plan that lays out clear steps (investments, new policies, standards and requirements development, changes in organizational structure, etc.) required to move an organization from its current state to a desired level of greater maturity.

- **Plan Implementation** – An organization must stick to its data plan, but remain flexible enough to adapt the plan as needed to respond to emerging threats and opportunities.

- **Re-evaluation** – At some regular interval (or in response to/preparation for a significant change in the operating environment) an organization should repeat the assessment process.

3.2 Components of Data Management

Many cross-cutting and recurring themes can be found throughout the literature regarding the attributes and components of effective data management; the same core values surrounding data management are relevant regardless of the field of expertise (transportation, health care, education, private sector, etc.). Although much has been written in regard to enriching data quality (accuracy, consistency, reliability, etc.), it is also apparent that effective data management requires a more comprehensive approach to how the data is defined, managed, and interpreted. The findings from the research on the components of effective data management are summarized below.

- **A Focus on Data Quality** – The most frequently mentioned topic in the literature reviewed was that of data quality, a term that encompasses the desired attributes of an effective data set. A core outcome of effective data management is an improvement in understanding, measuring, and monitoring the quality of one's data. The literature on this subject is extensive, in depth, and expansive in nature, but a few common attributes of data quality emerged:

    o **Accuracy** – The data should measure what it is intended to measure and errors should be minimized and easily identified;

    o **Consistency** – Data collection methods should be consistent and the data should remain consistent within and among data sets;

- o **Reliability** – Protocols for collecting and analyzing data should be in place to ensure that the data is reliable, regardless of who is using the information;

- o **Completeness** – Data sets should be appropriately inclusive and duplication should be minimized;

- o **Timeliness** – Data should be available within a time frame that supports its use;

- o **Integrity** – Data should be protected from deliberate bias or manipulation for personal, political, or other reasons;

- o **Currency** – Data should be appropriately up-to-date; and

- o **Confidentiality** – Appropriate security and disclosure procedures should be in place.

- **Alignment with Strategic Goals –** An effective data program must meet the strategic needs and support the mission, goals, and policies of an organization. For data management, this means qualifying and quantifying the value of data assets and programs to the core strategic goals of the organization. It also includes establishing processes though which decisions and recommendations can be based on data.

- **Clear Definitions** – Identifying clear and consistent definitions across data programs is important to ensure that the data sets are used and interpreted correctly. Definitions of a particular piece of data can often vary across organizational and programmatic boundaries. A common example is that of "revenue" – a finance department might have one definition of revenue, while an accounting department might have a different one, and a programming department might have yet another. These differences might be wholly appropriate given the applications within each department, but a clear and consistent identification of these differences is important to ensure that the data is managed used appropriately.

- **Ability to Segregate, Aggregate, and Analyze Data Longitudinally** – A necessary precursor to data analysis is a flexible data set that can be parsed, parceled, and constituted in whatever way is the most relevant to the desired application. Coordination in this area is necessary between and within organizations if data sets are to be shared across internal and external jurisdictions.

- **Regular Audits and Validation Procedures** – Errors will inevitably occur within any data program and audits and validation must be continuous and ongoing. As a part of data management, time and resources should be allocated to provide continuous examination of data for accuracy, consistency, reliability, etc. It also is important that discrepancies are properly documented to ensure that the process can be duplicated and reconstructed if necessary, and that corrections are carried throughout the affected programs and analyses.

- **Identification of Data Sources, Uses, and Users** – Understanding exactly where all of the data is coming from, who is using it, and for what purpose helps eliminate inefficiencies and demonstrates a clear tie between the sources and the expected outcomes. This mapping exercise is often done in the form of diagrams or flow charts.

- **Adaptable Data Management Plans** – The data management plan should be applicable even when policies, strategic goals, and leadership change. This does not alleviate the need to reevaluate data plans often to ensure that they are meeting the evolving needs of an agency, but

the core principles of the data management plan should be adaptable and scalable regardless of how circumstances change.

- **Clear and Appropriate Organizational Roles –** It is important that a data management plan be integrated within an organization. The roles of leadership, managers, data stewards, IT staff, data steering committees, etc. need to be clearly identified and articulated. Consideration should be given to the availability, capability, and training of appropriate staff. This might not apply just to those that are tasked with managing and analyzing the data. For example, an issue often discussed in education circles is that teachers and administrators are expected to use data to inform their approaches, yet many in this field have not been trained to adequately understand data analysis.

- **Consideration of Costs and Life Cycles –** Collecting, maintaining, and analyzing data comes with a cost. Data should be periodically analyzed to ensure that the time, staff, and monetary costs related to collecting a data set or managing a data program are appropriate. Inherent in this is recognition of when legacy data programs that are no longer providing a suitable benefit can be retired and an organization can identify opportunities to shift resources away from collecting and maintaining data that is not necessary.

- **Time and Resources for Conducting Analysis** – Data alone does not provide understanding. The missing link between the data managers and the decision makers is often the transformation of data into information. Good data analysis provides for an understanding of the relationships between data and desired outcomes through clear explanation and visualization techniques. A data plan should dedicate the necessary resources and provide guidance to ensure that data is being evaluated in a manner that is consistent with the desired application.

- **Mechanisms for Security, Privacy, and Ethical Considerations –** An effective data management plan must provide for adequate protection of data assets due to security, privacy, and proprietary concerns. Ethical considerations must also be applied when publishing analysis and visualizations to ensure that the results and conclusions are not manipulated to unfairly endorse a specific agenda.

3.3 Data Assessment Approaches

Literature on data assessment approaches and guidance largely focuses on the use of maturity modeling. As described in numerous articles and web pages, the concept of data maturity modeling grew out of the Capability Maturity Model (CMM), developed by the Software Engineering Institute, and can be applied to almost any data program, regardless of the application. Broadly defined, maturity modeling is the process of establishing a graduated path for improvement of an organization's data program management processes and activities, and then applying the framework to assess both where an organization currently is and where it needs to go. Examples of this desired progression include:

- From reactive to proactive data and information management;

- From project/ad hoc initiatives to enterprise-wide solutions;

- From "siloed" data to synchronized data and information;

- From localized systems with inconsistent levels of data classification and security to consistent data classification and standards based security; and

- Migration to the capability to build efficient information and knowledge management.

Potential outputs and benefits from using a maturity model approach could include:

- Credible identification, quantification/qualification, and articulation of potential benefits from improvements to data resources and data program management;

- Better organizational understanding of data management and its importance;

- Establishment of an incremental and evolutionary approach to improvement, including a process for determining performance measurement needs;

- A clear and objective assessment on the current state of an organization's data program management, including capabilities, strengths, weaknesses, opportunities and threats;

- A means to benchmark against peer organizations;

- Definition of applicable best practices and an incremental path that will help move an organization to the next stage of maturity;

- Establishment of short- and long-term data program goals and objectives;

- Establishment of the purpose and objectives of an assessment approach;

- Normalization of methods for continuously examining business processes and IT practices; and

- Documentation and centralization of reference information.

The original CMM identifies five levels of maturity; most of the evaluation approaches that build from the CMM suggest scales with four to six discrete steps that vary in complexity, focus, and comprehensiveness. There does not appear to be a specific "right" answer to the size and scope of a maturity scale, however, there are five common increments that are included in most approaches (a summary of several of the maturity models reviewed is summarized in the full literature in Appendix A):

1. **Undisciplined** – No formal data management approach;

2. **Ad hoc** – Data management occurs, but is typically in response to a specific problem or project rather than applied organization-wide;

3. **Standardized** – The organization has implemented some standard data management practices and roles;

4. **Proactive** – The organization has instituted formal data management policies and goals for all business units, has dedicated stewards for all business units, and has begun to both coordinate across the enterprise and measure/predict data program performance; and

5. **Optimized** – Data management is continually improved and used to set, communicate, and enforce business policies and strategic direction.

3.4 Challenges

Several challenges are associated with developing a data program self assessment tool and implementing it. As a result, development of a tool will likely need to be an evolutionary process that starts off imperfect and improves over time as solutions are developed and hurdles are overcome. Key concerns include:

- **Strategic Alignment** – Even after reviewing private sector experiences, there is little guidance on how to align data programs with strategic decision-making needs – the literature implies one "just do it." The issue generally is not that agencies do not know that they need to improve their alignment, in many cases, the problem is that they do not necessarily know how best to inform their decision-making with data.

- **Lack of Institutional Capacity** – Agencies may lack the internal capabilities to assess their own programs, and/or to implement improvements. Moreover, the workplace demographics of the industry will result in a significant brain drain due to retirement. The ability of agencies to retain information on legacies of their current data program will influence the effectiveness of self assessments and improvement initiatives.

- **Organizational Structures and Institutional Relationships** – While they share many similarities, transportation agencies tend to vary widely in terms of their responsibilities, organizational structures, resources, and relationships. Establishing a self assessment tool that accommodates the wide range of differences will not be easy and may require an extremely robust assessment mechanism.

- **Shortfalls in Financial Resources** – The deteriorating condition of transportation sector funding will likely put increasing strains on agency budgets; finding the resources to implement improvements and reforms in such an environment could be difficult.

- **Institutional Momentum** – Many elements of an agency's data program typically have a significant and complex history within the organization, thus there may be internal turf considerations, institutionalization of processes, and entrenched data audiences, all of which may create resistance to change.

- **Absence of Leadership Continuity** – Many transportation agencies (particularly DOTs) have fast turn-over in their senior leadership which can make it tough to "stay the course" with multi-year data program improvement initiatives.

3.5 Success Factors

There are several factors that will dictate the success of a data management framework and assessment. Many of these factors are reliant on strong organizational and leadership structures, and a clear understanding of the roles within the organization:

- **Strong Executive Leadership** – Like any comprehensive organizational initiative, leadership and support from the top of the agency is critical;

- **Solid Partnerships within the Organization** – Data management issues affect, and are affected by, many groups within an organization, collaboration and cooperation among information technology and the business units of the organization is key;

- **Good Communication** – It is important that the various communities of interest regarding the data and associated application systems that are used to collect, maintain, and report information are able to communicate and share ideas and resources;

- **Published Definitions and Standards** – Source data, metadata, and data reports from the various systems not only need to be clearly defined and standardized, but also formally adopted, published, and distributed; and

- **Data Champion** – Those within the organization that understand and appreciate the importance of data management need to be actively involved in the process and engaged in communicating the importance to others.

4 Transportation Agencies and Data Management

Data programs are clearly critical transportation agency resources, but the research shows that current agency management of these programs is at best immature and often non-existent. At the same time, a wide range of trends and considerations are driving the need for improved data program management by DOTs and MPOs. The following section discusses current transportation agency data practices, describes the factors that are driving data program requirements, and identifies the need for better data program self assessment.

4.1 Current Transportation Agency Data Programs

The state-of-the-practice for transportation agencies' data programs appears to vary widely, both from one agency to another, and between different categories of data resources. In some data program areas, such as system condition and safety data, approaches are reasonably mature – collection and analysis activities have been refined, historical data is available to support trend analysis and identify predictable cycles, data attributes have been improved upon, and users/uses have been defined. In these instances, most agencies perform at least some elementary level of data quality assessment, either as part of ongoing data improvement initiatives, or in conjunction with validation procedures for various modeling activities.

For many data areas, data program assessment seems to be fairly ad hoc, inconsistent, and (when it does exist) is too tactically focused. Few agencies have formal comprehensive data assessment programs and senior leaders are often frustrated that their data programs do not adequately support agency strategic-level decision-making. As a result, agency executives tend to feel "data rich, information poor." Similarly, data managers are often frustrated that senior management cannot articulate what they need from data programs. While data staff wants to better meet strategic data needs and generally improve both the quality and management of their data, they tend to lack adequate resources to do so and must contend with often competing (even conflicting) demands.

4.2 Forces for Change

One of the issues that complicate the process of data management is that many transportation data programs and associated program requirements have been changing rapidly in recent years. The increasing importance of data to strategic decision-making is at the heart of this change. In the past, data programs largely supported in-house activities such as project-level planning, various modeling efforts, and "after the fact" federal reporting requirements and system performance assessment. Now, data programs also need to inform more strategic activities such as complex trade-off decisions between different types of investment (e.g., modernization vs. expansion) and performance-based planning and programming. This, in turn, is changing both the type of data needed, requirements related to data attributes such as accuracy, timeliness, and completeness and the importance of translating data into information that can better inform decisions. Other factors that are causing changes to data programs include:

- **Technology** – Continuing improvements in computing capacity, data collection/processing methodologies and equipment, analytical tools, and communications technologies are changing the capabilities and efficiency of data programs, and increasing the range of functions data programs must support.

- **Collaboration and Integration** – In part because of technology, but also due to changing inter-jurisdictional relationships, expanded data needs, public sector financial constraints, and the emergence of private sector data resources, agencies must now both consider a broader audience of potential data customers and weigh external options for meeting data needs (and the implications of using them).

- **Public Demand for Accountability** – Growing public demand for increased access to information is increasing the need for data to be more accessible, better documented, and user-friendly.

- **Financial Considerations** – Expanding agency funding gaps, pressure to reduce administrative costs, and expanding demands combine to continually squeeze the resources data staff have available to fund data program activities and investments.

- **Federal Requirements** – Closely associated with the financial constraint consideration, federal data requirements (and in many cases state requirements as well) continue to expand. Most recently, the American Recovery and Reinvestment Act (ARRA) imposed a host of new data tracking requirements on agencies; many anticipate that federal performance reporting requirements will be established and further expand federally-driven data requirements.

- **Data Applications** – Data uses and data program requirements have grown over time, whether to support evolving modeling and analytical tools or to meet newly emerging data needs, such as support for traveler information activities or a shift toward performance-based planning and programming.

- **Changing Agency Roles and Relationships** – Data requirements have evolved as the transportation system has matured and agencies have shifted from an emphasis on building facilities to maintaining and operating them. Similarly, new agency strategic concerns, such as sustainability and environmental stewardship, and new collaborative relationships are changing what data is needed, how it is used, and how it must be managed.

4.3 Need for Self Assessment

The need for transportation agencies to do a better job planning and managing their data programs is universally recognized and a self assessment tool is widely accepted as an appropriate means to help agencies do so. Opinions differ on why a tool is needed. Data staff often focus on considerations associated with data quality – they want a tool that will help them improve their existing data resources and capabilities – while senior management want a tool that will help their agencies align data programs and data spending with their strategic decision-making needs and associated priorities. That said, the two "perspectives" also share many opinions about how a self assessment tool could help their agencies. The following is a complete list of these needs and/or potential benefits:

- **Better Understanding of Needs** – Help bridge the communication gap between data staff and senior management/decision-makers with respect to what data is needed to support strategic decisions, how these needs can be met and how future program needs may emerge;

- **Improved Risk Management** – Improve agency understanding of potential data program risks (e.g., inaccuracy, safety, security, confidentiality, technology/obsolescence, etc.) and identify needed and optional steps to mitigate these risks;

- **Cost Effectiveness** – Assist data staff in determining the benefit/cost of various data program options, and justifying selected strategies;

- **Better Informed Decisions –** Assist agency leaders and decision makers by improving data accuracy and analysis

- **Added Institutional Capacity** – Enable agencies to determine their human resource and other capability needs, assess their current and future ability to meet these needs, and identify strategies to address anticipated gaps;

- **Greater Collaboration/Integration** – Identify where collaborative and/or integrated approaches to meeting data needs may be possible and provide a means for weighing the pros and cons of such approaches;

- **Better Data Inventory and Quality** – Help agencies identify and understand what data resources they have and establish a means to catalog data uses, define what level of data quality is needed for specific data resources, and assess how well each resource is meeting these needs; and

- **Stronger Credibility and Transparency** – Establish a clear approach through which an agency can validate that it is following a defined and meaningful data management strategy.

While the literature is generally positive about the potential for organizations to establish a successful data management framework and conduct meaningful data assessments, most articles and papers also note that formidable barriers may need to be addressed or overcome. Moreover, there are several considerations that are unique or more pronounced in the public sector. The following is a composite of potential challenges organizations may need to address as they develop and implement data management reforms:

- **Management Indifference** – Management may fail to understand or appreciate the need for and relevance of data management and/or may not provide adequate support and sponsorship for assessment;

- **Weak Vision –** Management may be unwilling or lack the ability of to provide a clear strategic vision for an organization's data programs;

- **Management Turnover** – Frequent turnover in leadership and/or poor succession planning can hurt the continuity of data management improvement initiatives;

- **Absence of Leadership Continuity** – Many transportation agencies (particularly DOTs) have fast turn-over in their senior leadership which can make it tough to "stay the course" with multi-year data program improvement initiatives.

- **Short Management Horizon** – Organizations and leaders tend to focus on short term gains, preventing data program managers from addressing long-terms needs and solving problems from a life-cycle perspective;

- **Consensus – building** – Gaining agreement of all necessary parties regarding policies can be difficult and time consuming, particularly when there are competing priorities within the organization;

- **Capacity Limitations** – The development or procurement of new tools and software may be required to enable more effective data management;

- **Financial Constraints** – The cost of conducting data assessments and implementing management policies can be prohibitive, particular during periods when agencies are under significant pressure to cut costs;

- **System Incompatibility** – Legacy systems with unique architectures may be difficult to accommodate under universal management policies;

- **External Requirements** – Particularly in the public sector, factors such as statutory and administrative requirements, grant-related stovepipes, and public demands for transparency and access to information may preclude the application of practical data management solutions;

- **Cultural Barriers** – A comprehensive data management approach often requires culture change to adapt the organization;

- **Task Magnitude** – The amount of work involved in developing a data management plan is often underestimated;

- **Poor Implementation** – Data assessment and management efforts are often long on structure and policies, and short on action;

- **Inappropriate Pace** – Trying to move too fast from "no-data" management to enterprise-wide data management can lead to mistakes and missteps; and

- **Organizational Resistance** – Both data professionals and data users often resistance to migration of data from silos to an enterprise management system.

5 Proposed Framework

The purpose of the self assessment framework proposed in this report is two-fold. First and foremost, it is meant to provide a conceptual design and guidance for the future development and use of a self assessment tool. As such, it defines a potential structure for how the tool could be organized and identifies the work required to develop it. Secondly, the framework itself is intended to provide agencies with guidance and information they can use now to develop their own approach to assessing the quality and usefulness of their data programs, and for making improvements to existing data management processes.

While the goal of this study is to design a self assessment framework that is as straightforward and simple as possible, the effort also has sought to propose a methodology that can cover the diverse range of transportation data programs, and that can be adapted to fit the unique circumstances and assessment needs of individual agencies. In many ways, these goals are in conflict due to the inherent complexity of agency data resources and uses, and must be balanced.
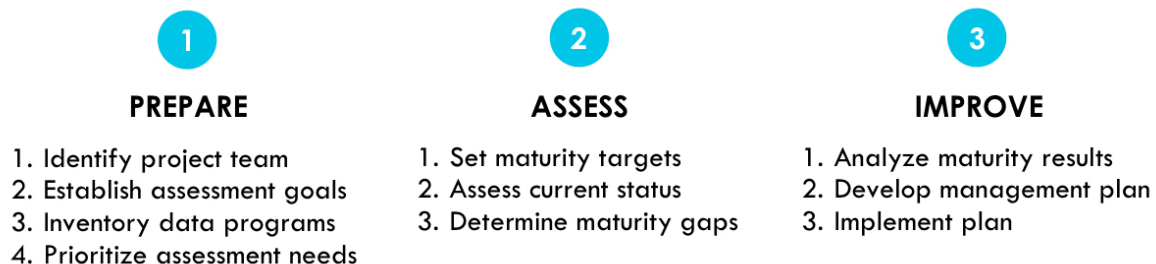
The following section introduces the proposed framework for a transportation data program self assessment and describes how it would be applied by agencies. Details about what will be required to actually develop and implement the framework are described in chapter 6.

5.1 Conceptual Overview

Based on the findings from the literature review, discussions with data practitioners, and input from the project panel, the project team determined that a maturity model-like approach should serve as the underlying structure for assessing the quality, effectiveness, and usefulness of agency data programs. This structure, which will use transportation industry information and benchmarks to define different levels of actual and potential data program performance, is highly versatile and can be adapted to provide assessment scales for both data quality considerations, as well as broader data program concerns such as the alignment of data resources with agency strategic needs and the effectiveness of data program management processes. It also is anticipated that this approach will allow agencies to customize the scope of their assessment and adjust assessment scales to suit their needs, resources, and interests.

The proposed framework is built upon the recognition that an assessment will not be meaningful if an agency does not adequately prepare for it, nor will it be useful if an agency fails to follow-up on the findings. The framework thus includes a three-step process that leads an agency through preparing for an assessment, conducting it, and then developing/implementing a plan for making needed refinements and improvements. This process is summarized in Figure 5.1.

Figure 5.1 – Framework Process

1

PREPARE

1. Identify project team
2. Establish assessment goals
3. Inventory data programs
4. Prioritize assessment needs

2

ASSESS

1. Set maturity targets
2. Assess current status
3. Determine maturity gaps

3

IMPROVE

1. Analyze maturity results
2. Develop management plan
3. Implement plan

5.2 Agency Data Program Typology

Transportation agencies currently have in place a vast number of data programs that range from traditional programs to collect and analyze travel patterns to new and emerging programs that measure the impacts of transportation on the environment and economy. Although the number and diversity of these data programs is nearly limitless, the research team recognizes that for purposes of assessment they should be categorized into a "typology." It is anticipated that this typology will provide an organizing structure for conducting a self assessment, allow for a nuanced maturity approach to be developed for unique data program elements, and make the assessment tool easier to use.[2] It is also expected that for some agencies a further breakdown within some categories may be necessary.

A recommended typology of the most common transportation data elements is demonstrated in Table 5.2. It is important to note that this typology is based on a somewhat limited level of research and industry outreach and focuses on areas that support transportation system decision-making. What is included captures the elements of transportation agency data programs that are most likely to require assessment, grouped into "data categories" based on the commonality of their sources, uses, and anticipated evaluation criteria.[3]

Table 5.2 – Recommended Data Program Categorization

Travel Data - Monitor and measure use and performance of the transportation system itself, the data is commonly collected through a combination of counting mechanisms and surveys: • Traffic volumes (historical and projected) • Congestion data • Speed data • Freight movements • Data for modeling (occupancy, travel time, trip type, turning movements, etc.) •	**System Inventory Data** - Record the location and describes the various components of the system: • Location data • System features and geometrics • Off-system data (such as land use, aerial photography, etc. - may require special considerations)
Financial Data - Monitor how an organization collects and spends its revenues; the data is generally collected and tracked in construction management systems and cash flow models: • Revenue and expenditures • Budget and cash flow • Labor/equipment/material costs	**Operational Data -** The most wide ranging of the data program categories, these data elements monitor how an organization uses it's time and resources: • Project schedule/status • Equipment management • Fuel and energy use • Weather/Road Condition (may require special considerations)

---

[2] The term "data element" is used throughout the document to refer to specific data resources such as speed and location data. The "data elements" are grouped into "data categories" such as "travel data" or "system inventory data." All data categories combined are an agency's "data programs."

[3] Modal data program, have been purposely omitted from this research due to the project size and available budget. It is expected that many of the principles used to define these categories will carry over to the modal programs. Similarly, human resources, and personnel data programs have also been left out.

System Condition Data - Track the physical condition of the system, often to support planning and programming decisions:	Safety Data - directly related to the recording of accidents, incidents, and behavior of system users:[4]
• Pavement and bridge conditions • Roadway feature condition (striping, signing, guardrail, etc.)	• Accident and fatality rates • Incident data • Enforcement • Behavioral (seat belt usage, etc.)
Customer Relations - Information that helps organizations to quantify customer satisfaction and effectiveness of their communications. • Surveys • Web Traffic • Stakeholder database	

5.3   Assessment Framework Step 1: Preparation

The first step of the assessment will guide agencies in planning and preparing for a self assessment. The purpose of this step includes four elements: 1) provide guidance on who to involve in the assessment and how to assemble the project team; 2) identify considerations that should be used to establish the goals of the self assessment; 3) outline the process by which an organization can take inventory of its existing data program elements; and 4) provide guidance on how to strategically prioritize the assessment. These preparatory activities are presented here in an order that will be suitable to most organizations, but the steps will likely inform each other, thus feedback loops between the sub steps may be necessary and prudent.

5.3.1   Establish a Project Team

Data program elements are often interwoven far and wide throughout an organization as data is collected, maintained, and used by a variety of people and groups within an organization. As a result, the planning and execution of the assessment should include representatives from different levels and disciplines throughout the organization. Who to formally include (and how) will vary by the intended purpose of the assessment and the unique structure/culture of an organization. The following are some general guidelines for establishing a project team:

- **Support from Leadership** – Ensuring full support and awareness from top management is a critical precursor to conducting an assessment. Leadership should be engaged early in the process to bring credibility to the initiative, improve support for the assessment and its recommendations, and facilitate communication between agency leadership, data users, managers, and data practitioners.

- **A Project Leader** – Project champions are essential for any reform effort within a large organization, especially one concerning an area as integral and pervasive as data. An effective leader who has support and creditability within the organization and a vested interest in the success and results of the assessment should be identified to lead and manage the effort.

- **Project Team** – A diverse team that represents the range of personnel involved with agency data should be assembled to guide an assessment initiative and ensure different perspectives and

---

[4] Note: data elements in other categories are also used in analyzing system safety but are not considered pure safety data elements and are thus not included in the safety data category (e.g., traffic volumes and roadway geometrics).

interests are adequately represented. A team might include those that collect data, manage data program elements, conduct analyses, provide technical/IT support, and use the data to make decisions.

- **Additional Assessment Participants** – Involving additional staff with intimate knowledge about the quality and management practices of specific data programs, or pulling in external end users of data-supported analyses to gain their opinions and insights, will likely be necessary in the later steps of the assessment (specifically, the maturity model evaluation described in Section 5.4). These participants do not need to all be identified at the beginning of the assessment; in fact, the need to include some individuals in various facets of the assessment will likely emerge when the data program inventory is conducted.

5.3.2   Establish Assessment Goals

One of the first responsibilities of the Project Team will be to define the goals of the assessment. Without a clear direction set forth at the beginning of the project, the assessment might easily lose focus, particularly in light of how far-reaching data programs and issues can be within the transportation industry. By collaboratively establishing a clear vision, the assessment will be more likely to produce the desired results. The goals should be outlined as clearly and specifically as possible and definition of the desired time frame, available budget, and staff resources for the effort should all be considered at this time to ensure that the goals are achievable and realistic.

When establishing the assessment goals, it is important to note that improving the quality or management of data in-and-of-itself is not valid or appropriate end goal of an initiative. Improving the data or data management processes might be *the means* to achieve a goal, but it should not *be* the goal. For example, an agency's assessment goal might be to determine how to improve the support current data programs and practices provide for core functions, and a finding might be that data quality improvements are need to improve to the level of support. In addition, the goals for an agency's assessment (and hence their overall data program) should consider both leadership needs and the information interests of key policy makers and "influencers" such as elected officials (and their advisors), the media, and interest groups. A major challenge in doing so is understanding the linkages between the data that is collected and the processed information that decision makers ultimately want, and then aligning the level of data program performance to adequately and cost effectively meet these demands.

For illustrative purposes, a few possible assessment goals are listed below. These are not intended to be all-inclusive, and it should be recognized that individual organizations must establish their own goals to ensure they reflect their unique assessment needs, ambitions, circumstances, and agency culture.

- **Business Function Support Goal –** To assess how well current data programs support the organization's core business practices, such as operating the transportation system, administering capital projects, tracking system conditions, administering federal and state funding, and providing vital information to the public.

- **Strategic Alignment Goal** – To assess how well the current data programs are aligned with and serve the strategic goals of the organization. Common strategic goals include safety, preservation, mobility, congestion, sustainability, and economic impacts.

- **Compliance Goal** – To assess how well the current data programs are adequately meeting mandates and requirements, such as federal system condition reporting mandates or state statutory performance reporting requirements.

- **Data Accessibility Goal** – To assess how effectively data and information is shared within an organization or among partners.

- **Effectiveness Goal** – To assess how efficient current data programs are designed and managed with regard to available staff, time, and budgetary resources.

5.3.3 Data Program Inventory

After establishing a project team and the goals of the assessment, an agency's next step will be to take stock of what data programs exist and assess how they are aligned within the organization. The main purpose of this activity is to establish a map of where data is gathered, analyzed, managed, and used within the organization. This information will serve as a foundation for conducting the assessment by providing insight into how the various data program elements are organized within an agency and by engaging and identifying the people within the organization (and perhaps externally as well) who use and manage the data.

The first task is to identify the relevant data programs to be included in the assessment. For a comprehensive agency-wide assessment, this step will involve developing a "laundry list" and categorization of all the data programs that exist in the organization. The categorization will vary, for some agencies, a high-level grouping of the programs may be sufficient (such as that in the typology presented in Figure 5.2) and for others a further breakdown may be necessary. For a more targeted assessment, such as one that seeks to assess how well data program elements support a certain strategic or core business need, only those elements that are directly involved with meeting theses needs should be identified and inventoried. It may also be important to identify other offices or agencies or partners that have similar data and to involve them in the inventory.

For each data program that is to be inventoried, the following items should be identified and recorded in a matrix (or some other template) such as the one illustrated in Table 5.3:

- Who is responsible for collection, management, and analysis?

- What is the source of the data?

- Who uses the data – both internally and externally?

- For what and how is the data is used for the following general categories:

  o Strategic level (input into system-level goal setting and other strategic decisions)?

  o Program level (input into project selection and resource allocation decisions)?

  o Project level (input into the planning, design, and construction of individual projects)?

  o Operational level (input to support into an organization's daily operations)?

  o Other applications (e.g., input to comply with mandates, support performance measurement, etc.)?

- Where is the data housed and archived (i.e., which databases is it stored in or connected to)?

- What strategic initiatives or goals does the data program element serve?

- What internal offices or other agencies/partners have similar data? Are there opportunities to reduce redundancies, improve consistency, and partner with others to reduce data costs and improve quality?

Table 5.3 – Data Program Inventory Example

Data Program: Travel Data Sub Program (if applicable): Traffic Counts	Strategic Initiavites: Preservation Economic Impacts Mobility Safety Sustainability	Core Business Practices: Develop statewide plan STIP development Administer MPO plans 511 highway Info Provide highway data
Data Location: XYZ Database	Sources: Traffic Counts ITS Live Counters	Collected by: Transportation Planning Urban ITS Centers
Users:	Categorical Uses:	Specific Uses:
Planning	Corridor Studies Forecasts Models	Project Level
Design	Road geometrics Bridge design Pavement design	Project Level
Safety	Intersection improvements	Project Level
Public Affairs	511 traffic conditions	Operational Level
Capital Programming	Project Selection	Program Level

After the individual data programs have been inventoried, it may be helpful to visualize how the data programs align within the organization as a whole. Developing an effective visualization will depend heavily on the goals and sophistication of the assessment. For example, if a goal of the assessment is to identify the strategic data needs of the agency, one should start not with the data programs but with the strategic initiatives. By visualizing which data programs are serving which initiatives, gaps and duplications may begin to emerge, providing insight into where opportunities exist for data systems integration, redundancy reductions, and new data collection. Other possible visualizations include data catalogues to outline what data exist and what they should and should not be used for and matrices or flow charts demonstrating how data programs are aligned within the context of the core business functions or where the data programs have similar sources, uses, and users. No one visualization technique or approach will meet the needs of every type of assessment, but an example visualization that might be used is illustrated in Figure 5.4.

Figure 5.4 – Inventory Visualization Example

Figure 5.4 – Inventory Visualization Example
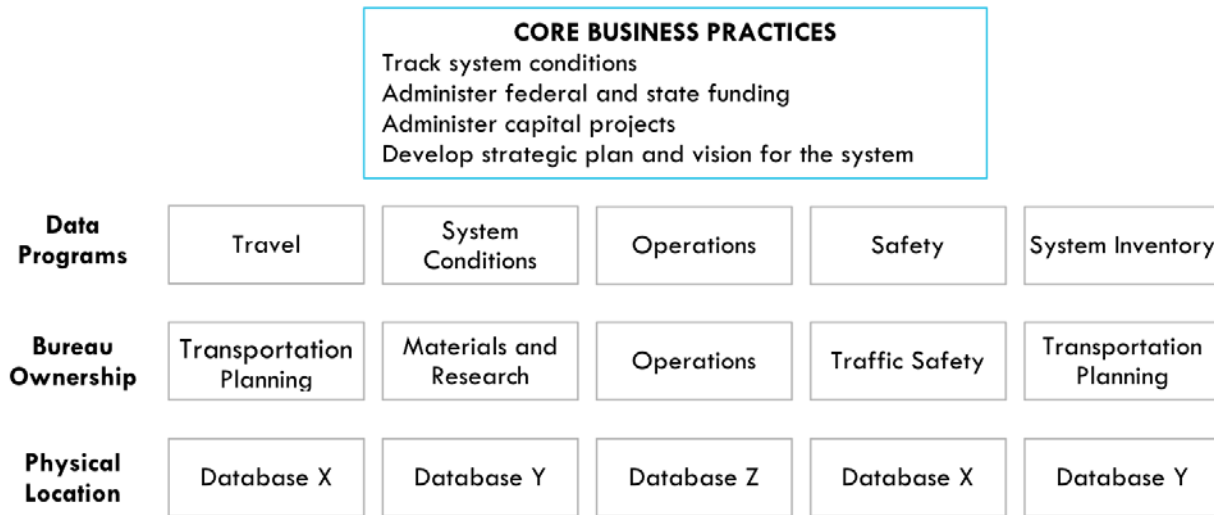
5.3.4   Data Assessment Prioritization

Conducting the data inventory described in the previous section should provide an agency with a solid foundation for the assessment without requiring a major expenditure of resources. The same cannot necessarily be said, however, for the subsequent steps of the assessment, as both the gap analysis and plan development can be resource intensive. To keep the assessment focused and appropriately scoped, it is recommended that the project team prioritize among the inventoried data programs based on the established assessment goals to identify those that are in most need of assessment.

Table 5.5 – Summary of Potential Assessment Categories

Strategic Alignment	Data Quality
• Alignment with strategic goals • Clear and appropriate organizational roles • Alignment with user needs • Identification of data sources, uses, and users • Data utilization and visualization	• Accuracy • Consistency • Reliability • Timeliness • Completeness • Currency • Integrity • Confidentiality
Data Program Administration • Clear definitions • Ability to segregate, aggregate, and analyze • Time and resources for conducting analysis and visualization • Regular audits and validation procedures • Consideration for program trade-offs, costs and life-cycles • Mechanisms for security, privacy, and ethical considerations • Data collaboration • Management continuity	

To conduct the prioritization, the project team should revisit the assessment goals and identify what data program elements are applicable to these goals and thus need to be assessed, and what about these elements need to be considered. As identified in Table 5.5, the framework breaks these considerations into three "assessment categories:*" Strategic Alignment, Data Program Administration,* and *Data*

Quality. Additionally, consideration should be given to the available resources, including staff and project budget, and the desired schedule. This step also may serve as a feedback loop to the goal setting task and lead to some changes to the overall direction of the assessment.

5.4 Assessment Framework Part 2: Gap Analysis

The second step in the framework focuses on evaluation of an agency's actual data programs. The purpose of this component is three-fold: 1) it identifies the process through which an agency could assess the current status and performance of its data programs; 2) it provides guidance and information to help an agency determine where it wants its data programs to be; and 3) it guides an agency in analyzing its various data program performance gaps and identifying options to address them, and establishes a means for an agency to clearly understand and articulate gaps in its data program capabilities, characteristics, and management to senior leadership. These objectives are addressed through the three sub steps discussed below: target setting, current status assessment, and gap determination and analysis.

The primary means for conducting this evaluation will be through the application of "maturity scales," which will use transportation industry data and other relevant information and standards to establish intervals that define different levels of performance for discrete data programs with respect to specific strategic alignment, data program administration performance, and data quality considerations (see Table 5.5). For example, to support an assessment of the *timeliness* of an agency's *traffic data*, the self assessment guide might include a maturity scale such as the one demonstrated in Figure 5.6.

Figure 5.6 – Illustrative Maturity Scale for Traffic Data Timeliness[5]

Maturity Interval	Description
Undisciplined	Counts are rarely taken, growth factors are applied
Ad Hoc	Annual counts are taken on up to XX% of the system and a few accurate counts can be gathered in a timely fashion if needed for a specific project Data is available for users with in X months of collection.
Standardized	Accurate counts are taken at appropriate intervals on XX% of the system and at least once every X years for the entire system Data is available for users with in X months of collection
Proactive	Continuous counts are taken on at least XX% of the system and accurate counts are taken at appropriate intervals at least once every X years Data is available for users with in X months of collection

5.4.1 Current Status Assessment

For each relevant data element and assessment category, an agency will need to apply the applicable maturity scale and associated guidance/tools to assess the current status of their data programs. Performing this task will require a time and resource-intensive effort on the part of an agency to identify, collect, and assemble the information needed to support the assessment. For example, to adequately apply maturity scales associated with traffic data accuracy, the data program assessment team might need to conduct interviews with the associated data collectors and users, review the data collection methodologies that are being used, and/or conduct independent tests to sample and assess the accuracy of the data. The requirements for these efforts, as well as guidance on how to conduct needed analyses, will presumably

---

[5] Based on the findings from the literature review, it may be appropriate to create a fifth interval for "optimization," once the industry data program management has improved and matured.

be provided as part of the resources that are developed to support implementation of the framework (see Chapter 6 for additional details).

As with target setting, the process for establishing the current maturity status of various data program elements will require a balance between qualitative analyses and subjective assessments depending on the nature of the data element and the assessment category.

5.4.2   Target Setting

For each unique data element that is evaluated, an agency will need to determine the optimum maturity interval it should strive to achieve for each applicable assessment category. While much of this effort will be driven by considerations unique to a given data elements, examples of general questions an agency might need to answer to establish their targets include:

- What are the minimum requirements to support the various uses and applications of the data element?
- What is anticipated benefit/cost of a higher level of performance?
- What level of performance is feasible or practical?

It also is important to recognize that the analytical rigor of the target setting process is likely to vary by assessment category. In areas related to data quality, strict benchmarks and metrics may be available and targets may be fairly clear, whereas the need to achieve a given level of performance for strategic alignment consideration may be murkier to determine. In such cases, targeting setting may be as much "art" as it is "science." Regardless of the assessment category, the target setting effort should be achieved through some form of consensus-based decision-making that involves individuals that collect, analyze, maintain, and use the associated data.

5.4.3   Gap Determination and Analysis

By providing agencies with a means to determine both where their data programs currently are and where they need to be, the self assessment framework will effectively facilitate identification of data program gaps and help agency officials translate these gaps into required investments, institutional capacity building, and new or revised processes and polices. To do so, the data assessment project team will need to analyze the assessment results to answer key questions such as how does an agency's performance in a given data areas compare to that of their peers? What is the magnitude of the gap and is it a major problem? How relevant is addressing the gap to agency leadership and external decision makers? What needs to be done? And, what are the anticipated benefits of addressing the gap and/or the implications of not making improvements? In addition to informing development of a data improvement plan, the answers to these questions should provide data staff and other data interests with improved information for communicating the implications of data performance gaps to senior management and other decision-makers.

It is anticipated that the primary means to conduct this analysis will be through a simple database tool, such as the example provided in Figure 5.7, which helps to visualize where the most severe maturity gaps exist, identifies trends in those gaps, and facilitates analysis of which gaps present the greatest threat to the success of the organization. This database tool also will allow the project team to generate customized reports of the results, and in turn sort them in a fashion consistent with the assessment goals.

The customized view of the assessment results should provide the project team with an important tool to help prioritize among the identified gaps. For example, the analysis might reveal that a certain data attribute, such as data accuracy, is consistently below standard, or that a particular set of business practices, such as the administration of capital projects, is plagued by inconsistent data. Identifying these

Figure 5.7 – Example Database Tool
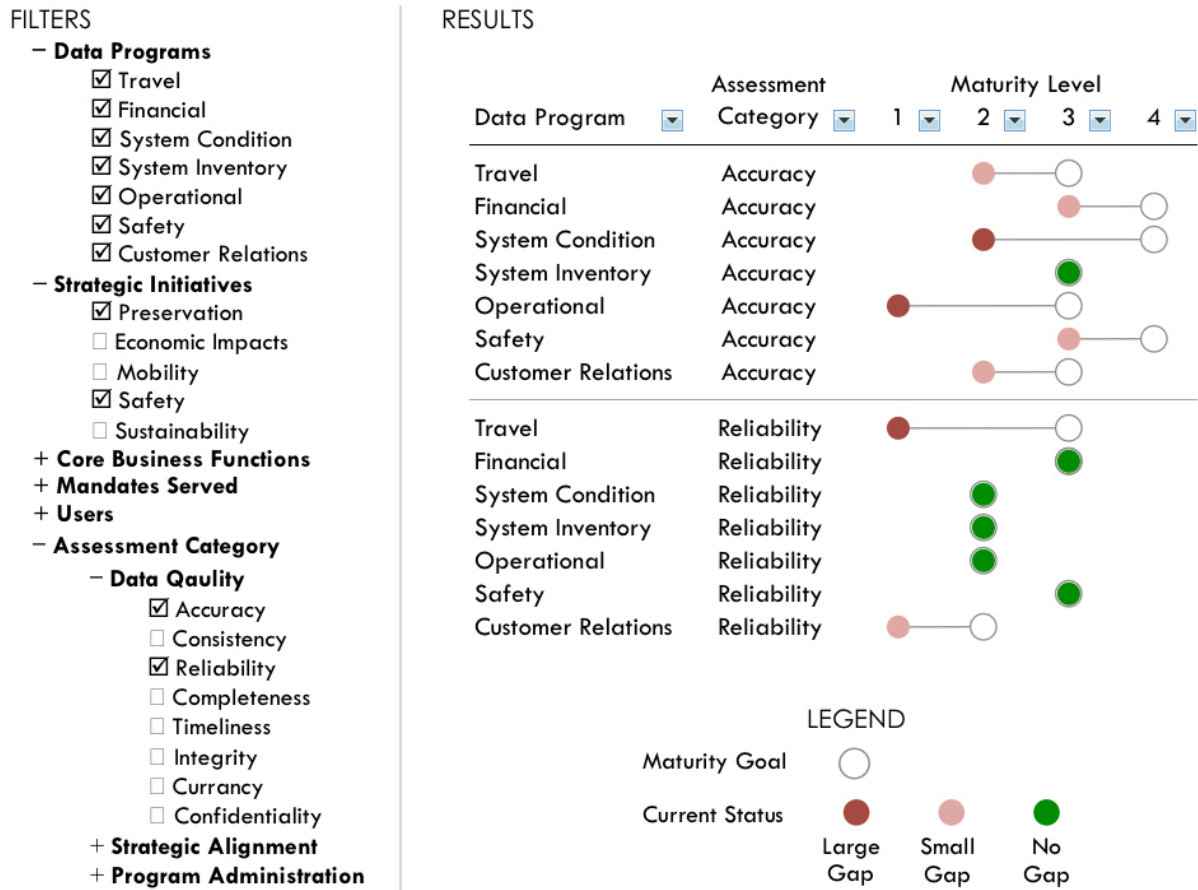
trends and comparing them with the established assessment goals will enable the project team to prioritize where to begin with plan development and implementation.

5.5 Plan Development and Implementation

Once the first two steps of the assessment have been completed, an agency should have a clear understanding of both how its various data program elements are aligned within the organization, and where gaps in data program maturity exist. The final framework step is designed to assist agencies in prioritizing needed data program improvements (as identified through the gap analysis step) and provide guidance for the development of new or revised processes and polices that constitute an effective long-term data management plan. This component consists of the following three activities: 1) analysis and prioritization of the identified maturity gaps; 2) development of a plan to address the gaps; and 3) implementation of that plan.

5.5.1 Plan Development

After the results of the maturity gap analysis have been evaluated, the next task in the framework is to develop a plan to improve the ways data programs serve the organization. The plan should be consistent with the original goals of the assessment and consider the time, staff, and budgetary resources that are available, and any existing agency data management processes. The details and focus of the plan will vary depending on the results of the assessment, but generally include the following topics:

- **Data Management Policies and Procedures** – Define the data quality standards and identify policies and processes to collect, maintain, analyze, and control the quality of data;

- **Organizational Roles** – Identify clear and appropriate roles for the management of data within the agency, including both typical data practitioner roles such as stewardship, analysis, management, and auditing, and roles for executive leadership as it relates to data management;

- **Definitions** – Establish clear definitions that will promote consistency across the organization and ensure that standard naming conventions and definitions are used;

- **Monitoring** – Establish audit processes to ensure the plan is being followed;

- **Availability –** Recommend strategies for sharing and integrating data within the organization through relationships and the use of technology; and

- **Needs Prioritization –** A process for identifying priorities and trade-offs among the various identified needs, such as investing in new data collection, technology improvement, or system integration.

5.5.2 Plan Implementation

The success of any plan ultimately lies in how completely and effectively it is implemented. Thus, simply put, the final task in the assessment framework is to implement the data management plan. Perhaps equally important, the data program assessment and the resulting data management plan should not be viewed as one-off events or short-term initiatives. Indeed, the assessment should produce a process that is repeatable and lead to a long-term plan that helps to institutionalize effective data management within an organization. The implementation process, like the plan itself, will vary from organization to organization and will be rely heavily on the goals of the assessment itself, but there are some general considerations that should be incorporated into most any implantation approach:

- **Communications** – Establish a communications structure that facilitates coordination within the organization and its partners. This may include a data management or governance council to identify coordination opportunities and to resolve any issues or conflicts that arise.

- **Human Resources** – Make a concerted effort to invest in staff training to ensure that those tasked with collecting, maintaining, analyzing, and managing the data are adequately equipped to do so.

- **Enforcement** – Create a means to maintain and enforce management policies and procedures. Just having the policies in place is a good first step, but long-term data management requires that those policies actually be followed and enforced. This will include conducting audits as outlined in a management plan, and monitoring the established organizational roles and responsibilities to ensure that the appropriate people are in place as staff changes occur.

6 Self Assessment Tool Development Needs

The prior chapter details a general framework and conceptual design for a data program self assessment process. How this framework is implemented and the resources that are developed to support its application could take on a variety of forms. A critical question is what the format of the final product should be.

On one end of the continuum, the final product of this initiative could be a simple guidebook that walks agency staff through the steps of a self assessment process and provides resource materials to support various analyses and decisions (e.g., a series of potential maturity scales for specific data program elements and assessment categories). Alternatively, the product could be a highly-automated model/software program, where users enter information and are provided assessment results and recommendations (e.g., based on responses to an on-line questionnaire, the tool would identify an agency's maturity needs, current maturity status, associated gaps for selected data programs, and recommended steps for improvement). An option somewhere in between is an online guidebook with access to selected support resources (e.g., spreadsheets, information on standards and practices, visualization tools, etc.).

While the "format question" will need to be addressed before an effort to move the data program self assessment initiative from concept to product can be fully completed, the framework proposed in this report provides significant direction on the research, data collection, analytical methodologies, and process development that will be needed to implement it, regardless of its final form. The following section provides an overview of these anticipated work requirements and is intended to serve as input for scoping development of the final self assessment product. It is generally organized by the various activities that make up the three steps of the proposed framework.

6.1 Data Program Typology

The typology of agency data programs identified in Section 5.2 is based on a fairly small sampling of state DOTs and MPOs and has not been tested in any practical fashion to assess whether it is a meaningful way to group and assess the various agency data programs. Fuller development and finalization of the typology will require:

- A complete (or at least extensive) survey of state DOT's and MPOs to identify, better understand, and characterize the full breadth of data elements that may need to be addressed as part of any transportation data program self assessment;

- A process to vet the proposed typology with agency data practitioners (e.g., a series of workshops); and

- Some form of beta test to ensure the final typology adequately supports the rest of the self assessment tool.

6.2 Assessment Planning and Team Formation Modules

The development of directions, advice, and background information to guide agencies through the initial stages of assessment planning will require a combination of common sense process development and transportation industry research. Specific development requirements will include:

- Guidance on the range of options for missions, roles, and structures (including selection of the day-to-day project manager) that could be adopted by agencies to establish assessment project teams, and an evaluation of the pros and cons of the different options;

- Interviews with agency personnel and analysis of agency organizational structures to identify the types of agency positions that could potentially need to be included in an assessment project team (or involved in the assessment in other fashions) based on the anticipated scope of the assessment initiative;

- Development of a questionnaire (and potentially a tool) to help agencies determine who to include on their assessment project teams;

- Interviews with agency senior managers and data practitioners to define a list of potential data assessment initiative goals that agencies can use as straw men to define their own goals; and

- Identification of other critical questions and issues an agency should strive to address to effectively plan their data program self assessment and, where possible, development of guidance to help answer and/or address these questions and issues.

6.3 Data Inventory Module

The development of a data program inventory will be a critical input to an agency's overall data program self assessment effort. To provide adequate direction, advice, and resources to support this important activity, development of the data inventory component of the framework should include the following:

- Development of step-by-step directions on how to conduct a data program inventory and research on appropriate background information to support the inventory (e.g., the product should define the various categories included in the inventory, explain their relevance, and identify the type of agency positions that should be interviewed as part of the inventory process);

- Development of questionnaires (either for individual or multiple related data elements) that can be used to conduct inventory interviews (developing these questionnaires themselves will likely require industry research and outreach);

- Development of templates that can be used to record and organize inventory findings. This could range from a simple template to an on-line form with drop-down tables that adjust based on specific selections (e.g., once you select a data category, you would be provided an option to select applicable sub categories); and

- Identification of recommended approaches for visualizing how the inventoried data programs align within the organization as a whole (e.g., generic flow charts and matrices), analysis of the pros and cons of the different approaches, and development of guidance on how the different approaches can be applied.

6.4 Assessment Prioritization Module

The materials to support the assessment prioritization part of the framework will need to guide users through the steps for prioritizing the issues and identify considerations to be addressed in a data assessment effort. Specifically, it will need to instruct users how to apply the identified assessment goals

to the results of the data program inventory to refine the scope of the self assessment. Development of the guidance for this module will thus need to focus on two elements:

- Identifying the key questions an agency should ask of itself (and perhaps of its partners and stakeholders) to determine its assessment priorities; and

- Identification, description, and development of guidance on the use of prioritization decision frameworks that can be used by agencies to support their assessment prioritization efforts. These could include quantitative approaches that use weighted scores for various criteria, or visual approaches such as ones that plot data elements on an X-Y axis based on two key considerations (e.g., strategic importance and ability to achieve improvements).

6.5 Maturity Gap Analysis Step

The bulk of the work associated with supporting and facilitating the maturity gap analysis step will be centered on the development of the maturity scales (discussion separately below). The remainder of the work associated with supporting this step will be fairly straightforward and include the following:

- Development of questionnaires (and potentially an analytical tool) to help agencies establish their desired or needed level of maturity on individual scales, and to determine where they currently are on the scales; and

- The development of guidance to help agencies adjust their desired or needed maturity levels based on practical considerations such as resource limitations and institutional barriers.

6.6 Development of Maturity Scales

In large part, the degree to which the self assessment tool will be able to assist an agency in identifying and addressing data program improvement needs will be determined by the quality, applicability, meaningfulness, and adaptability of the maturity scales that are developed for the tool. Establishing such scales will require a significant development effort that includes the following elements:

- Identification and definition of the specific maturity scales that will need to be established to support the model (this will be driven by the development of the final data program typology);

- Research to identify and assemble the information needed to inform the definition of different levels of maturity for discrete data elements; this will likely need to include surveys, interviews, and considerable analysis to establish various industry benchmarks and commonalities;

- Development of the actual maturity scales; and

- Outreach and consensus building with agency leaders and data practitioners to validate the maturity scales.

> **What should the scales be?** There is no definitive right answer for either how many maturity intervals should be established, or what level of performance each level should seek to define. The focus should be on developing scales that will facilitate the most meaningful assessment and planning. The scales for assessment categories associated with strategic alignment and data program administration could be fairly simple (say three intervals), while the scales for data quality considerations will likely need to be more complex.

6.7　Plan Development and Implementation Step

The final element of the framework that will need to be implemented is development of the guidance and information that agencies could use to help transform self assessment findings into a meaningful data management plan. Doing so will require the following:

- Developing guidance on how to characterize maturity gaps in different data areas and talk to agency leadership about their relevance and implications;

- Conducting research and analysis to develop mechanisms for quantifying and/or qualifying the benefits that could result from data program performance gaps;

- Conducting research and establish informational resources on strategies for addressing data program performance gaps; and

- Developing frameworks to help agencies assess the costs and benefits of gap closure strategies and prioritize associated actions.

6.8　Potential Next Steps

As reflected in both the typology of data program categories and the litany of potential considerations associated with transportation data programs, the scope of developing a comprehensive data program self assessment tool for transportation agencies will be quite large and complex. the following is a proposed three-phased approach to framework refinement and implementation:

- **Phase I: Framework Review and Refinement** – As a precursor to developing the actual tool, an effort should be made to vet the proposed framework with a large cross section of senior agency managers, data practitioners, and other interested parties. In addition, it could provide an opportunity to address some of the validation needs identified above, such as confirming that typology of data program categories is "right." An appropriate level of outreach could presumably be achieved through a targeted survey of state DOTs and MPOS, and a series of three to five half-day workshops, perhaps held in conjunction with appropriate AASHTO meetings, the AMPO national conference, or TRB-related meetings.

- **Phase II Prototype Development/Proof of Concept** – A logical second step would be to develop and implement a prototype or targeted version of the proposed framework that could be used as a proof of concept and a learning mechanism to inform development of a boarder, more comprehensive product down the road. Such an approach would likely develop a self assessment guide/tool for one or two of the more mature data program categories, such as *system inventory* and *safety*, and a selected subsection of data assessment categories. This effort also could include development of a somewhat generic assessment mechanism that could be used to assess strategic alignment considerations for any data element.

- **Full Model Development** – Once a prototype has been developed, tested, and proven, the initiative could then be expanded to accommodate broad range of transportation data program elements. It is also anticipated that the tool would essentially need to be a "living resource" and evolve as technology, the state of the practice, and agency needs change. Development of the full model would thus likely need to have either an ongoing maintenance component or require occasional updates.

Appendix A: Literature Review

Overview

The literature review produced a wealth of information that can support the development of a framework for conducting transportation data program assessments. Much has recently been written about the need for transportation decision-making to become more performance-driven and thus more data based. The literature also does an excellent job at identifying the importance and potential benefits of good data governance and data program assessment practices for both public and private sector enterprises. In addition, our research indentified several different (albeit similar) models and structures for instituting data governance and conducting data assessment. Finally, the body of materials we reviewed provides extensive insight into the challenges organizations may encounter, and the factors that will contribute to success, as they develop and implement data governance and program assessment initiatives.

The primary weakness of the literature lies in the link between strategic data needs and governance/data assessment approaches. The materials on transportation trends and issues point to the need for better data and data analysis to support strategic decision-making, but say little about how data and data governance needs should change. Data governance and assessment information guidance typically identify "determining strategic intent" as an important first step, but offer limited insight into how you determine strategic data needs (most approaches imply you simply ask your leadership). As result, there is an apparent gap in the literature on organizational approaches for improving the interface between leadership and those that manage and implement data programs.

Part I: Transportation Trends and their Implications for Data Programs

The first part of the literature review focused on a range of recent policy statements, reports, and meeting proceedings that have identified or commented on major trends and issues in surface transportation, with an emphasis on considerations that will or could have implications for state transportation data programs. A common topic that emerges from this literature is the growing overall emphasis on more performance-based, and by extension more data-driven, decision-making. Key elements of this trend include the following overarching industry issues:

- **Performance-based Decision-Making** – A wide range of recent publications, including the National Transportation Policy Project's (NTPP) 2009 report entitled Performance Driven: A New Vision for U.S. Transportation Policy and the National Surface Transportation Policy and Revenue Commission's (the Policy Commission) 2008 report entitled Transportation for Tomorrow, have emphasized the need for states to better integrate system performance considerations into their decision-making processes. In *Performance Driven*, for example, the authors conclude that "*substantial efforts will be needed in the area of data quality and data collection to support the rigorous and meaningful application of metrics, and to refine and update them periodically.*"

  Proceedings from subsequent forums that have assessed the implications of moving toward performance-based decision making in transportation include most notably the AASHTO/FHWA-sponsored *Executive Roundtable on Performance-based Planning and Programming* in 2009, the NTPP-sponsored *Transitioning to a Performance-Based Federal*

Surface Transportation Policy workshop in 2010, the AASHTO/FHWA-sponsored *Peer Workshop on Data Needs and Considerations Related to Performance-based Planning and Programming* in 2010, and the AASHTO/AMPO/APTA/FHWA/FTA-sponsored *National Forum on Performance-based Planning and Programming* in 2010, which have all identified the need for better and expanded transportation data to support decision-making.

- **Performance-based Management** – There is a growing focus on performance-based management of transportation systems and agencies, and better data will be needed to support it. As noted in a read-ahead document for the 2009 AASHTO CEO Leadership Forum entitled Performance-Based Management: State-of-the-Practice White Paper, "*High-quality, consistent data are critical to successful performance management and, by extension, to achieving the overall goals and objectives of the agency. Complex, system-level transportation decisions require timely, understandable, and standardized data. Conversely, data that are uncertain or inaccurate reduce the management value of the performance measures they inform.*" Numerous recent NCHRP reports, particularly those in the 20-24 series on asset management and transportation agency management have also noted the need for improved data quality and data program governance to support future agency management needs.

- **Emphasis on Transparency and Accountability** – A more technology and web savvy public is demanding increased transparency and accountability from government at all levels. This means increased demand both for raw and interpreted data and a heightened need to provide data enriched evidence for decisions. This is evident in AASHTO's 2009 reauthorization position papers including *Creating America's Future Transportation System*, which states, "To improve accountability for achieving national objectives, each state will be called on to develop specific measures through which they can track and report on results." As with performance-based decision-making and performance-based management, satisfying this new emphasis will require both improved data quality and timeliness, as well as create new data program demand for access, security, and standardization.

- **Funding Shortfalls** – Revenues for transportation have faced steady, real and inflation-based decline in recent years. As noted in the Policy Commission Report and the National Surface Transportation Infrastructure Financing Commission's 2009 report entitled Paying Our Way, the federal gas tax has not increased since 1993 and many states are facing growing funding gaps. Tighter budgets not only create more pressure for data driven decisions, they make it more difficult for agencies to commit to basic investments in data collection, analysis and maintenance, much less make broader improvements to their overall data governance.

- **Linking National, State, and Local Decisions** – The proceedings from the three performance-based planning forums all discuss the need for greater inter-jurisdictional coordination and collaboration to address challenges and emerging issues such as funding shortages, greenhouse gas reductions, and changing population demographics. These factors are resulting in relationship changes that are already creating an increased emphasis on data sharing, coordination, and standardization.

- **Growing Complexity of Transportation Decision-making** – As discussed in many of the sources cited above as well as other studies and reports on transportation planning and programming, transportation officials are expected to make decisions that incorporate an ever-

growing spectrum of considerations. The focus on economic impacts, livability, climate change, and sustainability requires different data and analysis than has been used in the past when decisions were focused more on just the condition and performance of the transportation system itself. As noted by Alan Pisarski at the *2010 AASHTO/FHWA Peer Workshop on Data Needs and Considerations Related to Performance-based Planning and Programming*, "data collection and analysis for our purposes are reaching a critical moment. Increasingly, we do not have the content, tools, or mechanisms for effective data collection."

Part 2: Data Program Assessment

The second part of the literature review focused on three areas: 1) articles, reports, and other information on topics related to data assessment such as data governance, data management, and data stewardship; 2) approaches and guidance for data program assessment, and 3) examples of data assessments and associated materials used in both the transportation industry and non-transportation fields such as health care and education. The following section summarizes applicable lessons and guidance from the literature that could support development of a framework for self-assessment transportation data programs and is organized into the following elements:

- The Case for data governance;
- Governance structures and the role of data assessment;
- Components of data governance;
- Data assessment approaches;
- Conducting data assessments; and
- Challenges and barriers to good data governance;

Data Program Terminology

As some of the literature explicitly acknowledges, the terminology associated with the management of data programs is undisciplined, overlapping, and often confusing. In particular, terms such as *data governance*, *data management*, *data plans, and data stewardship*, are often used interchangeably by data professionals and data customers alike, yet the terms also appear to mean different things to different people. For purposes of the discussion in this paper, we are assuming the following definitions for these terms:

- *Data Governance* – the overall polices, processes, and structures that define how an organization's data assets are managed;
- *Data Management* – the establishment of data standards and requirements;
- *Data Plans* – the plan for implementing data policies and standards; and
- D*ata Stewardship* – defining who has responsibility for implementing data plans and establishing the authority for them to do so.

The Case for Data Governance

A few universal themes can be found throughout the vast array of articles, reports, whitepapers and web pages that address the overlapping topics of data management, data governance, data stewardship and data plans:

- Data has become an important enterprise asset in both the public and private sector;

- Data is critical to informing strategic decision-making and ensuring the success of organizations, yet it is often highly flawed and nearly all enterprises could benefit from better quality data;

- Organizations often own significant data resources, but fail to convert the data into meaningful information that can drive decision-making;

- Most organizations are currently doing a poor job at managing their data programs with respect to ownership/stewardship, security, consistency, use, general risk management, maintenance, etc;

- Managing data well requires a consistent and on-going process to manage, monitor, maintain, and improve data resources and programs; and

- The need for sound data governance is quickly evolving from elective to mandatory, in many cases driven by laws, regulations, or other requirements.

Taken together, these themes clearly point to a need for both public and private sector leaders and managers to initiate efforts to improve the quality of their data, and to either establish or expand data governance in their organizations. To this end, the literature identified a variety of ways that data governance can help organizations improve performance and succeed. Perspectives on why data governance and assessment is important, and what an organization should strive to achieve through its assessment efforts, however, are different for the public and private sectors. For private sector enterprises, the need for data governance can be directly tied to an organization's bottom line – high quality data and good data governance can help mitigate risk, optimize business and revenue, and control costs. For public sector organizations, however, the need for better is data is tied to improving goods and services the entities provide the public. These differences notwithstanding, there are also several areas where the cited benefits from good data governance cut across private and public sector lines:

- Broadened perspective and better understanding of the strategic and operational importance of data;

- Better, data-driven support decision-making, project delivery, and problem solving;

- Improved alignment between data resources/quality and data user requirements;

- Greater usability and sharing of data;

- Added credibility for data and associated analysis/applications;

- Better balance between long-term/strategic and short-term/tactical improvements in data programs;

- Enhanced organizational capacity to quickly respond to threats and opportunities (e.g., through standardization of data);

- Creation of a framework and culture for continuous improvement of data resources and management (e.g., helps move from a project to an enterprise perspective of data programs);

- Improved quality, maintenance, and sustainability of data;

- Clearer definition of data investment and spending priorities;

- Added prevention of future data problems and risk mitigation;

- Promotion of better and more efficient data usage; and

- Creation of opportunities for cost savings and better allocation of limited data resources.

Governance Structures and the Role of Data Assessment

As noted in a 2007 paper developed by the IBM Data Governance Council, "*many companies are just learning to examine their data governance practices, and searching for industry benchmarks and common frameworks to ground their approach.*" Firms that specialize in data governance issues and public sector entities alike are finding they need to address:

- A lack of cross-organizational data governance structures, policy-making, and risk assessment causing a disconnect between business goals and data programs;

- Governance policies that are not linked to data requirements, data collection, or analysis needs and capabilities;

- Risks are not addressed from a lifecycle perspective with common data repositories, policies, standards and calculation processes;

- Metadata and business glossaries are not used to track data quality, bridge semantic differences and demonstrate the business value of data; and

- Controls, compliance, and architecture are deployed before long-term consequences are properly evaluated.

Clearly, there has been no shortage of thinking about how to respond to these issues – many articles and papers present ways to organize, explain, and conduct data governance initiatives. Stepping back from individual sources, the overall body of literature implies good data governance must be structured as an evolutionary and cyclical process. Consolidating information from the various flow-charts, tables, diagrams, and lists that accompany many of the articles on governance structure, the following appear to be the key steps in a meaningful data governance structure (although the lines between these steps are often blurred):

- **Leadership Support and Direction** – An organization's leaders must recognize and embrace the need for data governance. This means providing adequate resources, granting needed authority, providing input on direction and priorities, and supporting necessary policy and management changes.

- **Assessment –** Developing good governance requires a comprehensive and honest assessment of the current "as is" situation with respect to the various complements of data governance (discussed in detail in the next section). Such an assessment is typically driven by the content and details of a data maturity model.

- **Valuation and Risk** – Assessing the value of data resources to an organization and evaluating the risk different data resources pose to the success and health of an entity.

- **Goal Setting** – An organization must define where it wants to be along the maturity scale with the various components of data governance. Preferably, goals and objectives are established based on fiscal and other practical realities, and reflect both strong input and buy-in from senior leadership.

- **Defining Roles and Relationships** – A key part of governance centers around defining and assigning roles and establishing clear lines of responsibility relative to data resources.

- **Developing Data Plans** – The results of the prior steps should be incorporated into a well-defined data plan that lays out clear steps (investments, new policies, standards and requirements development, changes in organizational structure, etc.) required to move an organization from its current state of governance to a desired level of greater maturity.

- **Plan Implementation** – An organization must stick to its data plan, but remain flexible enough to adapt the plan as needed to respond to emerging threats and opportunities.

- **Re-evaluation** – At some regular interval (or in response to/preparation for a significant change in the operating environment) an organization should repeat the governance/data program assessment process.

Components of Data Governance

Many cross-cutting and recurring themes can be found throughout the literature regarding the attributes and components of effective data governance; it is clear that the same core values surrounding data governance are relevant regardless of the field of expertise (transportation, health care, education, private sector, etc.). Although much has been written in regard to enriching data quality (accuracy, consistency, reliability, etc.), it is also apparent that effective data governance requires a more comprehensive approach to how the data is defined, managed, and interpreted. The findings from the research on the components of effective data governance are summarized below.

- **A Focus on Data Quality** – The most frequently mentioned topic in the literature reviewed was that of data quality, a term that encompasses the desired attributes of an effective data set. A core outcome of effective data governance is an improvement in understanding, measuring, and monitoring the quality of one's data. The literature on this subject is extensive, in depth, and expansive in nature, but a few common attributes of data quality emerged. These components include:

    o **Accuracy** – The data should measure what it is intended to measure and errors should be minimized and easily identified;

    o **Consistency** – Data collection methods should be consistent and the data should remain consistent within and among data sets;

    o **Reliability** – Protocols for collecting and analyzing data should be in place to ensure that the data is reliable, regardless of who is using the information;

    o **Completeness** – Data sets should be appropriately inclusive and duplication should be minimized;

    o **Timeliness** – Data should be available within a time frame that supports its use;

- o **Integrity** – Data should be protected from deliberate bias or manipulation for personal, political, or other reasons; and

- o **Confidentiality** – Appropriate security and disclosure procedures should be in place.

- **Alignment with Strategic Goals –** An effective data program must meet the strategic needs and support the mission, goals, and policies of an organization. For data governance, this means qualifying and quantifying the value of data assets and programs to the core strategic goals of the organization. It also includes establishing processes though which decisions and recommendations can be based on data.

- **Clear Definitions** – Identifying clear and consistent definitions across data programs is important to ensure that the data sets are used and interpreted correctly. Definitions of a particular piece of data can often vary across organizational and programmatic boundaries. A common example is that of "revenue" – a finance department might have one definition of revenue, while an accounting department might have another, and a programming department might have yet another. These differences might be wholly appropriate given the applications within each department, but a clear and consistent identification of these differences is important to ensure that the data is used appropriately.

- **Ability to Segregate, Aggregate, and Analyze Data Longitudinally** – A necessary precursor to data analysis is a flexible data set that can be parsed, parceled, and constituted in whatever way is the most relevant to the desired application. Coordination in this area is necessary between and within organizations if data sets are to be analyzed and distributed across internal and external jurisdictions.

- **Regular Audits and Validation Procedures** – Errors will inevitably occur within any data program and audits and validation must be continuous and ongoing. As a part of data governance, time and resources should be allocated to provide continuous examination of data for accuracy, consistency, reliability, etc. It is also important that discrepancies are properly documented to ensure that the process can be duplicated and reconstructed if necessary, and that corrections are carried throughout the affected programs and analyses.

- **Identification of Data Sources, Uses and Users** – Understanding exactly where all of the data is coming from, who is using it, and for what purpose helps eliminate inefficiencies and demonstrates a clear tie between the sources and the expected outcomes. This mapping exercise is often done in the form of one or more diagram or flow chart.

- **Repeatable Data Governance Plans –** The data governance plan should be repeatable even when policies, strategic goals, and leadership change. This is not to say data plans should not be reevaluated often to ensure that they are meeting the evolving needs of an agency, but the core principles of the data governance plan should be adaptable regardless of how circumstances change.

- **Clear and Appropriate Organizational Roles –** It is important that a data governance plan be integrated within an organization. The roles of leadership, managers, data stewards, IT staff, data steering committees, etc. need to be clearly identified and articulated. Consideration should be given to the availability, capability, and training of appropriate staff. This might not apply just to

those that are tasked with managing and analyzing the data. For example, an issue often discussed in education circles is that teachers and administrators are expected to use data to inform their approaches, yet many in this field have not been trained to adequately understand data analysis.

- **Consideration of Costs and Life-Cycles –** Collecting, maintaining, and analyzing data comes with a cost. Data should be periodically analyzed to ensure that the time, staff, and monetary costs related to collecting a data set or managing a data program are appropriate. Inherent in this is recognition of when legacy data programs that are no longer providing a suitable benefit can be retired and opportunities can be identified that shift resources away from collecting and maintaining data that is not necessary.

- **Time and Resources for Conducting Analysis** – Data alone does not provide understanding. The missing link between the data managers and the decision makers is often the transformation of data into information. Good data analysis provides for an understanding of the relationships between data and desired outcomes through clear explanation and visualization techniques. A data plan should dedicate the necessary resources and provide guidance to ensure that data is being evaluated in a manner that is consistent with the desired application.

- **Mechanisms for Security, Privacy, and Ethical Considerations –** An effective data governance plan must provide for adequate protection of data assets due to security, privacy, and proprietary concerns. Ethical considerations must also be applied when publishing analysis and visualizations to ensure that the results and conclusions are not manipulated to unfairly endorse a specific agenda.

Data Assessment Approaches

Literature on data assessment approaches and guidance largely focuses on the use of maturity modeling. As described in numerous articles and web pages, the concept of data maturity modeling grew out of the Capability Maturity Model (CMM), developed by the Software Engineering Institute in 1984, and can be applied to almost any data program, regardless of the application. Broadly defined, maturity modeling is the process of establishing a graduated path for improvement of an organization's data program governance activities (policies, standards, assignment of responsibilities, etc.), and then applying the framework to assess both where an organization currently is and where it needs to go. Examples of this desired progression include:

- From reactive to proactive data and information management;
- From project/ad hoc initiatives to enterprise-wide solutions;
- From "siloed" data to synchronized data and information;
- From localized systems with inconsistent levels of data classification and security to consistent data classification and standards based security; and
- Migration to the capability to build efficient information and knowledge management

Based on the literature, potential outputs and benefits from using a maturity model could include:

- Better organizational understanding of data governance and its importance;

- Establishment of an incremental and evolutionary approach to improvement, including a process for determining performance measurement needs;

- A clear and objective assessment on the current state of an organization's data program governance, including capabilities, strengths, weaknesses, opportunities and threats;

- A means to benchmark against peer organizations;

- Definition of applicable best practices and an incremental path that will help move an organization to the next stage of maturity;

- Establishment of short- and long-term data program goals and objectives;

- Identification, quantification/qualification, and articulation of potential benefits from improvements to data resources and data program management;

- Establishment of the purpose and objectives of an assessment approach;

- Normalization of methods for continuously examining business processes and IT practices; and

- Documentation and centralization of reference information.

The original CMM identifies five levels of maturity; most of the maturity models that build from the CMM suggest maturity scales with four to six discrete steps that vary in complexity, focus, and comprehensiveness. There does not appear to be a specific "right" answer to the scale and scope of a maturity model, however, there are five common increments that are included in most approaches (a summary of several of the maturity models reviewed is summarized in table 1):

6. **Undisciplined** – No formal data governance approach;

7. **Ad hoc** – Data governance occurs, but is typically in response to a specific problem or project rather than applied organization-wide;

8. **Standardized** – Organization has implemented some standard data governance practices and roles;

9. **Proactive** – Organization has instituted formal data governance policies and goals for all business units, has dedicated stewards for all business units, and has begun to both coordinate across the enterprise and measure/predict data program performance; and

10. **Optimized** – Data governance is continually improved and used to set, communicate, and enforce business policies and strategic direction.

Table 1: Summary of Data Governance Maturity Models

Maturity Model	Maturity Levels	Emphasis/Attributes
Gartner Enterprise Information Management Maturity Model	1. Unaware 2. Aware 3. Reactive 4. Proactive 5. Managed 6. Effective	• Need to manage enterprise data as a business asset • Identification of action items for each level • Defines 5 core goals – data integration across IT portfolios, unified content, integrated mater data domains, seamless information flows, and metadata management and semantic reconciliation
IBM Data Governance	1. Initial 2. Managed 3. Defined	• Direct link to original CMM levels • Focus on 11 "data domains" organized as

Council Maturity Model	4. Quantitatively Managed 5. Optimizing	"outcomes,""enablers,""core disciplines," and "supporting disciplines" • Importance of continual improvement
the EWSolutions Data Governance Maturity Model	1. Informal Processes 2. Emerging Processes 3. Engineered Processes 4. Controlled Processes 5. Optimized Processes	• Model as a tool for communication and training • Performance monitoring • Defining roles
DataFlux Data Governance Maturity Model	1. Undisciplined 2. Reactive 3. Proactive 4. Governed	• Reducing "bad data" risk • Phases defined in terms of people, policies, technology and risk • Need to manage enterprise data as a business asset • Defines steps to get to next level
Knowledge Logistics Data Governance Maturity Model	1. Initial 2. Repeatable 3. Defined 4. Managed 5. Optimized	• Close link to original CMM levels • Predictive performance
MDM Institute Data Governance Maturity Model	1. Basic (anarchy) 2. Foundational (IT monarchy) 3. Advances (business monarchy) 4. Governed (federalist)	• Migration from the initial state • Leveraging service-oriented architecture • Master data management
Oracle Institute Data Governance Maturity Model	1. Marginal 2. Stable 3. Best Practice 4. Transformational	• Aligning people, processes, and technology

Data Assessment Examples

As data-driven approaches to decision making have become more prevalent, many organizations have developed assessment tools to evaluate their data quality and governance capabilities. The emphasis of a particular assessment varies depending on its scope and purpose. The one objective that is consistent across all of the assessments reviewed, however, was improved data quality. Every example assessment was concerned with data quality, some almost exclusively, and the general approach to assessing data quality by way of analyzing processes and policies was relatively consistent across the literature. Less consistent was the issue of aligning data quality and governance policies with the strategic intent of the organization – some assessments include a strong strategic component, others assume strategic intent has already been established, and still others nearly ignore the issue altogether.

Based on examples of data assessment experiences, organizations frequently initiate their assessments by conducting multiple choice surveys that makes statements or asks questions about the conditions of various data governance and management aspects within an organization. Some questionnaires are organized around a formal maturity model approach; others are similar in nature to a maturity model but less structured. The most comprehensive assessments also include guidance for analyzing the results and for developing an implementation plan for improvement.

In reviewing the literature, there was a noticeable lack of case studies or post-assessment documentation about how successful a particular data assessment approach was. This lack of "assessment of

assessments" makes it difficult to determine the strengths and weaknesses of different approaches. One factor that diminishes this concern is that while the approaches vary slightly, at their core most are quite similar.

Education Example: Assessment /Accountability Comprehensive Center Data Needs Assessment

The transportation industry has seen an increased focus on accountability in recent years, but even more scrutiny has been put on the education system, due in no small part to the passage of *No Child Left Behind* and its goal of academic proficiency for all students. As a consequence, the education system has seen an increase in the need for data to track and demonstrate improvement in student achievement.

Not surprisingly, the increased focus on data has led to significant efforts to improve the quality and understanding of education data. One of the tools that has emerged is the Data Needs Assessment developed by the Assessment and Accountability Center (AACC), a component of a federal technical assistance system established to develop assessment and accountability methods to diagnose needs, guide instruction, monitor student and program progress, and support instructional and administrative needs.

The Needs Assessment is a tool developed for use by state-level Regional Comprehensive Centers (RCCs), District- and site-level administrators, and teachers to help evaluate the extent to which "key capacities" for data use are in place. The Needs Assessment includes five areas of focus corresponding to each of the five key capacities. These key capacities are as follows:

- **Leadership** – The capacity of leadership as it relates to data;

- **Quality Assurance** – the quality of data and capacity for its use;

- **Infrastructure** – The existence of tools, technical personnel, and professional development to support effective data use;

- **Communication** – The ability to build support for data-driven decisions through the educational system; and

- **Action for Learning** – The capacity to use data to take positive steps to actually improve learning.

The Needs Assessment itself provides a series of questions within each key capacity. Each question is answered on a sliding scale as to the extent that a statement is true within an organization. The questionnaire is available for download, but is also presented via an online submission form that automatically generates a report complete with visualization of the results, a sample of which is shown in.

After the questionnaire is completed, resources are available to interpret the results and provide direction on how to build capacity in each area. The resources encompass a spectrum of research-based reports and more practical how-to guides. Finally, a template is provided to establish a transformative and monitoring plan that outlines the processes, procedures, resources, and anticipated outcomes of improving the organization's key data capacities.
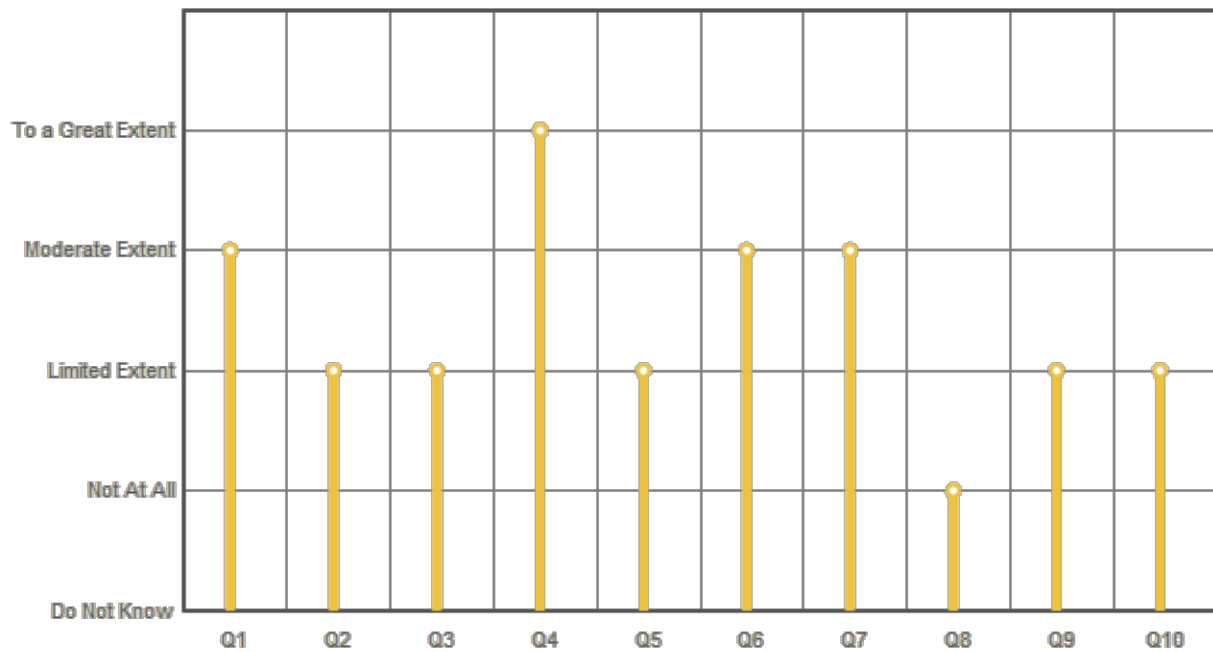
General Needs Assessment – Individual Score: 13 / 27

Figure 2 for illustrative purposes

Cases studies from four states, Arkansas, Iowa, New York, and Montana, which discuss the implementation of the Data Needs Assessment, were also reviewed. In each case, participants were given guided practice in how to access, analyze, and make statements about their educational data. The New York case study offered several lessons learned, which can be summarized as follows:

- The Needs Assessment was particularly successful at determining if sufficient resources are present to use data;

- Training plays a significant role in not only assessing data needs, but also in the implementation of the resulting transformative plan. To be successful, staff must develop a firm foundation of data literacy and a clear understanding of the assessment needs, goals, and implementation process.

- The identification of leaders responsible for coaching and supervising the implementation was also important;

- Goals for improvement should be realistic and recognition should be given to the time necessary to successfully implement the plan; and

- Data does not replace expertise and should be viewed as just one tool in a broader decision making methodology.

International Health Care Example: Global Fund to Fight Aids, Tuberculosis, and Malaria - Routine Data Quality Assessment Tool

Although the focus of this assessment is on a few specific components of overall data governance, the goals of the assessment and scale of the collaboration required to develop and implement it are impressively grand in nature. The Global Fund to Fight Aids, Tuberculosis, and Malaria partnered with numerous national and international partners, including USAID, the World Health Organization, the World Bank, and others to develop the "Routine Data Quality Assessment Tool" (RDQA), the purpose of which is to provide a common approach for assessing and improving overall data quality associated with these diseases.

The worldwide fight against diseases such as AIDS, Tuberculosis, and Malaria is heavily reliant on data. Understanding where these diseases are most prevalent and who is contracting them is critical to effectively combating them. Because the effort spans so many organizations and countries, it is essential that this data be collected consistently so that it may "roll up" from the local level to regional, national and international levels. The primary challenges are that those collecting and reporting the data worldwide have a wide variety of backgrounds and resources, and although formal audits do take place, they are not frequent enough to comprehensively manage data programs that literally span the globe.

The RDQA tool was designed to allow for the self-assessment of data quality, to build capacity for data governance within an organization, and to prepare for external formal audits. Specifically, the objectives of the tool are to verify the quality of reported data and to assess the ability of data management systems to collect and report quality data. The tool also provides the means to develop action plans to improve both of these areas.

The RDQA tool comes with a very detailed and comprehensive step-by-step guide that includes specific technical instructions, such as how to select a suitable sample size and how to perform necessary statistical analysis, and process guidance such as how much time to allow for each step of the self-assessment process. The assessment itself consists of two separate components, one for assessing the data management and reporting systems, and one for assessing the quality of the data itself.

The first component, for assessing the data management and reporting systems, is similar in nature to a maturity model. A series of thirteen questions are asked within five distinct categories. The categories and questions are as summarized in Figure 3.

Category	Question
1. Staff	1. Are data management staff clearly identified and assigned responsibilities? 2. Have the majority of staff received the required training?
2. Definitions and Reporting	3. Are the definitions meeting relevant standards and consistently followed? 4. Is there clear documentation of what is to be reported when and how?
3. Data Collection and Reporting Forms and Tools	5. Are there standard data-collection forms that are used systematically? 6. Is data recorded with sufficient precision and detail? 7. Are data maintained with international confidentiality guidelines? 8. Are source documents kept and made available?
4. Data Management Process	9. Does clear documentation of collection, aggregation, and analysis exist? 10. Are data quality challenges identified and addressed? 11. Are there clearly defined procedures to identify discrepancies? 12. Are there clearly defined procedures to verify source data?
5. Link to National Reporting System	13. Does the data collection and reporting system link to the National Reporting System?

The assessment is designed to be utilized at four different levels within the organization. The first assessment occurs during a review of a program's documentation. The other three reviews are all conducted on-site at different levels of the program: on the ground where the data is collected, at the district level where data is initially reported, and at the regional level where it is further aggregated. Each of the questions is answered at each level of the organization with a simple "Yes, completely," "Partly," or "No, not at all" response. Answers are then given a "3," "2," or "1," respectively for the purposes of scoring them. The tool then provides simple visualization techniques of aggregating the results

Figure 4: RDQA Assessment Results
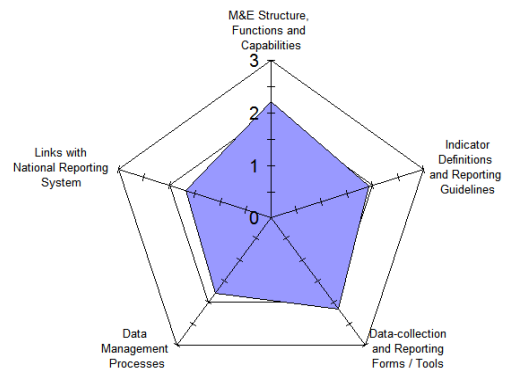
for comparison across functional areas, one of which is shown in Figure 4.

The second component of the assessment provides a worksheet and technical documentation to assist an organization in determining if data has been accurately recorded and reported. The worksheets provide specific step-by-step instructions to re-aggregate reported numbers from collection sites to determine if the data is accurate and consistent. The guidance includes clear and concise instructions to assist in the necessary statistical analysis, and the worksheets provide clear indications of where inaccuracies and inconsistencies occur.

The RDQA tool stresses that the purpose of the assessment and the subsequent result is to provide a means of prioritizing system strengthening activities, much in the way of a traditional maturity model. Example findings and recommendations are also provided to assist an organization in developing a plan to improve data management systems and quality.

Although the scope of the AACC Data Needs Assessment is probably too specific to translate directly to a general framework for an assessment of transportation data, it provides several lessons about how to develop and conduct a meaningful assessment:

- To be applied correctly, an assessment should include concise instructions and training for staff on how to apply and use the assessment;

- Data assessments can and should be applied comprehensively throughout multiple levels of an organization;

- Clear and simple visualization techniques should be used to assist in understanding the assessment results; and

- An effective assessment should be accompanied by resources that help guide the development of a data management plan and the implementation of that plan.

Transportation Example: Alaska Department of Transportation & Public Facilities Data Business Plan

The Program Development Division of the Alaska Department of Transportation and Public Facilities (ADOT&PF) is currently in the midst of implementing a comprehensive Data Business Plan to guide management of strategic data programs that are used to support the core business functions of the Division. These core functions, highway safety, traffic, road weather management, 511 traveler information, and GIS services, support the Transportation Planning Mission to *"optimize state investment in transportation by means of data-driven recommendations and meet federal and state requirements through effective data collection, analysis, planning, public involvement, and documented decisions."*

Although the Data Business Plan is not scheduled to be fully developed until 2011, many of the key assessment steps and implementation strategies are already in place. The process began in 2007 with an in-depth assessment of the performance measures currently used to monitor the core data systems, including detailed analysis of what data is needed in each program area. Subsequently, three other white papers have been published that further inform the Data Business Plan. *Assessing the Value of the ADOT&PF Data Programs* explores why the core business functions exist and what data programs are used to support those functions. *Data Sharing and Data Deliver Methods and Tools* focuses on data and

information sharing within the organization. The third paper, *Data Governance, Standards, and Knowledge Management* examines how a data governance model and knowledge management structure can be implemented.

To develop the Data Business Plan, the agency used a maturity model to develop an understanding of the data management goals and how data programs aligned with the organization's goals and mission. This activity also provided a long-range view of data management programs. A key step was to establish well-defined roles and responsibilities for all data program participants and to clearly map each of the data programs to their uses, users, and sources.

These efforts led to a data governance and knowledge management plan with several implementation recommendations, which are summarized below:

- Establish a data governance board and identify executive leadership;

- Establish a data governance charter;

- Identify internal data work groups and external stakeholders and communities of interest;

- Publish and annually review a Data Governance Manual and offer implementation training to staff;

- Publish and continue to update any data standards, policies, procedures, definitions that result from the use of data governance;

- Conduct a risk assessment of existing data programs to determine priorities for addressing negative impacts due to unavailability of data systems; and

- Develop and implement an integrated data warehouse that will streamline and simplify knowledge management by minimizing duplication of data and improving access to data.

The efforts of ADOT&PF to establish a Data Business Plan are extensive and are based on the perspective that good data governance and management takes commitment from agency leadership, a long-term vision, and a firm and steady application of implementation strategies.

Challenges and Success Factors to Data Governance and Data Program Assessment

While the literature is generally positive about the potential for organizations to establish a successful data governance framework and conduct meaningful data assessments, most articles and papers also note that formidable barriers may need to be addressed or overcome. Moreover, there are several considerations that are unique or more pronounced in the public sector. Based on the reviewed materials, the following is a composite of potential challenges organizations may need to address as they develop and implement data governance reforms:

- **Management Indifference** – Management may fail to understand or appreciate the need for and relevance of data governance and/or provide inadequate support and sponsorship assessment;

- **Weak Vision –** Management may be unwilling or lack the ability of to provide a clear strategic vision for an organization's data programs;

- **Management Turnover** – Frequent turnover in leadership and/or poor succession planning can hurt the continuity of data governance improvement initiatives;

- **Short Management Horizon** – Organizations and leaders tend to focus on short term gains, preventing data program managers from addressing long-terms needs and solving problems from a life-cycle perspective;

- **Consensus – building** – Gaining agreement of all necessary parties regarding policies can be difficult and time consuming, particularly when there are competing priorities within the organization;

- **Capacity Limitations** –The development or procurement of new tools and software may be required to enable data governance;

- **Financial Constraints** – The cost of conducting data assessments and implementing governance policies can be prohibitive, particular during periods when agencies are under significant pressure to cut costs;

- **System Incompatibility** – Legacy systems with unique architectures may be difficult to accommodate under universal governance policies;

- **External Requirements** – Particularly in the public sector, factors such as statutory and administrative requirements, grant-related stovepipes, and public demands for transparency and access to information may preclude the application of practical data governance solutions;

- **Cultural Barriers** – A comprehensive data governance approach often requires culture change to adapt the organization;

- **Task Magnitude** – The amount of work involved in developing a data governance plan is often underestimated;

- **Poor Implementation** – Data assessment and governance efforts are often long on structure and policies, and short on action;

- **Organizational Resistance** – Both data professionals and data users often resistance to migration of data from silos to an enterprise management system;

The literature also speaks to several factors that will dictate the success of a data governance framework and assessment. Many of these factors are reliant on strong organizational and leadership structures, and a clear understanding of the roles within the organization. Based on the literature, the following is a summary of the factors that will determine the success of a data governance plan:

- **Strong Executive Leadership** – Like any comprehensive organizational initiative, leadership and support from the top of the agency is critical;

- **Solid Partnerships within the Organization** – Data governance issues affect, and are affected by, many groups within an organization, Collaboration and cooperation amongst information technology and the business units of the organization is key;

- **Good Communication** – It is important that the various communities of interest regarding the data and associated application systems that are used to collect, maintain, and report information are able to communicate and share ideas and resources;

- **Published Definitions and Standards** – Source data, metadata, and data reports from the various systems not only need to be clearly defined and standardized, but also formally adopted, published, and distributed; and

- **Data Champion** – Those within the organization that understand and appreciate the importance of data governance need to be actively involved in the process and engaged in communicating the importance to others.

Observations and Conclusions

The literature review revealed that the strategic importance of data to transportation leaders is clearly growing and is driving the need for new data, improved data quality, and better data governance. A wealth of information is also available about the general importance and benefits of good data governance and data assessment practices. Finally, our research identified several possible models and approaches for initializing and implementing a data governance framework.

The increased emphasis on data in our society has meant that most organizations, both public and private, can and must dramatically improve their data assessment and governance approaches. Aside from the obvious improvements to data quality and usability, there are compelling benefits of a good data governance structure, such as the ability to make better decisions and to deliver better projects. Conversely, there are also risks for not adequately governing one's data, such as losses in credibility with a public that expects more accountability, and an inability to quickly respond to threats and opportunities as they arise.

As the needs for better data governance are beginning to resonate within public and private organizations, the resources available to address and implement better data governance are in many cases tightening. The continued pressure on DOTs to continue to cut costs, save money, and defer investments in institutional resources and capabilities will affect their ability to implement effective data governance.

The literature review has uncovered several implications for the development of a framework for data self-assessment tool for transportation agencies. The framework must be intertwined with a proposed structure for general data governance – you cannot have one without the other. There are several well-established models for governance and assessment that we can build on for transportation, both in terms of data governance structures and maturity models. However, there is limited guidance from the literature, and limited experience within the DOT world, on how to actually build a governance structure and maturity model that is tailored to the unique needs of the transportation industry.

In order for transportation agencies to be successful in developing and deploying good governance and data assessment frameworks, a host of challenges and barriers must be overcome. Many of these barriers are organizational and cultural in nature, and defeating them will require strong leadership and long term commitment. It may be most important of all to recognize that implementation of an effective data governance and management plan is an iterative process, and application of this study and any subsequent efforts will take a long-term commitment to achieving a mature and governed data structure. The benefits for doing so are clear, but the endeavor is not something that happens quickly and without significant investment.

Bibliography

AASHTO, *Creating America's Future Transportation System*, 2009

Cambridge Systematics, Inc., *Assessing the Value of the ADOT&PF Data Programs,* Prepared for the Alaska Department of Transportation and Public Facilities, Juneau, Sep 2009.

Cambridge Systematics, Inc., *Data Governance, Standards, and Knowledge Management,* Prepared for the Alaska Department of Transportation and Public Facilities, Juneau, Sep 2009.

Cambridge Systematics, Inc., *Data Sharing and Data Delivery Methods and Tools,* Prepared for the Alaska Department of Transportation and Public Facilities, Juneau, Sep 2009.

Cambridge Systematics, Inc., High Street Consulting Group, and Paul N. Bay, P.E., *Performance-Based Planning: A State-of-the-Practice Summary,* Prepared for the National Cooperative Highway Research Program, Washington, D.C., Sep 2010.

Cambridge Systematics, Inc., *Performance-Based Management: State-of-the-Practice White Paper,* Prepared for the University of Minnesota Center for Transportation Studies for the 2009 AASHTO CEO Leadership Forum, Apr 2009.

Comprehensive Everglades Restoration Plan Quality Assurance Systems Requirements, U.S. Army Corps of Engineers and South Florida Water Management District, Mar 2009.

Data-Driven Decision Making for District and School Level Data Teams, Connecticut State Department of Education, Hartford, 2010.

Data Governance Part II: Maturity Models – A Path to Progress, National Association of State Chief Information Officers, Lexington, Kentucky, 2009.

Data Use for Improved Learning Needs Assessment, Assessment and Accountability Comprehensive Center, 2010. < http://datause.cse.ucla.edu/needs.php>

DataFlux Corporation, *A Do-It-Yourself Guide for High-Value Data Across the Enterprise,* A DataFlux White Paper, 2010.

DataFlux Corporation, *The Data Governance Maturity Model: Establishing the People, Policies and Technology That Manage Enterprise Data,* A DataFlux White Paper, 2010.

Dyché, J., *Five Models for Data Stewardship,* Baseline Consulting Group, 2009.

Ferguson, M., *Data Ownership and Enterprise Data Management: Implementing a Data Management Strategy,* A DataFlux White Paper, 2010.

Goulden, M., *Report on the Institutional Data Management Government Survey,* Prepared for the University of California Berkeley, May 2009.

FHWA Strategic Plan, Federal Highway Administration, U.S. Department of Transportation, Washington, D.C., Mar 2010.

How Ready Are You? A Special Tool for School District Leaders to Assess Readiness to Use Data Effectively, The Consortium for School Networking, Washington, D.C., Aug 2010

IBM Corporation, *The IBM Data Governance Council Maturity Model: Building a Roadmap for Effective Data Governance,* 2007.

Kahn, B., D. Strong, R. Wang, "Information Quality Benchmarks: Product and Service Performance," *Communications of the ACM,* Apr 2002.

Lee, Y., and D. Strong, "Knowing-Why About Data Processes and Data Quality," *Journal of Management Information Systems,* Winter 2003.

Lee. Y., L. Pipino, D. Strong, and R. Yang, "Process-Embedded Data Integrity," *Journal of Database Management,* Jan-Mar 2004.

National Surface Transportation Policy and Revenue Commission, *Transportation for Tomorrow*, 2008

National Transportation Policy Project, *Performance Driven: A New Vision for U.S. Transportation Policy*, 2009

National Transportation Policy Project, *Transitioning to a Performance-Based Federal Surface Transportation Policy – Workshop Summary Proceedings*, 2010

Practice Brief: Data Quality Management Model, American Health Information Management Association, Chicago, Illinois, Jun 1998.

Reform for the Federal Surface Transportation Program, American Association of State Highway and Transportation Officials, Washington, D.C., 2009.

Resource Systems Group, Inc., *Peer Workshop on Data Needs and Considerations Related to a National Performance-Based Planning Process*, Summary Report, Prepared for Federal Highway Administration, Held in San Antonio, Texas, Jul 2010.

Routine Data Quality Assessment Tool: Guidelines for Implementation, The Global Fund to Fight Aids, Tuberculosis and Malaria, 2008.

Smith, A., "Data Governance Maturity – An Overview," *EIMInsight,* Dec. 2007.

University of London Computer Centre, *The AIDA Self-Assessment Toolkit,* Prepared for the Joint Information Systems Committee, Bristol, UK, Feb 2009.

Wayman, J., S. Midgley, and S. Stringfield, *Collaborative Teams to Support Data-Based Decision Making and Instructional Improvement,* Prepared for the American Educational Research Association, Washington D.C., 2005.

Appendix B: Project Participants

Transportation sector practitioners interviewed for this project included the following:

- Tom Barry, Senior Vice President of Transportation at PBS&J
- Jennifer Finch, Transportation Development Director, Colorado DOT
- Mark Hodges, Director of Transportation Analysis, Arizona DOT
- Daniel Holm, Solutions Architect at Interworks, Inc.
- Jonette Kreideweis, Director of Transportation Data and Analysis, Minnesota DOT
- Vladimir Livshits, System Analysis Program Manager, Maricopa Association of Governments
- Deb Miller, Secretary of the Kansas DOT
- Hyun-A Park, President of Spy Pond Partners
- Alan Pisarski, Transportation Consultant
- Scott Richrath, Performance & Policy Analysis Unit Manager, Colorado DOT
- Alan Spicer, Assistant Bureau Chief in Transportation Planning, Kansas DOT